DIGITAL MORPHOGENESIS VIA SCHELLING SEGREGATION

GEORGE BARMPALIAS, RICHARD ELWES, AND ANDREW LEWIS-PYE

Abstract. Schelling's model of segregation looks to explain the way in which particles or agents of two types may come to arrange themselves spatially into configurations consisting of large homogeneous clusters, i.e. connected regions consisting of only one type. As one of the earliest agent based models studied by economists and perhaps the most famous model of self-organising behaviour, it also has direct links to areas at the interface between computer science and statistical mechanics, such as the Ising model and the study of contagion and cascading phenomena in networks.

While the model has been extensively studied it has largely resisted rigorous analysis, prior results from the literature generally pertaining to variants of the model which are tweaked so as to be amenable to standard techniques from statistical mechanics or stochastic evolutionary game theory. In [5], Brandt, Immorlica, Kamath and Kleinberg provided the first rigorous analysis of the unperturbed model, for a specific set of input parameters. Here we provide a rigorous analysis of the model's behaviour much more generally and establish some surprising forms of threshold behaviour, notably the existence of situations where an *increased* level of intolerance for neighbouring agents of opposite type leads almost certainly to *decreased* segregation.

George Barmpalias

State Key Lab of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing, China. School of Mathematics, Statistics and Operations Research, Victoria University of Wellington, New Zealand.
E-mail: barmpalias@gmail.com. *Web:* http://barmpalias.net

Richard Elwes

School of Mathematics, University of Leeds, LS2 9JT Leeds, United Kingdom.
E-mail: r.h.elwes@leeds.ac.uk. *Web:* http://richardelwes.co.uk

Andrew Lewis-Pye

Department of Mathematics, Columbia House, London School of Economics, Houghton Street, London, WC2A 2AE, United Kingdom.
E-mail address: A.Lewis7@lse.ac.uk. *Web:* http://aemlewis.co.uk

An extended abstract of this paper with most proofs and some results omitted, has appeared in the Proceedings of the IEEE Symposium on Foundations of Computer Science (FOCS) 2014. Authors are listed alphabetically. Barmpalias was supported by the 1000 Talents Program for Young Scholars from the Chinese Government No. D1101130. Partial support was also received from the China Basic Research Program No. 2014CB340302. Lewis-Pye was supported by a Royal Society University Research Fellowship.

1. Introduction

While Alan Turing is best known within the mathematical logic and computer science communities for his work formalising the algorithmically calculable functions, it is interesting to note that his most cited work [26] is actually that relating to morphogenesis. Turing wanted to understand certain biological processes: the gastrulation phase of embryonic development, the process whereby dappling effects arise on animal coats, and phyllotaxy, i.e. the arrangement of leaves on plant stems. One can consider the more general question, however, as to how morphogenesis occurs – how structure can arise from an initially random, or near random configuration. Along these lines, one of the major contributions of the economist and game theorist Thomas Schelling was an elegant model of segregation, first described in 1969 [23], which turns out to provide a very simple model of such a morphogenic process. This model looks to describe how individuals of different types come to organise themselves spatially into segregated regions, each of largely one type. Today it has become perhaps the best known model of self-organising behaviour, and was one of the reasons cited by the committee upon awarding Schelling his Nobel memorial prize in 2005.

For Schelling, part of the significance of his model was that it provided evidence for a recurrent theme in his research – especially as elaborated upon in his influential work [24] – that individuals acting according to their local interests can produce global results which are undesired by all. Running small simulations of his model, for example, he observed that large levels of segregation can result in communities, when each individual has no preference for segregation, but still requires a certain low proportion of their own local neighbourhood to be of their own type. Although the explicit aim was initially to model the kind of racial segregation observed in large American cities, Schelling himself pointed out that the analysis is sufficiently abstract that any situation in which objects of two types arrange themselves geographically according to a certain preference not to be of a minority type in their neighbourhood, could constitute an interpretation. As pointed out in [28], for example, Schelling's model can be seen as a finite difference version of differential equations describing interparticle forces and applied in modelling cluster formation. Many authors (see for example [25, 9, 22, 12, 19]) have pointed out direct links to spin-1 models used to analyse phase transitions – by introducing noise into the dynamics of the underlying Markov process one can arrive at the Boltzmann distribution for the set of possible configurations, with the 'energy' typically corresponding to some measure of the mixing of types. From there one can immediately deduce that the modified (now ergodic) process spends a large proportion of the time in completely segregated states, with this proportion tending to 1 as the analogue of the temperature is taken to 0. So had Schelling been aware of these connections to variants of the Ising model, he could have based his work on a long history of physics research.

Our own avenue into these questions, however, came via the work of computer scientists [5] and the study of cascading phenomena on networks as studied by Barabasi, Kleinberg and many others, a good introduction to which can be found in [16]. The dynamics of the Schelling process (as will be clear once it has been formally defined below) are almost identical to many of those used to model the flow of information or behaviour on large social or physical networks such as the internet or the web of social contacts via which disease may spread – the principal differences here being the initial conditions considered and the use of a much simpler underlying graph structure. An immediate concern, given the results of this paper, is as to whether techniques developed here can be extended and applied to understand emergent phenomena on the various (normally more complex) random graph structures studied by those in the networks community. Along these lines, Henry, Prałat and Zhang have described a simple but elegant model of network clustering [13], inspired by the Schelling model. Their model doesn't display the kind of involved threshold behaviour, however, that one might expect to be exhibited by a direct translation of the Schelling process to random underlying graph structures.

In [8] Clark and Fossett give an account of the role that Schelling's model has had on the debate concerning anti-discrimination policy and the ongoing battle against residential segregation, along ethnic lines and others. Certainly there has been increased governmental recognition of the fact that segregation has become one of the most important socio-political and public economic issues, with residential segregation having knock-on effects in terms of education, underachievement and disadvantages in the labour market [18] and healthcare [4]. In order to inform policy choices in combatting segregation it is essential to understand the extent to which this is an outcome caused by voluntary residential preferences on the one hand, or by constraints on choices resulting from discrimination inherent the system on the other. Schelling's model (and more sophisticated variants, see for example [31]) have been used to argue that

one might expect segregation to persist even in the absence of systemic discrimination. While the simplicity of the model might initially suggest that one should be wary of rushing to draw real world conclusions, the robustness of the segregation phenomenon is certainly striking and occurs in all versions of the model which have been considered. Somewhat counter-intuitively, Zhang [32, 33] and Pancs and Vriend [21] have even shown segregation to result in variants of the model in which individuals *actively prefer* integration.

We concentrate here on the one-dimensional version of the model, as in [5]. The model works as follows. One begins with a large number n of nodes (individuals) arranged in a circle. Each node is initially assigned a *type*, and has probability $\frac{1}{2}$ of being of type α and probability $\frac{1}{2}$ of being of type β (the types of distinct individuals being independently distributed). We fix a parameter w, which specifies the 'neighbourhood' of each node in the following way: at each point in time the neighbourhood of the node u, denoted $\mathcal{N}(u)$, is the set containing u and the w-many closest neighbours on each side – so the neighbourhood consists of $2w + 1$ many nodes in total. The second parameter $\tau \in [0, 1]$ specifies the proportion of a node's neighbourhood which must be of their type for them to be happy. So, at any given moment in time, we define u to be *happy* if at least $\tau(2w + 1)$ of the nodes in $\mathcal{N}(u)$ are of the same type as u. One then considers a discrete time process, in which, at each stage, one pair of unhappy individuals of opposite types are selected uniformly at random and are given the opportunity to swap locations. Following [5], we work according to the assumption that the swap will take place as long as each member of the pair has at least as many neighbours of the same type at their new location as at their former one (note that for $\tau \leq \frac{1}{2}$ this will automatically be the case). The process ends when (and if) one reaches a stage at which no further swaps are possible.

Much of the difficulty in providing a rigorous analysis stems from the large variety of absorbing states for the underlying Markov process. Various authors have therefore worked with variants of the model in which perturbations are introduced into the dynamics so as to avoid this problem, i.e. the model is altered by introducing a further random element, allowing individuals to sometimes make moves which are detrimental with respect to their utility function. Such changes in the model dramatically simplify the dynamics, and might be justified by the assumption that we are dealing with individuals of 'bounded rationality' – one might consider that precise information concerning the racial composition of each neighbourhood is not available, for example. As remarked above, these discussions are often couched in the language of statistical mechanics. In fact though, Young used techniques from evolutionary game theory – an analysis in terms of stochastically stable states – to develop the first results along these lines [30], and these ideas were then substantially developed in a number of papers by Zhang [31, 32, 33]. While the language used may differ from that of the physicists, the basic analysis is essentially equivalent: Zhang establishes a Boltzmann distribution for the set of configurations, and then his stochastically stable states correspond to ground states.

In [5], Brandt, Immorlica, Kamath and Kleinberg used an analysis of locally defined stable configurations, combined with results of Wormald [29], to provide the first rigorous analysis of the unperturbed one-dimensional Schelling model, for the case $\tau = \frac{1}{2}$. The results obtained there are very different to those for the perturbed models: in the final configuration the average length of maximal segregated region is independent of n and only polynomial in w. So now the local dynamics do not induce global segregation in proportion to the size of the society but instead only a small degree of segregation at the local level. The suggestion is that these results are in accord with empirical studies of residential segregation in large populations [7, 11, 27].

Our contribution. In this paper we shall consider what happens more generally for $\tau \in [0, 1]$ (for the unperturbed model), and we shall observe, in particular, that some remarkable threshold behaviour occurs. While some aspects of the approach from [5] remain, in particular the focus on locally defined stable configurations which can be used to understand the global picture, the specific methods of their proof (the use of 'firewall incubators', and so on) apply only to the case $\tau = \frac{1}{2}$, and so largely speaking we shall require different techniques here.

We are very grateful to one of the anonymous referees for pointing out close connections to the work of Durrett and Steif [10]. Their analysis of linear threshold voter models has several points of similarity with the current study of Schelling models. At the same time, there are substantial differences between the two situations. For $\tau > 0.5$ (which

corresponds to $\tau < 0.5$ in the notation of [10]) their voter model and the Schelling model we work with have very different behaviours. For $\tau \leq 0.5$ the continuous time version of their voter model is closer to the Schelling model than the discrete time version of their model is (where all nodes are given the opportunity to change at each stage), and in fact the introduction of the discrete time model in [10] is motivated by the fact that they are unable to establish the main results of the paper for the continuous model.[1] The relationship between their results and those described here is as follows: Theorem 1.1 essentially follows from the work in [10], while Theorems 1.2 and 1.4 do not.

The picture which emerges is one in which one observes different behaviour in five regions. For κ which is the unique solution in $[0, 1]$ to:

$$\left(\frac{1}{2} - \kappa\right)^{1-2\kappa} = (1-\kappa)^{2-2\kappa},$$

($\kappa \approx 0.353092313$) these regions are: (i) $\tau < \kappa$, (ii) $\tau = \kappa$, (iii) $\kappa < \tau < \frac{1}{2}$, (iv) $\tau = \frac{1}{2}$, (v) $\tau > \frac{1}{2}$. In fact we shall not consider the case $\tau = \kappa$, but the behaviour for all other values of τ is given by the theorems below. Perhaps the most surprising fact is that, in some cases, increasing τ almost certainly leads to decreased segregation. The assumption is always that we work with $n \gg w$, i.e. all results hold for all n which are sufficiently large compared to w.[2] A *run* of length d is a set of d-many consecutive nodes all of the same type. *Complete segregation* refers to any configuration in which there exists a single run to which all α nodes belong.

The first theorem deals with low values of τ, and formally establishes the (perhaps rather intuitive) idea that very low levels of intolerance lead to low levels of segregation:

Theorem 1.1. *Suppose $\tau < \kappa$ and $\epsilon > 0$. For all sufficiently large w, if a node u is chosen uniformly at random, then the probability that any node in $\mathcal{N}(u)$ is ever involved in a swap is $< \epsilon$. Thus there exists a constant d such that, for sufficiently large w, the probability u belongs to a run of length $> d$ in the final configuration is $< \epsilon$.*[3]

Theorem 1.1 can easily be extrapolated from the arguments in [10].

As one increases τ beyond the threshold κ, however, the dynamics of the process qualitatively change:

Theorem 1.2. *Suppose $\tau \in \left(\kappa, \frac{1}{2}\right)$ and $\epsilon > 0$. There exists a constant d such that (for all w and all n such that $w \ll n$) the probability that u chosen uniformly at random will belong to a run of length $\geq e^{w/d}$ in the final configuration, is greater than $1 - \epsilon$.*

So, to summarise Theorem 1.2 less formally, increasing τ beyond κ suddenly causes high levels of segregation, in the form of run-lengths which are exponential in w. Furthermore, by analysing the proof of Theorem 1.2, we shall be able to prove (a formalised version of the statement) that increasing τ in this interval actually decreases run-lengths. So while Theorem 1.2 provides lower bounds on the run lengths, we shall later also be able to modify the proof of Theorem 1.2 in order to establish upper bounds on run lengths as well – this discussion takes places at the end of Section 4. The next case is discussed in [5], where one sees polynomially bounded run-lengths:

Theorem 1.3 ([5])**.** *Suppose $\tau = \frac{1}{2}$. There exists a constant $c < 1$ such that for all $\lambda > 0$, the probability that u chosen uniformly at random will belong to a run of length greater than λw^2 in the final configuration, is bounded above by c^λ.*

[1] In fact, their continuous time model is closer to the *simple* version of the Schelling model that we shall introduce shortly.

[2] Precisely what this means is that when a result is claimed to hold for a certain w, in fact there exists N_w such that the result holds for w whenever $n \geq N_w$. How much larger n must be than w is therefore not specified, and the claim is simply that the result holds whenever n is sufficiently large compared to w.

[3] A comment on conventions concerning quantification may be in order here. In the statement of Theorem 1.1 we fix τ and ϵ and then conclude that, for all sufficiently large w, a certain statement holds. It is to be understood here that how large one has to take w may depend upon τ and ϵ. Similarly, in Theorem 1.2, how large one has to take w, and then how large n must be in comparison to w, may depend upon τ and ϵ. The upshot of this is that, where we subsequently make multiple uses of the weak law of large numbers, there shall be no requirement of *uniform* convergence for the random variables in question.

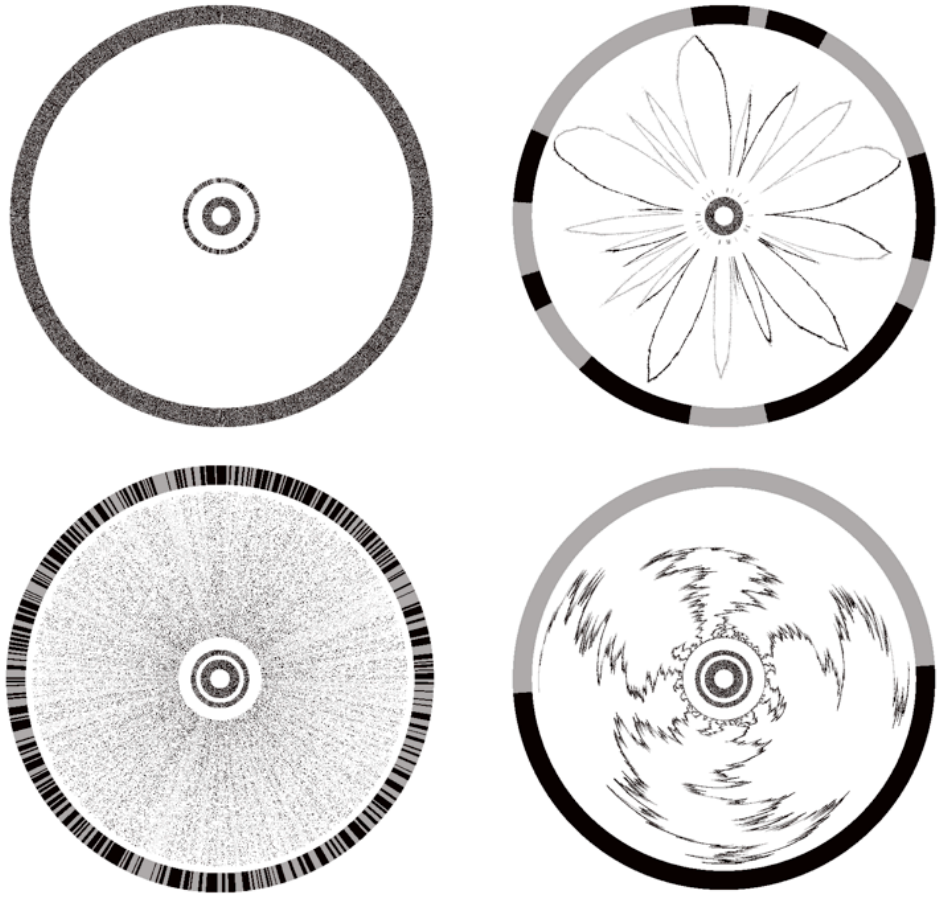

FIGURE 1. Threshold behaviour: top left $\tau = 0.3$; top right $\tau = 0.38$; bottom left $\tau = 0.5$; bottom right $\tau = 0.7$.

It is worth noting that Theorem 1.3 provides upper bounds but no lower bounds on run lengths. The final case is when $\tau > \frac{1}{2}$. Here a combinatorial argument can be used to argue that complete segregation will eventually occur. Note that, if $\frac{1}{2} < \tau \leq \frac{w+1}{2w+1}$, then the process is identical to that for $\tau = \frac{1}{2}$, since in both cases a node requires $w + 1$ many nodes of its own type in its neighbourhood in order to be happy.

Theorem 1.4. *Suppose that $\tau > \frac{1}{2}$, and that w is sufficiently large that $\tau > \frac{w+1}{2w+1}$. Then, with probability tending to 1 as $n \to \infty$, the initial configuration will inevitably lead to complete segregation.*[4]

We have constructed a program which efficiently simulates the process. The outcomes of some simulations are illustrated in Figure 1. In the processes depicted here the number of nodes $n = 100000$, $w = 60$ and in the diagrams individuals of type α are coloured light grey and individuals of type β are coloured black (much larger simulations are also illustrated in Section 5). The inner ring displays the initial mixed configuration (in fact the configuration is sufficiently mixed that changes of type are not really visible, so that the inner ring appears dark grey). The outer ring displays the final configuration. Just immediately exterior to the innermost ring are second and third inner rings, which display individuals which are unhappy in the initial configuration and individuals belonging to 'stable' intervals in the initial configuration respectively (stable intervals will be defined subsequently, and in fact, this third ring

[4]Note that, while Theorem 1.4 requires $\tau > \frac{w+1}{2w+1}$, Theorem 1.2 does not. Briefly, this is because the choice of d can be made to so as to deal with finitely many small w anyway, meaning that Theorem 1.2 is essentially a statement about what happens for large w.

is empty in these examples except for the case $\tau = 0.3$). The process by which the final configuration is reached is indicated in the space between the inner rings and the outer ring in the following way: when an individual changes type this is indicated with a mark, at a distance from the inner rings which is proportional to the time at which the change of type takes place. In fact, for the case $\tau > \frac{1}{2}$ one has to be a little careful in talking about the 'final' configuration – there will, almost certainly, always be unhappy individuals of both types able to swap, but once a completely segregated configuration is reached all future configurations must remain completely segregated.

We shall also be interested in a variant of the model, which we shall call the *simple model* (also see [1]), and which proceeds in the same way except that at each stage *one* unhappy node u is selected uniformly at random and changes type so long as this does not cause it to have less nodes of its own type within $\mathcal{N}(u)$. The process ends when no more legal changes are possible. One might justify interest in this version of the model in various ways. Firstly there are a number of situations in which it is much easier to work with (in one instance here, and also in [5], results are proved for the standard model by first considering the simple model and then arguing that the proof can be extended to the standard case). The simple variant of the model also makes sense as soon as one drops the assumption that we are working within a closed system. One might suppose that unhappy individuals living in a city will move to a location in the same city if one should be available, but will move elsewhere otherwise, and similarly that individuals will move into the city to fill locations becoming vacant. Lastly, this simple variant of the model is also closer to the various spin-1 models normally studied in statistical mechanics (although versions of the Ising model with 'Kawasaki' dynamics involving particles which swap location are also considered there). In all cases we shall prove the same results for both models, except for the case $\tau > \frac{1}{2}$, where the simple model eventually yields a society of nodes all of one type. The rough conjecture is that the simple model accurately describes local behaviour for the standard model.

In Section 2 we shall discuss terminology and we shall make some easy observations about how the evolution of the model can be understood. In Section 3 we prove Theorem 1.1. In Section 4 we prove Theorem 1.2. In Section 5 we discuss the case $\tau = \frac{1}{2}$, which was already dealt with in [5], and in Section 6 we prove Theorem 1.4. Sections 7, 8 and 9 deal with deferred proofs from Sections 3, 4 and 6 respectively.

Subsequent work. Subsequent to the first version of this paper, the authors have considered variants of the model in which the two types do not occur with equal probability in the initial configuration [2], in which the two types occur with equal probabilities in the initial configuration but have different parameters τ_α and τ_β [1], and also the latter scenario for the 2-dimensional case [3]. In both [1] and [3], however, a version of the model which is easier to work is used, in which unhappy nodes simply switch type rather than swapping with other unhappy nodes – the 'simple' model, as defined in Section 2. In [15], Immorlica, Kleinberg and Lucier also consider the 2-dimensional version of the same model, but deal with the case that both types have the same parameter τ. For $\tau = (1-\epsilon)/2$ with ϵ small they are able to establish similar results to Theorem 1.2, giving exponentially large regions in which all nodes are of the same type in the final configuration. These results have also been improved on recently by Omidvar and Franceschetti [20].

2. Some notation, terminology and some easy observations

In describing the model earlier, we talked in terms of nodes or individuals swapping locations at various stages of the dynamic process. To work in this way, however, requires one to draw a distinction between individuals and the locations they occupy at each stage, and to maintain a bijective map at each stage between these two sets. In fact, it is notationally easier to consider a process whereby one simply has a set of n nodes, with two unhappy nodes of opposite type selected at each stage (if such exist), which may then both change type (when this occurs we shall still refer to the nodes as 'swapping', but now they are swapping type rather than location). Thus nodes are identified with indices for their locations amongst the set $\{0, 1, ..., n-1\}$, and unless stated otherwise, addition and subtraction on these indices are performed modulo n. In the context of discussing a node u_1, for example, we might refer to the immediate neighbour on the right as node $u_1 + 1$. Since we work modulo n it is worth clarifying some details of the

interval notation: for $0 \leq b < a < n$, we let $[a, b]$ denote the set of nodes ('interval') $[a, n-1] \cup [0, b]$ (while $[b, a]$ is, of course, understood in the standard way).

As noted before, for any node u, we let $\mathcal{N}(u)$ denote the neighbourhood of u, which is the interval $[u-w, u+w]$. For any set of nodes I, suppose that x is the number of α nodes in I, while $y = |I| - x$. Then $\Theta(I) := x - y$ and is called the *bias* of I. By the bias of a node we mean the bias of its neighbourhood. Recall that by a *run* of length $m+1$ we mean an interval $[u, u+m]$ in which all nodes are of the same type.

We shall be particularly interested in local configurations which are stable, in the sense that certain nodes in them can never be caused to change type. Note that if an interval of length $w+1$ contains at least $\tau(2w+1)$ many α nodes, then each of those α nodes is happy so long as the others do not change type, meaning that, in fact, no α nodes in that interval will ever change type. We say that such an interval of length $w+1$ is α-stable (and similarly for β). An interval of length $w+1$ is *stable* if it is either α-stable or β-stable. We shall also make use of a particular kind of stable interval which was used in [5]: a *firewall* is a run of length at least $w+1$. We write 'for $0 \ll w \ll n$', to mean 'for all sufficiently large w and all n sufficiently large compared to w'.[5]

Arguing that the process ends. We define the *harmony index* corresponding to any given configuration to be the sum over all nodes of the number of their own type within their neighbourhood. For $\tau \leq \frac{1}{2}$, this harmony index is easily seen to strictly increase whenever an unhappy node changes type, which combined with the existence of an upper bound $n(2w+1)$, implies that the process must terminate after finitely many stages. For $\tau > \frac{1}{2}$, we shall argue that with probability tending to 1 as $n \to \infty$, the initial configuration is such that complete segregation eventually occurs with probability 1. Once complete segregation has occurred it is easy to see that all future states must be completely segregated, but that 'rotations' can occur, i.e. if the nodes of type α are precisely the interval $[a, b]$ at stage s, then at stage $s+1$ they must be either $[a-1, b-1]$, $[a, b]$ or $[a+1, b+1]$.

High level outline of the proofs of Theorems 1.1 and 1.2. The threshold κ which appears in Theorems 1.1 and 1.2 is pivotal because it is the point at which unhappy nodes become more likely than stable intervals. If $\tau < \kappa$ and we pick a node u uniformly at random, then for large w, we can expect stable intervals of both types on either side of u which are closer to u than any unhappy nodes. From there one can inductively argue that our randomly chosen u can never change type, and infer that most nodes will not change throughout the process. When $\tau > \kappa$, however, and for large w, one can expect many unhappy nodes on either side of u and which are nearer to u than any stable intervals. One can then show that each of these unhappy nodes has some reasonable chance of causing the creation of a firewall, as defined above. This happens by a straightforward cascading effect whereby an unhappy α node, for example, changes type to β, causing further α nodes to become unhappy. These nodes then change type, causing further nodes to become unhappy, and so on. The crucial observation is then this: firewalls will spread until they meet stable intervals of the opposite type. So if one of the unhappy nodes which is nearer to u than any stable interval causes a firewall, then this firewall (or another) will spread to consume u, meaning that u ultimately belongs to a run of long length. In order to make this argument work we also have to rule out various outcomes, such as the possibility that an unhappy node near u might fail to lead to the creation of a firewall, but might create a stable interval which blocks subsequent firewalls from consuming u.

3. The case $\tau < \kappa$

The analysis here is identical for the standard and simple models. Of course, we are yet to explain how κ comes to be defined as described previously – we shall do so shortly.

Outline. As described above, the basic idea is that we wish to find κ, which is that value of τ at which stable intervals become more likely than unhappy nodes in the initial configuration. For such τ, taking w large, we shall have that stable intervals are *much* more likely than unhappy nodes in the initial configuration. With a little bit more work one

[5] When we say that a result holds for $0 \ll w \ll n$, this means that there exists W such that for all $w \geq W$ there exists N_w such that the result holds for w when $n \geq N_w$.

can then show that, for a node u which is selected uniformly at random, we shall almost certainly find stable intervals of both types on either side of u before any unhappy element. The following lemma then suffices to establish that, given such an initial configuration, u can never change type.

Lemma 3.1. *Suppose that, in the initial configuration, u_1 and u_2 each belong to (possibly different) α-stable intervals, and that there are no unhappy α nodes in $[u_1, u_2]$. Then no α node in $[u_1, u_2]$ will ever become unhappy. A similar result holds for β.*

Proof. Suppose otherwise, and let v be the first α node in the interval $[u_1, u_2]$ to become unhappy. In order for v to become unhappy, another α node $v' \in \mathcal{N}(v)$ must change to type β. Since $v' \notin [u_1, u_2]$, we either have $v' < u_1 \leq v$, or else $v \leq u_2 < v'$. Suppose that the first case holds, the other is similar. Then, together with the fact that any α-stable interval to which u_1 belongs is of length $w + 1$, $v' \in \mathcal{N}(v)$ implies that v belongs to any stable interval to which u_1 belongs, and so cannot become unhappy. This gives the required contradiction. □

Finding κ. We are interested in the probability that a randomly chosen node belongs to a β-stable interval in the initial configuration. As a first approximation, however, we begin by asking the following slightly simpler version of that question: given some node u, chosen uniformly at random and of type β, say, what is the probability that $[u, u + w]$ is β-stable in the initial configuration? This can be modelled with a binomial distribution $X \sim b(w, \frac{1}{2})$, from which we are interested in the probability $P_{\text{stab}} = \mathbf{P}(X \geq (2w + 1)\tau - 1)$.

Similarly, an α node is unhappy if its neighbourhood contains more than $(2w+1)(1-\tau)$ many β nodes. We model this as $Y \sim b(2w, \frac{1}{2})$, from which we are interested in the probability $P_{\text{unhap}} = \mathbf{P}(Y > (2w + 1)(1 - \tau))$. We are interested in finding conditions on w and τ which ensure:

$$\frac{P_{\text{unhap}}}{P_{\text{stab}}} < 1 \quad \text{or} \quad \frac{P_{\text{unhap}}}{P_{\text{stab}}} < \varepsilon.$$

First, notice that if $\tau \leq \frac{1}{4}$, then $P_{\text{stab}} \geq \frac{1}{2}$ for all w, while $P_{\text{unhap}} \to 0$ as $w \to \infty$. Thus for all large enough w, we shall have $\frac{P_{\text{unhap}}}{P_{\text{stab}}} < \varepsilon$. Thus, as we look to find κ, we shall assume that $\frac{1}{4} < \tau < \frac{1}{2}$. Our key probabilistic fact will be the following result concerning the binomial distribution, which is presumably known, and which is proved in Section 7:

Lemma 3.2. *Suppose $h : \mathbb{N} \to \mathbb{N}$ and $p \in (0, 1)$ are such that there exists $k \in (0, 1)$ so that for all large enough N, we have $\left(1 + \left(\frac{1}{p} - 1\right)k\right)h(N) > N \geq h(N) > pN > 0$. Then for all large enough N, if $X_N \sim b(N, p)$, we have*

$$P\left(X_N = h(N)\right) \;\leq\; \boldsymbol{P}\left(X_N \geq h(N)\right) \;\leq\; \left(\frac{1}{1-k}\right) \cdot \boldsymbol{P}\left(X_N = h(N)\right).$$

That is to say in asymptotic notation, $\boldsymbol{P}\left(X_N \geq h(N)\right) = \Theta\left(\boldsymbol{P}\left(X_N = h(N)\right)\right)$.[6]

Applying Lemma 3.2 in our setting with $p = \frac{1}{2}$, $N = w$, and $1 > k > \frac{\frac{1}{2}-\tau}{\tau}$ gives us that $P_{\text{stab}} \approx \dfrac{1}{2^w}\dbinom{w}{h}$ where $h = \lceil (2w + 1)\tau \rceil - 1$ and where by "$\approx$" we mean the asymptotic notion Θ.

Similarly, taking $N = 2w$ and $1 > k' > \frac{\tau}{1-\tau}$ gives us $P_{\text{unhap}} \approx \dfrac{1}{2^{2w}}\dbinom{2w}{h'}$, where $h' = \lfloor (2w + 1)(1 - \tau)) \rfloor + 1$. Thus

$$\frac{P_{\text{unhap}}}{P_{\text{stab}}} \approx \frac{1}{2^w} \cdot \frac{\dbinom{2w}{h'}}{\dbinom{w}{h}}.$$

We now employ Stirling's formula, that $n! \approx n^{n+\frac{1}{2}}e^{-n}$. Then, the powers of e cancel and we see:

[6] Note that in this paper Θ is usually used to denote bias rather than being the standard asymptotic notation. It is only in discussing Lemma 3.2 that we shall use Θ to denote the standard asymptotic notation, so no confusion should result.

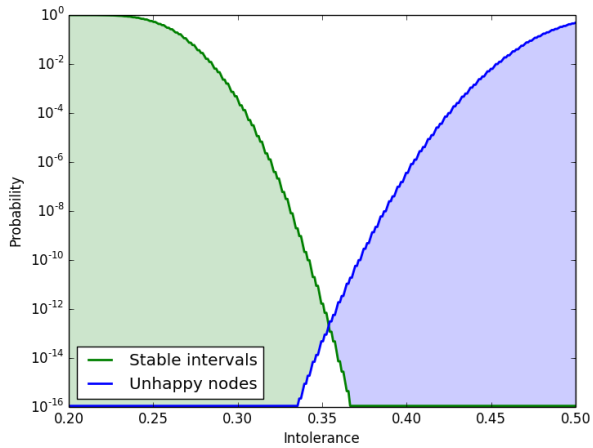

FIGURE 2. Probabilities (log scale) for $w = 300$.

$$\frac{P_{\text{unhap}}}{P_{\text{stab}}} \approx \frac{1}{2^w} \cdot \frac{(2w)^{2w+\frac{1}{2}}(h)^{h+\frac{1}{2}}(w-h)^{w-h+\frac{1}{2}}}{w^{w+\frac{1}{2}}(h')^{h'+\frac{1}{2}}(2w-h')^{2w-h'+\frac{1}{2}}}. \tag{3.0.1}$$

Now, approximating h by $2w\tau$ and approximating h' by $2w(1-\tau)$ (which we shall justify shortly) we get:

$$\frac{P_{\text{unhap}}}{P_{\text{stab}}} \approxeq (2w)^w \cdot \frac{(2w\tau)^{2w\tau+\frac{1}{2}}\,(w(1-2\tau))^{w(1-2\tau)+\frac{1}{2}}}{(2w(1-\tau))^{2w(1-\tau)+\frac{1}{2}}(2w\tau)^{2w\tau+\frac{1}{2}}}.$$

and so

$$\frac{P_{\text{unhap}}}{P_{\text{stab}}} \approxeq \left(\frac{\left(\frac{1}{2}-\tau\right)^{(1-2\tau)}}{(1-\tau)^{2(1-\tau)}}\right)^w. \tag{3.0.2}$$

Here, we write $f \approxeq g$ to mean that there are rational functions P and Q such that $P(w), Q(w) > 0$ and $P(w)g(w) \leq f(w) \leq Q(w)g(w)$.

Now put $f(x) = x^{2x}$. Then the value κ we are looking for is exactly τ such that

$$f\left(\frac{1}{2}-\tau\right) = f(1-\tau).$$

In other words $x = \frac{1}{2} - \kappa$ is exactly such that

$$f(x) = f\left(x + \frac{1}{2}\right).$$

Since f has a unique turning point at e^{-1}, it follows that κ is unique, and numerical analysis gives

$$\kappa \approx 0.353092313$$

(which is just slightly less than $\sqrt{2}/4$).

It remains to justify the approximations introduced for h and h' after 3.0.1. Well, $h = 2w\tau + \delta$ where $-1 < \delta < 1$. Thus $h^{h+\frac{1}{2}} = (2w\tau+\delta)^{\delta+\frac{1}{2}} \cdot (2w\tau)^{2w\tau} \cdot \left(1 + \frac{\delta}{2w\tau}\right)^{2w\tau}$. The final term lies within $[e^{-1}, e]$. Thus the necessary correction amounts

to multiplying or dividing $(2w\tau)^{2w\tau+\frac{1}{2}}$ by at most a polynomial term in w. Similar small polynomial corrections may be required for the other terms in 3.0.1, but the exponential nature of 3.0.2 guarantees that these correcting terms will play a negligible role for large w.

There were various other simplifications which were made in the analysis above, and which must now be attended to. In defining P_{stab}, we considered the probability that a β node belongs to a β-stable interval, rather than the probability that a node chosen uniformly at random belongs to a β-stable interval. We also considered a specific interval of length $w+1$ to which a node might belong and which might be stable, namely $[u, u+w]$, and so did not overtly take account of the fact that a given node belongs to $w+1$ many intervals of length $w+1$ which might be stable, namely $[u-\ell, u-\ell+w]$ for $0 \le \ell \le w$. Again, the exponential nature of (3.0.2) accommodates all such modifications, which only cause changes to P_{unhap} and P_{stab} which are are polynomially bounded in w. While (3.0.2) was derived under the assumption that $\frac{1}{4} < \tau < \frac{1}{2}$, in fact we can deduce that the following statement holds more generally:

($\dagger_0$) Consider the initial configuration. For u chosen uniformly at random, let P_{stab} be the probability that u belongs to an α-stable interval, and let P_{unhap} be the probability that u is an unhappy α node. If $\tau < \kappa$ and $k, r > 0$, then for all sufficiently large w, $P_{\text{stab}} > kw^r \cdot P_{\text{unhap}}$. When $\tau > \kappa$, $P_{\text{unhap}} > kw^r \cdot P_{\text{stab}}$. A similar result holds for β.

In order to see that ($\dagger_0$) holds outside the interval $\frac{1}{4} < \tau < \frac{1}{2}$, (for the non-trivial cases) one can simply compare the relevant probability ratios with those for τ just inside the interval.

Completing the argument. As hinted previously, we need something more than ($\dagger_0$) though. For any u, let x_u be the first node to the left[7] of u which, in the initial configuration, is either an unhappy α node, or else belongs to an α-stable interval. We must show that when $\tau < \kappa$ and w is sufficiently large, it is almost certainly the case that x_u belongs to an α-stable interval (and similarly for β and to the right of u). With this in place, the result then follows directly from Lemma 3.1. When combined with ($\dagger_0$), the following lemma (which is stated in such a way as to be general purpose, so we can apply it again later) therefore completes our proof, and is proved in Section 7.[8]

Lemma 3.3. *Let P_u and Q_u be events which only depend on the neighbourhood of u in the initial configuration, meaning that if the neighbourhood of v in the initial configuration is identical to that of u (i.e. for all $i \in [-w, w]$, $u+i$ is of the same type as $v+i$), then P_u holds iff P_v holds and Q_u holds iff Q_v holds. Suppose also that:*

(i) $\boldsymbol{P}(P_u) \neq 0$ and $\boldsymbol{P}(Q_u) \neq 0$.

(ii) For all k, for all sufficiently large w, $\boldsymbol{P}(P_u)/\boldsymbol{P}(Q_u) > kw$.

For any u, let x_u be the first node to the left of u such that either P_{x_u} or Q_{x_u} holds. For any $\epsilon > 0$, if $0 \ll w \ll n$ then the following occurs with probability $> 1-\epsilon$ for u chosen uniformly at random: x_u is defined and for no node v in $[x_u - 2w, x_u]$ does Q_v hold.

An analogous result holds when 'left' is replaced by 'right'.

4. The case $\kappa < \tau < \frac{1}{2}$

We work first with the simple model, and then supply the necessary modifications for the standard case.

In what follows we shall work with some fixed τ in the interval $\left(\kappa, \frac{1}{2}\right)$, some fixed $\epsilon > 0$, and we shall assume that n is large compared to w. We want to show that there exists a constant d such that for all sufficiently large w the probability that a randomly chosen node will belong to a run of length $\ge e^{w/d}$ (in the final configuration) is greater

[7] By the first node to the left of u satisfying a certain condition we mean the first in the sequence $u, u-1, u-2, \cdots$ which satisfies the condition.

[8] For the reader who is concerned that we need to make four applications of Lemma 3.3 to various different events P_u and Q_u which are not independent (with P_u either being the event that u belongs to an α-stable interval or that u belongs to a β-stable interval, and with Q_u either being the event that u is an unhappy α node or the event that it is an unhappy β-node), note that the probability that any one of four unlikely events occurs is at most the sum of their individual probabilities.

than $1-\epsilon$. Of course, proving the result for all sufficiently large w suffices to give the result for all w since one can simply adjust the choice of d to deal with finitely many small values, but we shall make frequent use of the fact that we need only work for all sufficiently large w in what follows and so stating the theorem in this way is instructive.

Our entire analysis takes place relative to a node u_0, chosen uniformly at random. As described previously, the basic aim is to establish that in the final configuration u_0 very probably belongs to a firewall of considerable length. The argument consists essentially of two main parts: first we consider what can be expected from the vicinity[9] of u_0 in the initial configuration, and then we consider how events are likely to develop in subsequent stages.

Before we begin with the technicalities, let us consider very informally what can be expected from the vicinity of u_0 in the initial configuration. Since $\kappa < \tau < \frac{1}{2}$, for large w we shall have that unhappy nodes are much more likely than stable intervals, but that unhappy nodes themselves are few and far between. Starting at u_0 and moving to the left, (since n is large) we can expect to find a first unhappy node, l_1 say, and it will very likely be the case that $[l_1, u_0]$ is an interval of considerable length, containing no stable intervals. To the left of l_1 and inside $\mathcal{N}(l_1)$, there may be some other unhappy nodes. If we move now to $l_1-(2w+1)$, however, and repeat the process (with $l_1-(2w+1)$ taking the place of u_0), then so long as w is large enough, we can expect the same to happen again, i.e. we find a first unhappy node l_2 and $[l_2, l_1-(2w+1)]$ is an interval of considerable length, containing no stable intervals. In fact, for any fixed k which does not depend on w, if we repeat this process k many times, then so long as w is large enough (and how large we have to take w will depend on k) we can be pretty sure that the same thing will happen at every one of those k-many steps. Now let us establish this informal picture more carefully.

The initial configuration. Recall that for any set of nodes I, $\Theta(I)$ is the *bias* of I and that by the bias of a node we mean the bias of its neighbourhood. In general, if $x_1, ..., x_k$ are independent random variables with $\mathbf{P}(x_i = 1) = \mathbf{P}(x_i = -1) = \frac{1}{2}$ when $1 \le i \le k$, then letting $X = \sum_{i=1}^{k} x_i$ Hoeffding's inequality gives, for arbitrary $\lambda > 0$:

$$\mathbf{P}(|X| > \lambda\sqrt{k}) < 2e^{-\lambda^2/2}.$$

Now we use this to bound the probability that a node u has bias in the initial configuration which will cause it to be unhappy, should u be of the minority type in its neighbourhood. So, we wish to bound the probability that the number of α nodes in $\mathcal{N}(u)$ is $> (1-\tau)(2w+1)$ or the number of β nodes in $\mathcal{N}(u)$ is $> (1-\tau)(2w+1)$. This corresponds to a bias $\Theta(\mathcal{N}(u))$ of $> (1-2\tau)(2w+1)$ or $< -(1-2\tau)(2w+1)$.

Definition 4.1. *When $|\Theta(\mathcal{N}(u))| > (1-2\tau)(2w+1)$ we say that u has **high bias**, denoted $\mathtt{Hb}(u)$. If this holds for the initial configuration, we say that $\mathtt{Hb}^*(u)$ holds.*

Definition 4.2. *For the remainder of this section we define $d = \frac{2}{(1-2\tau)^2}$.*

Lemma 4.3 (Likely happiness)**.** *Let u be a node chosen uniformly at random. For any $\epsilon' > 0$ and for all sufficiently large w, the probability that $\mathtt{Hb}^*(u)$ holds is $< \epsilon' e^{-w/d}$.*

Proof. Putting $\lambda\sqrt{2w+1} = (1-2\tau)(2w+1)$ in Hoeffding's inequality above, we get

$$\lambda^2/2 = (1-2\tau)^2\frac{(2w+1)}{2}, \quad \text{so} \quad e^{-\lambda^2/2} = e^{-w/d^*} \quad \text{where} \quad d^* = \frac{2w}{2w+1}\cdot\frac{1}{(1-2\tau)^2}.$$

We chose $d > \frac{1}{(1-2\tau)^2}$, which means that for any $\epsilon' > 0$ and for all sufficiently large w, the probability that u has high bias in the initial configuration is $< \epsilon' e^{-w/d}$, as required. □

[9] To be clear, the term 'vicinity' of u_0 is used informally here, to mean an interval containing u_0, which may be large compared to w but which is small compared to n.

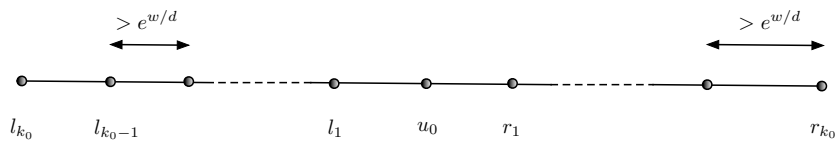

FIGURE 3. Picking out nodes of high bias in the vicinity of u_0.

Defining the nodes l_i and r_i. For now, we fix some $k_0 > 0$. We shall choose a specific value of k_0 which is appropriate later – for now, however, we make the promise that our choice of k_0 will not depend on w.

For $1 \le i \le k_0$ we define a node l_i to the left of u_0 and also a node r_i to the right. We let l_1 be the first node v to the left of u_0 such that $\mathtt{Hb}^*(v)$ holds, so long as this node is in the interval $[u_0 - \frac{1}{4}n, u_0]$ (otherwise l_1 is undefined). Then, given l_i for $i < k_0$ we let l_{i+1} be the first node v to the left of $l_i - (2w+1)$ such that $\mathtt{Hb}^*(v)$ holds, so long as no nodes in the interval $[l_{i+1}, l_i]$ are outside the interval $[u_0 - \frac{1}{4}n, u_0]$ (otherwise l_{i+1} is undefined). We let r_1 be the first node v to the right of u_0 such that $\mathtt{Hb}^*(v)$ holds, so long as this node is in the interval $[u_0, u_0 + \frac{1}{4}n]$. Given r_i for $i < k_0$ we let r_{i+1} be the first node v to the right of $r_i + (2w+1)$ such that $\mathtt{Hb}^*(v)$ holds, so long as no nodes in the interval $[r_i, r_{i+1}]$ are outside the interval $[u_0, u_0 + \frac{1}{4}n]$.

The reason for considering the intervals $[u_0 - \frac{1}{4}n, u_0]$ and $[u_0, u_0 + \frac{1}{4}n]$ in the above, is that we wish to be able to move left from u_0 to l_{k_0} without meeting any of the nodes r_i.

Definition 4.4 (Good spacing). *Let d be as in Definition 4.2. It is notationally convenient to let $l_0 = u_0 = r_0$. We let **Good spacing** be the event that for $1 \le i \le k_0$:*

(i) l_i and r_i are both defined, and;

(ii) $|l_i - l_{i-1}| > e^{w/d}$ and $|r_i - r_{i-1}| > e^{w/d}$.

Note that for any fixed w, as $n \to \infty$ the probability that any l_i or r_i is undefined (for $i \le k_0$) goes to 0. By Lemma 4.3, and since the probability that any node in an interval I has high bias in the initial configuration is at most $\Sigma_{u \in I}\mathbf{P}(\mathtt{Hb}^*(u))$, for any $\epsilon' > 0$ and for any fixed $k_0 \ge 1$ we can ensure that $\mathbf{P}(\textbf{Good spacing}) > 1 - \epsilon'$ by taking w sufficiently large (and by taking n sufficiently large compared to w). Thus, for $0 \ll w \ll n$, the picture we are presented with is almost certainly as in Figure 3.

Building the informal picture. Recall that, by a *run* of length $m+1$ we mean an interval $[u, u+m]$ in which all nodes are of the same type, and that a *firewall* is a run of length at least $w+1$. The basic observation on which we now wish to build is as follows: if the interval $[u-w, u]$ is a firewall of type α, then when $u+1$ is of type β, it cannot be happy unless the interval $[u+1, u+w+1]$ is β-stable. So[10], since we are dealing with $\tau \le \frac{1}{2}$:

Firewalls will spread until they hit stable intervals of the opposite type.

With this in mind, let us now consider informally what can be expected to happen in the neighbourhood of l_i. Suppose that l_i is initially of type β and is unhappy in the initial configuration. Then with probability close to 1 for sufficiently large w, there will not be any unhappy nodes of type α in the neighbourhood of l_i in the initial configuration. If l_i changes type, then this will make the bias in its neighbourhood still more positive, which may cause further nodes of type β to become unhappy. If these change to type α then this will further increase the bias, potentially causing more nodes to become unhappy, and so on. The following definitions formalise some of the ways in which this

[10]When dealing with the standard rather than the simple model, this is only true so long as there exist unhappy nodes of both types. Later, our analysis using Wormald's machinery will address this issue.

process might play out, and in particular the possibility that this process might play out without interference from what happens in other neighbourhoods $\mathcal{N}(l_j)$ or $\mathcal{N}(r_j)$.

Definition 4.5. *For $0 < i < k_0$ we say that l_i **completes at stage** s if both:*

(1) No node in $\mathcal{N}(l_i)$ is unhappy at stage s, and this is not true for any $s' < s$.

(2) There exist x_0 and x_1 with $l_{i+1} + 2w < x_0 < l_i - 2w < l_i + 2w < x_1 < l_{i-1} - 2w$, such that by the end of stage s, no node in $[x_0 - w, x_0]$ or $[x_1, x_1 + w]$ has changed type.

*We say that l_i **completes** if it completes at some stage. We also define completion for r_i analogously.*

Definition 4.6. *We say that l_i (or r_i) **originates a firewall** if it completes at some stage s and (i) it belongs to firewall at stage s, and (ii) all type changes in $\mathcal{N}(l_i)$ (or $\mathcal{N}(r_i)$) at stages $\leq s$ are of the same kind (i.e. all α to β, or all β to α).*[11]

The informal idea, is that we now wish to show that each l_i and each r_i has some reasonable chance of originating a firewall (and that this reasonable chance is bounded below by some value which doesn't depend on w). Then we can choose k_0 so that the probability none of the l_i originate a firewall or none of the r_i originate a firewall is $\ll \epsilon$, i.e. with probability close to 1 firewalls will originate either side of u_0 within the interval $[l_{k_0}, r_{k_0}]$. Then, letting i_1 be the least i such that l_i originates a firewall, and letting i_2 be the least i such that r_i originates a firewall, we wish to show that with probability close to 1 the firewalls originated at l_{i_1} and r_{i_2} will spread until u_0 is contained in one of them. Since these two firewalls have originated at nodes which are at distance at least $e^{w/d}$ apart, u_0 ultimately belongs to a firewall of at least this length. So to sum up:

> The approximate reason Theorem 1.2 holds is that u_0 can be expected to join a firewall which – *precisely because* unhappy nodes are rare in the initial configuration – originated at a long distance from u_0.

In order to make this basic picture work, however, we need to be careful about the formation of stable intervals in $[l_{k_0}, r_{k_0}]$. As noted above, firewalls will spread until they hit stable intervals of the opposite type. Now suppose that, with i_1 and i_2 as above, $i_1 = i_2 = 2$ and, for now, suppose that α-firewalls are originated at both l_2 and r_2. In order to show that these two firewalls will spread until they meet each other, it will be helpful first of all, to be able to assume that in the initial configuration there are no stable subintervals of $[l_{k_0}, r_{k_0}]$. This will follow quite easily for large w, from our previous analysis of the ratio between the probability of unhappy nodes and stable intervals. A further danger that we have to be able to avoid, however, is that, while l_1 and r_1 do not originate firewalls, they *do* get as far as creating β-stable intervals.

Definition 4.7. *Given i with $0 < i < k_0$, let $u = l_i$ or $u = r_i$. Let $u_1 = u - (2w + 1)$ and $u_2 = u + (2w + 1)$. We say that u **subsides** if it completes at some stage s, and:*

- *There are no nodes in $[u_1, u_2]$ belonging to stable intervals at stage s;*
- *No nodes in $\mathcal{N}(u_1)$ or $\mathcal{N}(u_2)$ have been unhappy at any stage $\leq s$.*

So we need to be able to show, in fact, that with probability close to 1 each l_i and r_i either originates a firewall or subsides. To do this clearly involves a careful analysis what is likely to happen in each of the neighbourhoods $\mathcal{N}(l_i)$ and $\mathcal{N}(r_i)$. First of all, the large distances between these nodes mean that, for fixed k_0 and sufficiently large w, we can expect all of the l_i and r_i (for $0 < i < k_0$) to complete, so that one can understand the early stages of the process for each of these neighbourhoods by considering each in isolation. We then wish to show:

> The required dichotomy: in the neighbourhood of each l_i and r_i, either a small number of type changes will occur before completion, or else a large number of type changes will occur before completion and a firewall will be created.

[11] This definition might initially seem to neglect the possibility, for example, that l_i completes at some stage and does not belong to a firewall at that stage, but that, nevertheless, the sequence of type changes in its neighbourhood and surrounding neighbourhoods which have led to completion have caused the creation of a firewall. In fact we shall be able to ensure (Lemma 4.21) that with probability close to 1 (for large w) such events do not occur.

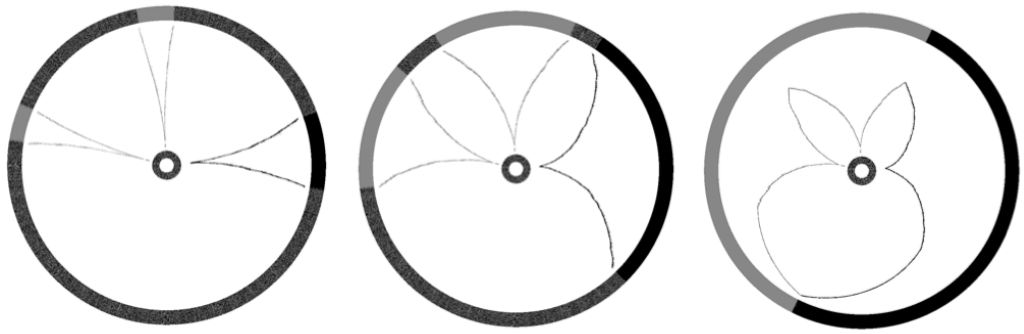

FIGURE 4. An example of the developing process with $n = 100000, w = 75, \tau = 0.37$, at stages $s = 5000, 15000$ and 24986.

Now, if we strengthen our original requirement that there are no stable subintervals of $[l_{k_0}, r_{k_0}]$ in the initial configuration, to a requirement that there are no subintervals which are 'close' to being stable in the initial configuration (where 'close' is to be made precise in such a way as to ensure that when a small number of type changes occur in the neighbourhood of l_i before completion, these are not enough to create any stable intervals), then we shall have that with probability close to 1 each l_i and each r_i either originates a firewall or else completes without creating stable intervals.

Once all this is in place, there is then one further hurdle. In the above, we assumed that the firewalls originating at l_2 and r_2 are both α-firewalls. If they are firewalls of opposite type, however, we still have some work to do in order to prove that u_0 will almost certainly end up belonging to one of these two firewalls.

Formalising the intuitive picture. Thus far we have defined an event, **Good spacing**, which depends upon the value k_0. We are yet to specify k_0, but have promised that this choice will not depend on w. For any $\epsilon' > 0$ and for any fixed $k_0 \geq 1$ we can ensure that $\mathbf{P}(\textbf{Good spacing}) > 1 - \epsilon'$ by taking w sufficiently large (and by taking n sufficiently large compared to w).

In order to specify how the type changing process can be expected to develop in the vicinity of u_0, we shall now proceed to define a finite number of other events of this kind. This finite set of events (having **Good spacing** as a member) we shall call Π. Informally speaking, each of the events in Π will be something which can be expected to occur for large w, and which will turn out to be useful in describing how the type changing process is likely to unfold. So while **Good spacing** is one element of Π, another might be an event asserting conditions on the neighbourhoods of each l_i and r_i in the initial configuration, for example. Our aim will be to show that, for any $\epsilon' > 0$, the probability that all the events in Π occur is greater than $1 - \epsilon'$ for all sufficiently large w (and for n sufficiently large compared to w). Suppose that this is established for some $\Pi' \subset \Pi$, and that we want next to establish the result for $\Pi' \cup \{Q\}$ with $Q \in \Pi - \Pi'$. In showing that Q holds with probability close to 1, it might be convenient to assume that *all* of the events in Π' hold, or it might be more convenient to assume only that some specific subset of Π' holds, and we can make whichever choice is more convenient. More formally, to establish the result for $\Pi' \cup \{Q\}$ with $Q \in \Pi - \Pi'$, it then suffices to prove for each $\epsilon' > 0$ and all sufficiently large w, that the probability of Q given P is greater than $1 - \epsilon'$, where P is any conjunction (possibly empty) of the events in Π'. The point is that we can choose P which is most convenient to work with.

In discussing the 'required dichotomy' above, a requirement was suggested, that there should be no subintervals of $[l_{k_0}, r_{k_0}]$ which are 'close' to being stable in the initial configuration. In fact an appropriate formalisation of this idea is easy to describe, and we now do so.

Definition 4.8. *For any $\tau' \in (0,1)$, we say that an interval of length $w+1$ is τ'-**stable**, if it contains $\tau'(2w+1)$ many α-nodes or $\tau'(2w+1)$ many β nodes.*

Definition 4.9 (**Stable clear**)**.** *Once and for all, fix some τ_0 with $\kappa < \tau_0 < \tau$. Let **Stable clear** $\in \Pi$ be the event that l_i and r_i are defined for all $1 \le i \le k_0$, and there do not exist any τ_0-stable subintervals of $[l_{k_0}, r_{k_0}]$ in the initial configuration (note that l_i and r_i are defined in terms of τ rather than τ_0) .*

The following lemma has a simple proof, given in Section 8:

Lemma 4.10. *For fixed k_0 and $\epsilon' > 0$, **P**(**Stable clear**) $> 1 - \epsilon'$ for all w sufficiently large (and all n sufficiently large compared to w).*

Now, in order to establish the required dichotomy, we need to build up a clear picture of what the neighbourhoods $\mathcal{N}(l_i)$ and $\mathcal{N}(r_i)$ can be expected to look like. The distributions for these intervals in the initial configuration are a little difficult to attack directly, however, due to the nature of their definition. In choosing l_1 we move left until we find the *first* node which has high bias – this gives an asymmetry to the given information concerning the neighbourhood. Roughly, we might expect something like a hypergeometric distribution, but how good is this as an approximation? What we shall do, in fact, is first of all to understand what can be expected from the neighbourhood of a node which is chosen uniformly at random from among those with borderline bias:

Definition 4.11. *Let us say that a node u has **borderline bias**, denoted $\mathsf{Bb}(u)$, if:*

$$|\Theta(\mathcal{N}(u))| = Min\{\theta \in 2\mathbb{N}+1 : \theta > (1-2\tau)(2w+1)\},$$

i.e., u has high bias but decreasing the modulus of the bias by the minimum possible amount of 2 would cause it not to have high bias. We say that $\mathsf{Bb}^*(u)$ holds if u has borderline bias in the initial configuration.

The following is worth emphasising:

($\dagger_1$) Assuming **Good spacing** each of the nodes l_i and r_i ($0 < i \le k_0$) has borderline bias.

In what follows it is often convenient to work with some fixed $k \ge 1$ and to divide an interval $I = [a,b]$ into k parts of equal length. This occasions the minor inconvenience that the length of the interval might not be a multiple of k, motivating the following definition:

Definition 4.12. *Let $I = [a,b]$ and suppose $k \ge 1$. We define the subintervals $I(1:k) := \left[a, a + \left\lfloor \frac{b-a}{k} \right\rfloor\right] := [I(1:k)_1, I(1:k)_2]$ and*

$$I(j:k) := \left[a + \left\lfloor \frac{(j-1)(b-a)}{k} \right\rfloor + 1, a + \left\lfloor \frac{j(b-a)}{k} \right\rfloor\right] := [I(j:k)_1, I(j:k)_2]$$

for $2 \le j \le k$.

In Definition 4.12 the intervals are counted from left to right, but it is also useful sometimes to work from right to left:

Definition 4.13. *Let $I = [a,b]$ and suppose $k \ge 1$. For $1 \le j \le k$ we define $I(j:k)^- = I(k-j+1:k)$, $I(j:k)^-_1 = I(k-j+1:k)_1$ and $I(j:k)^-_2 = I(k-j+1:k)_2$.*

Lemma 4.14 (Smoothness Lemma)**.** *Suppose u is such that the proportion of α nodes in $I := \mathcal{N}(u)$ is θ, and that u is selected uniformly at random from nodes with this property. Then for any fixed $k \ge 1$ and $\epsilon' > 0$, for all sufficiently large w the following holds with probability $> 1 - \epsilon'$: for every j with $1 \le j \le k$, the proportion of the nodes in $I(j:k)$ which are α nodes, lies in the interval $[\theta - \epsilon', \theta + \epsilon']$.*

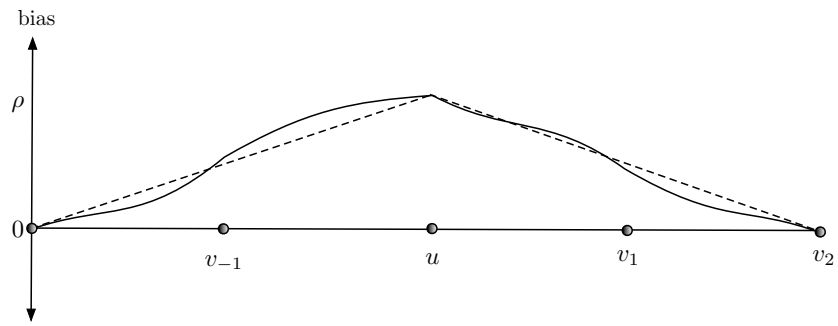

FIGURE 5. Smooth bias change for $k = 2$.

Proof. Once we are given that θ is the proportion of the nodes in I which are of type α, the nodes in this interval cease to be independently distributed, and the distribution becomes hypergeometric. Since we consider fixed k and ϵ' and take w large, it suffices to prove the result for given j with $1 \leq j \leq k$, i.e. if $P_1, ..., P_k$ are events, each of which occurs with probability tending to 1 as $w \to \infty$, then their conjunction also occurs with probability tending to 1 as $w \to \infty$. For a given j, the result follows directly, however, from an application of Chebyshev's inequality and standard results for the mean and variance of a hypergeometric distribution. Let x be the number of α nodes in the interval $I(j : k)$ and let ℓ be the length of the interval, so that $|\ell - (2w + 1)/k| \leq 1$. Then we have:

$$\mathbf{P}(|x/\ell - \theta| > \epsilon') < \ell^{-2}\epsilon'^{-2} \operatorname{Var}(x) = O(1)\ell^{-1}.$$

□

Lemma 4.14 basically tells us that if we choose a node u with borderline bias uniformly at random, then for large w we can expect the bias to move towards 0 fairly smoothly as we move to $u + (2w + 1)$ or $u - (2w + 1)$. In order to see roughly why this is true, suppose that $|\Theta(\mathcal{N}(u))| = \rho$ and let θ be the proportion of the nodes in $\mathcal{N}(u)$ which are of type α. Let $I = [u, u + (2w + 1)]$ and, for some k, consider the sequence of evenly spaced nodes $v_j = I(j, k)_2$. Now in forming the neighbourhood of v_j, we lose an interval of length (almost exactly) $\lfloor j(2w+1)/k \rfloor$ from the neighbourhood of u, which by Lemma 4.14 we can expect to have a proportion of α nodes very close to θ. We also gain an interval of the same length from outside $\mathcal{N}(u)$, which we can expect to have a proportion of α nodes very close to $\frac{1}{2}$. This means a bias for v_j close to $\rho\frac{k-j}{k}$.

The following definition allows us to express this more formally:

Definition 4.15. *Suppose that* $\mathtt{Hb}^*(u)$ *holds. Let* $I_1 = [u - (2w + 1), u]$ *and* $I_2 = [u, u + (2w + 1)]$. *Let* $|\Theta(\mathcal{N}(u))| = \rho$ *and let* θ *be the proportion of the nodes in* $\mathcal{N}(u)$ *which are of type* α. *Suppose that* $k \geq 1$ *is even and* $\epsilon' > 0$. *For* $1 \leq j \leq k$ *let* $v_j = I_2(j : k)_2$ *and let* $v_{-j} = I_1(j : k)_1^-$. *We say that* $\mathtt{Smooth}_{k,\epsilon'}(u)$ *holds if both:*

- *For every* j *with* $1 \leq |j| \leq k$, $|\Theta(\mathcal{N}(v_j)) - \rho\frac{k-j}{k}|/w < \epsilon'$.
- *For every* j *with* $1 \leq j \leq k/2$ *the proportion of the nodes in* $I_1(j : k)^-$ *which are of type* α *lies in the interval* $[\theta - \epsilon', \theta + \epsilon']$, *and similarly for* $I_2(j, k)$.

We say that $\mathtt{Smooth}^*_{k,\epsilon'}(u)$ *holds if* $\mathtt{Smooth}_{k,\epsilon'}(u)$ *holds in the initial configuration. Figure 5 illustrates the picture for* $k = 2$.

Corollary 4.16 (Smoothness Corollary)**.** *Suppose* u *is selected uniformly at random among those nodes such that* $\mathtt{Bb}^*(u)$ *holds. For all* $k \geq 1$ *and* $\epsilon' > 0$, *and for all sufficiently large* w, $\mathtt{Smooth}^*_{k,\epsilon'}(u)$ *holds with probability* $> 1 - \epsilon'$.

Proof. Let $u_{-1} = u - (2w + 1)$ and $u_1 = u + (2w + 1)$. The fact that u has borderline bias has no impact on the distributions for $\mathcal{N}(u_{-1})$ and $\mathcal{N}(u_1)$. Let $I_3 = \mathcal{N}(u_{-1})$ and $I_4 = \mathcal{N}(u_1)$. By the weak law of large numbers, for large w

we can expect the proportion of α nodes in each $I_3(j:k)$ and $I_4(j:k)$ to be close to $\frac{1}{2}$. The result then follows from Lemma 4.14. □

While Smoothness Corollary 4.16 tells us what can be expected from the neighbourhood of a node chosen uniformly at random from those with borderline bias, this does not immediately allow us to infer anything about what can be expected from the neighbourhood of each l_i and r_i. What we need is that, if we choose a node u uniformly at random and then move left (or right) until we find a first node v with high bias, then with probability close to 1 $\texttt{Smooth}^*_{k,\epsilon'}(v)$ holds. Ideally, we would like to be able to apply Lemma 3.3, but that would involve establishing an appropriate form of condition (ii) from the statement of the lemma. In order to work around this condition, one has to take a slightly circuitous route via several intermediate notions, but the following lemma can be established (the proof is given in Section 8).

Lemma 4.17 (Smoothness for l_i and r_i)**.** *For any node u, let x_u be the first node to the left of u which has high bias in the initial configuration. For any $\epsilon' > 0$ and $k \geq 1$, if $0 \ll w \ll n$ and u is chosen uniformly at random, then x_u is defined and $\texttt{Smooth}^*_{k,\epsilon'}(x_u)$ holds with probability $> 1 - \epsilon'$.*

An analogous result holds when 'left' is replaced by 'right'.

In order to see that Lemma 4.17 suffices to establish probable smoothness for all of the l_i and r_i, note first that k_0 is fixed while we take w large. At step i of the iteration which defines the sequence $l_1, l_2, ..$, the fact that l_i has borderline bias tells us nothing about the neighbourhood of $l_i - (2w+1)$ or the nodes to the left of this neighbourhood (but at a distance small compared to n).

We are now ready to define the third event in Π:

Definition 4.18 (**Smooth**)**.** *Let τ_0 be as in Definition 4.9. Once and for all, choose k_1 such that $\frac{1}{k_1} \ll \tau - \tau_0$, and choose k_2 and ϵ_0 such that $k_1 \ll k_2 \ll \frac{1}{\epsilon_0}$ and k_2 is a multiple of k_1. We define **Smooth** to be the event that all the l_i and r_i are defined for $1 \leq i \leq k_0$, and that when $u = l_i$ or $u = r_i$, $\texttt{Smooth}^*_{k_2,\epsilon_0}(u)$ holds.*

By Lemma 4.17, when $k_0 \ll w$ the probability that **Smooth** does not occur is $\ll \epsilon$.

The process to completion. Having established a clearer picture of what can be expected from the initial configuration, we now look to understand what will happen in the early stages, in the neighbourhood of each l_i or r_i. First of all, we must establish that these nodes can be expected to complete. The proof of Lemma 4.19 is given in Section 8, but the reader can probably see immediately how the proof might go. A happy node cannot become unhappy without changes to their neighbourhood – meaning that unhappiness has to spread one neighbourhood at a time. Given the large distances between the l_i and r_i, failure of completion for some l_i would involve a large number of events causing the spread of unhappiness from l_{i-1} or l_{i+1}, all occurring during stages when there are still unhappy nodes in $\mathcal{N}(l_i)$. At each stage the likelihood of such an event is not that much more than an unhappy node in $\mathcal{N}(l_i)$ being chosen to swap.

Lemma 4.19 (l_i and r_i complete)**.** *For any $\epsilon' > 0$, if $0 \ll w \ll n$ then for all $i \in [1, k_0)$, l_i and r_i will (be defined and will) complete with probability $> 1 - \epsilon'$.*

Definition 4.20 (**Completion**)**.** *We define **Completion** ($\in \Pi$) to be the event that all the l_i and r_i (for $1 \leq i < k_0$) are defined and complete.*

We are now ready to prove the required dichotomy.

Lemma 4.21 (The required dichotomy)**.** *Suppose that w is large and that **Good spacing**, **Stable clear**, **Smooth** and **Completion** all hold. Then for $i < k_0$, l_i and r_i will each either subside or originate a firewall.*

Proof. We prove the result for l_i, and the proof for r_i is essentially identical.

Note first that the choice of k_1 in Definition 4.18 means, in particular, that $10w/k_1$ type changes in any given neighbourhood cannot create stable intervals, given that **Stable clear** holds (the numbers here are fairly arbitrary). Note

also, that satisfaction of **Smooth** suffices to ensure it is not the case that there are unhappy nodes of both types in the neighbourhood of l_i in the initial configuration. Now suppose that l_i completes at stage s and has positive bias in the initial configuration (the case for negative bias is essentially identical).

It is useful at this point to establish names for a number of relevant intervals. We let $u_1 = l_i - (2w+1)$ and $u_2 = l_i + (2w+1)$. Then we define:

- $J = \mathcal{N}(u_1) \cup \mathcal{N}(l_i) \cup \mathcal{N}(u_2)$.
- $I = [u_1, u_2]$.
- $I_1 = [u_1, l_i]$, $I_2 = [l_i, u_2]$.
- $K_j^1 = I_1(j : k_1)^- \cup I_2(j : k_1)$.
- $K_j^2 = I_1(j : k_2)^- \cup I_2(j : k_2)$.

In Definition 4.18 we assumed that k_2 is a multiple of k_1, so we may let m be such that $k_2 = mk_1$. It is convenient to assume that k_2 is even. Now we divide into two cases.

l_i **subsides**. First of all, suppose that at stage s, there is a β-node u in the interval K_1^1. Then u must be happy at stage s. The fact that **Smooth** is satisfied, together with the fact that l_i completes at stage s, means that prior to stage s, the only type changes in the interval J are from type β to type α, so that u must be happy at *every* stage $\leq s$. Now, since $k_1 \ll k_2 \ll \frac{1}{\epsilon_0}$ and $\texttt{Smooth}^*_{k_2,\epsilon_0}(l_i)$ holds, any nodes in $I - (K_1^1 \cup K_2^1)$ have lower bias than u, and hence are happy, in the initial configuration. It then follows by induction on the stages $\leq s$ that no node in $J - (K_1^1 \cup K_2^1)$ changes type prior to stage s. In order to see this suppose that it holds prior to stage $s' \leq s$. Then at stage s', if $v \in I - (K_1^1 \cup K_2^1)$, it still has lower bias than u and so cannot change type from β to α, since u is happy so v must be. If $v \in J - I$, then v being the first in $J - I$ to change type would require it to be made unhappy by a type change outside J, contradicting the fact that l_i completes and condition (2) from the definition of completion.

We therefore get at most $|(K_1^1 \cup K_2^1)| < 10w/k_1$ many type changes in the interval J prior to completion. As observed above, this means that no stable intervals are created and l_i subsides, as required.

l_i **originates a firewall**. So suppose instead that, at stage s, all nodes in the interval K_1^1 are of type α. Given m as above, another way of putting this, is that all nodes in $\bigcup_{j \leq m} K_j^2$ are of type α at stage s. We now show by induction on $r \geq m$ that, when $r \leq k_2/2$, any nodes in K_r^2 must be of type α at stage s – i.e. that all nodes in $\mathcal{N}(l_i)$ are α nodes at stage s. So suppose that $m \leq r < k_2/2$ and that the hypothesis holds for all $r' \leq r$. Consider $u \in K_{r+1}^2$. Let ρ be the bias of l_i in the initial configuration. First let us form a lower bound for the bias of u in the initial configuration. The fact that **Smooth** holds means that the leftmost and rightmost nodes in K_r^2 have bias at least $\rho - \frac{r}{k_2}\rho - \epsilon_0 w$. Then, since the bias can change by at most 2 if we move left or right one node, we conclude that in the initial configuration u has bias

$$\rho_1 \geq \rho - \left(\frac{r}{k_2}\rho + \epsilon_0 w + \frac{2(2w+1)}{k_2}\right).$$

Now we have to take into account all of the β nodes in $\bigcup_{j \leq r} K_r^2$ which have changed type. In fact, so that we can be sure that each change of type affects the bias of u, we shall consider just those which lie between l_i and u. Let θ be the proportion of the nodes in $\mathcal{N}(l_i)$ which are α nodes in the initial configuration (recalling that l_i has borderline bias at that point) – so that $\rho = (2\theta - 1)(2w+1)$. Then the number of β nodes in $\bigcup_{j \leq r} K_j^2$ in the initial configuration, which lie between l_i and u, is at least:

$$(1 - \theta - \epsilon_0)(2w+1)r/k_2.$$

Each change of type for one of these nodes means an increase of 2 in the bias of u, so that at stage s, u has bias:

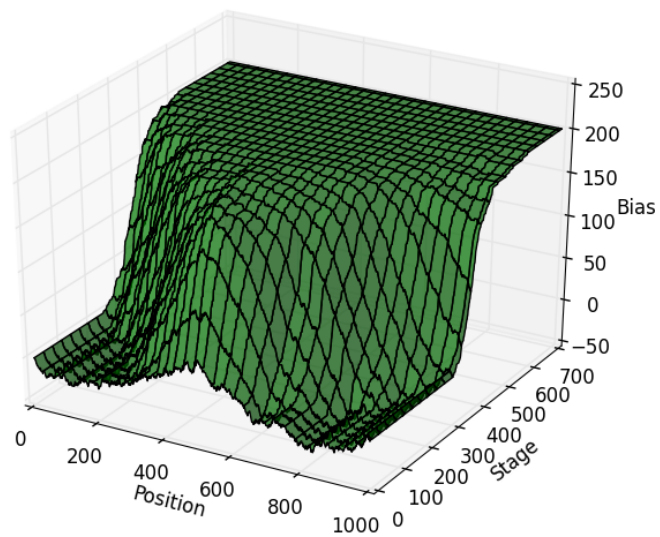

FIGURE 6. The creation of a firewall (extracted from simulation).

$$\rho_2 \geq \rho - \left(\frac{r}{k_2}\rho + \epsilon_0 w + \frac{2(2w+1)}{k_2}\right) + (1-\theta-\epsilon_0)2(2w+1)r/k_2.$$
$$= \rho + (2w+1)\left(\frac{(1-\theta)2r}{k_2} - \frac{\epsilon_0 w}{2w+1} - \frac{2}{k_2} - \frac{r}{k_2}(2\theta-1) - \frac{\epsilon_0 2r}{k_2}\right)$$

We are left to compare the terms

$$\frac{(1-\theta)2r}{k_2},\ \frac{\epsilon_0 w}{2w+1},\ \frac{2}{k_2},\ \frac{r}{k_2}(2\theta-1) \text{ and } \frac{\epsilon_0 2r}{k_2}.$$

Since $1/k_2 \gg \epsilon_0$, the second term is much smaller than the third. Since $\theta \in (0.5, 0.65)$, $r/k_2 \geq 1/k_1$ and $k_1 \ll k_2$, the first term is much larger than the third (to see that $\theta \in (0.5, 0.65)$ recall that l_i has borderline bias in the initial configuration and $\tau > \kappa > 0.35$). Since ϵ_0 is small, the first term is also much larger than the last. The result then follows for large w, since $2 - 2\theta$ is always more than double $2\theta - 1$ for $\theta \in (0.5, 0.65)$, meaning that the first term is more than double the fourth term, and thus $\rho_2 \geq \rho$, meaning that if u is a β node, it will be unhappy. □

Lemma 4.22 (Reasonable chance of firewall)**.** *Suppose that* ***Good spacing*** *holds. There exists $\delta > 0$ which does not depend on w, such that if $1 \leq i < k_0$, then l_i originates a firewall with probability $> \delta$ (and similarly for r_i).*

Proof. Rather than working directly with l_i, to begin with we consider u, which is chosen uniformly at random from amongst the nodes with borderline bias in the initial configuration. We extend the definitions of completion and firewall origination to apply to u in the obvious way, and we establish that there exists $\delta > 0$ which does not depend on w, such that u originates a firewall with probability $> \delta$. Let θ be the proportion of the nodes in $\mathcal{N}(u)$ which are of type α in the initial configuration. Without loss of generality, suppose that u has positive bias ρ. Let $K_1^1(u)$ be as defined in the proof of Lemma 4.21, but relative to u, i.e. with u replacing l_i in that definition. We first look to

establish that there exists $\delta' > 0$ which does not depend on w, such that the probability of u completing with all nodes in $K_1^1(u)$ being of type α, is greater than δ'.

Initially let us adopt the approximation that the nodes in $\mathcal{N}(u)$ are i.i.d. with probability θ of being an α node.

We consider a process, consisting of $\ell := \lceil \frac{2w+1}{k_1} \rceil + 1$ many steps.

Step 0. If u is of type β then define $\rho_0 = \rho + 2$, and otherwise define $\rho_0 = \rho$.

Step $s > 0$ (with $s \leq \ell$). Consider the two nodes $u - s$ and $u + s$. Let x_s be the number of these which are β nodes (so $x_s \in \{0, 1, 2\}$). Let y_s be $\Theta^*(\mathcal{N}(u - s)) - \Theta^*(\mathcal{N}(u - s + 1))$. Then define $\rho_s = \rho_{s-1} + 2x_s + y_s$.

So, given the approximation that the nodes in $\mathcal{N}(u)$ are i.i.d. with probability θ of being an α node, this gives a biased random walk. We may also consider the mirror image process, in which y_s is defined in terms of the bias at $u + s$ and $u + s - 1$ instead of the bias at $u - s$ and $u - s + 1$. Suppose that this process gives a set of values ρ'_s. If $\rho_s > \rho$ and $\rho'_s > \rho$ for all s then let us say that u is *prone to firewall origination* - the idea is that if u also completes and only has one type of unhappy node in $\mathcal{N}(u)$ in the initial configuration (though we do not yet assume these conditions), then being prone to firewall origination will mean that all the nodes in $K_1^1(u)$ must be α nodes at completion. To show that u has some reasonable chance of being prone to firewall origination, it suffices (modulo our false assumption that the nodes in $\mathcal{N}(l_i)$ are i.i.d.) to show that $\rho_s - \rho_{s-1}$ is more likely to be positive than negative, for $0 < s \leq \ell$.

The probability that $\rho_s - \rho_{s-1} = -2$ is $\frac{1}{2}\theta^3$. The probability that $\rho_s - \rho_{s-1} = 0$ is $\frac{1}{2}\theta \cdot 2\theta(1-\theta) + \frac{1}{2}\theta^2 = \theta^2(1-\theta) + \frac{1}{2}\theta^2$. Therefore the probability that $\rho_s - \rho_{s-1} \in \{2, 4, 6\} = 1 - \frac{1}{2}\theta^3 - \theta^2(1-\theta) - \frac{1}{2}\theta^2 > \frac{1}{2}$, given that l_i has positive borderline bias in the initial configuration, and so $\theta \in (0.5, 0.65)$.

Now we must drop the false assumption of independence which was made earlier, and work with the hypergeometric distribution on $\mathcal{N}(u)$. In the random walk described above, let $p > \frac{1}{2}$ be the probability that $\rho_s - \rho_{s-1} > 0$. Now choose p' with $\frac{1}{2} < p' < p$. For sufficiently large k_1,[12] when we drop the assumption of independence, the actual probability that $\rho_s - \rho_{s-1} > 0$ at each step is greater than p', no matter what has occurred at previous steps.

So far then, we have been able to conclude that there exists $\delta' > 0$, which does not depend on w, such that u has probability $> \delta'$ of being prone to firewall origination. Next, note that by Corollary 4.16 and by the same argument as in the proof of Lemma 4.19, the probability that u completes and satisfies $\texttt{Smooth}^*_{k_2,\epsilon_0}(u)$ tends to 1 as $w \to \infty$. Satisfaction of these conditions combined with being prone to firewall origination, means that all the nodes in $K_1^1(u)$ must be α nodes at completion. Then, as observed in the proof of Lemma 4.21, this latter condition combined with satisfaction of $\texttt{Smooth}^*_{k_2,\epsilon_0}(u)$ means that, in fact, all nodes in $\mathcal{N}(u)$ must be α nodes when u completes.

Finally, we move to consider l_i rather than u. Given l_i, choose v uniformly at random from amongst the nodes in $\mathcal{N}(l_i)$ with borderline bias. Now the information we have about v completely specifies the probability that v originates a firewall, and at this point, since we have only assumed **Good spacing**, the only information we have about v is: $(*)$ v has borderline bias, belongs to a configuration which satisfies **Good spacing** relative to *some* node (which happened to be u_0 in this case), and for the node labelled l_i within that configuration, belongs to $\mathcal{N}(l_i)$. For u chosen uniformly at random from amongst those with borderline bias, the probability that u satisfies $(*)$ tends to 1 as $w \to \infty$. So if u has probability $\delta' > 0$ of originating a firewall, then for any $\delta < \delta'$, we may conclude that a node chosen uniformly at random from amongst those which satisfy $(*)$, has probability $> \delta$ of originating a firewall so long as w is sufficiently large. Thus v has probability $> \delta$ of originating a firewall, so long as w is sufficiently large. Now we assume that **Completion** and **Smooth** both hold. Then v must belong to $K_1^1(l_i)$, meaning that if v originates a firewall then l_i belongs to that same firewall when v completes, and so must also originate a firewall. □

The following definition gives our final event in Π and also specifies k_0.

Definition 4.23 (Defining **Firewall** and choosing k_0)**.** *We let **Firewall** be the event that one of the l_i $i < k_0$ is defined and originates a firewall, and that the same holds for some r_j, $j < k_0$. According to Lemma 4.22, given $\epsilon > 0$ we*

[12] This works for all $k_1 > c$ for some constant c (which can actually be chosen so as not to depend even on τ, given that $\kappa < \tau < \frac{1}{2}$), so strictly speaking we should have required in Definition 4.18 that $k_1 > c$. Since this would have been confusing at that point, we mention it now instead, and observe that there is no circularity, i.e. c certainly need not depend on k_1 or any values defined in terms of k_1.

can choose k_0 once and for all, which is large enough such that the probability ***Firewall*** *does not occur is $\ll \epsilon$ for $0 \ll w \ll n$.*

The following lemma completes the proof of Theorem 1.2 for the simple model. The proof is given in Section 8.

Lemma 4.24 (u_0 ultimately joins a firewall)**.** *Suppose that all events in Π hold. Let i be the least such that l_i is defined and originates a firewall, and let j be least such that r_j is defined and originates a firewall. For any $\epsilon' > 0$, for all sufficiently large w, with probability $> 1 - \epsilon'$, u_0 will eventually be contained in one of the two firewalls originated at l_i and r_j.*

The standard model. The difficulty that arises when one moves to the standard model, is that when there are different numbers of unhappy α and β nodes, it is no longer true that every unhappy node is equally likely to be chosen as part of a swapping pair. If there are more unhappy α nodes than unhappy β nodes at a given stage, for example, then unhappy β nodes belong to more unhappy pairs of opposite type than do their unhappy α counterparts, and so are more likely to be chosen as part of an unhappy pair. This potentially complicates our proof of Lemma 4.19, that each l_i and r_i will almost certainly complete, for example, since it may make 'steps towards completion' (as defined in that proof) less likely. The solution, just as in [5], is to use technology developed by Wormald [29] in order to show that we can sufficiently accurately model the discrete process with a continuous one which is governed by a system of differential equations, and thereby demonstrate that the number of unhappy nodes of each type actually remains very evenly balanced. Our use of the Wormald machinery is very similar to the corresponding argument in [5]. We give the full proof since there are some minor differences, and to make the paper as self-contained as possible. We aim to prove the following fact, which shows that our proof of Theorem 1.2 for the simple model suffices for the standard model as well (the proof of ($\diamond$) appears in Section 8):

($\diamond$) Suppose w, τ (with $\kappa < \tau < \frac{1}{2}$) and $\epsilon > 0$ are fixed (where ϵ is the value we have fixed throughout this section, and which played a role in the definition of k_0). Let l_{k_0} and r_{k_0} be as defined previously. For any $\epsilon' > 0$, when n is sufficiently large the following holds with probability $> 1 - \epsilon'$: there exists a first stage at which there are no unhappy nodes in the interval $[l_{k_0}, r_{k_0}]$ and at all stages up to this the total number of unhappy α nodes divided by the total number of unhappy β nodes lies in the interval $[1 - \epsilon', 1 + \epsilon']$.

Increasing τ in the interval $[\kappa, \frac{1}{2}]$ decreases run-lengths. Let us fix $\epsilon > 0$ and τ_1 and τ_2 such that $\kappa < \tau_1 < \tau_2 < \frac{1}{2}$. For some large w and $n \gg w$ suppose that we run two version of the process, one for τ_1 and the other for τ_2, and then let ℓ_1 and ℓ_2 be the corresponding run-lengths to which u_0 (chosen uniformly at random) belongs in the final configuration. Our aim in this subsection is to observe that, so long as w is sufficiently large, the probability that $\ell_1 > \ell_2$ is greater than $1 - \epsilon$.

In order to see this we may reason as follows. In the proof of Theorem 1.2, we chose k_0 in such a way we could be almost certain at least one of the l_i and at least one of the r_i would originate firewalls. We could equally well have chosen k_0 so that the following fails to occur with probability $\ll \epsilon$: two of the l_i originate firewalls of opposite type, and similarly two of the r_i originate firewalls of opposite type. Our previous analysis then suffices to show that the following fails to occur with probability $\ll \epsilon$ for sufficiently large w: u_0 ultimately belongs to a firewall of length $> \min\{|u_0 - l_1|, |u_0 - r_1|\}$ and of length $< [l_{k_0}, r_{k_0}]$. So it suffices to show that with probability $> 1 - \epsilon$, the length of the interval $[l_{k_0}, r_{k_0}]$ for the process with τ_2 is less than the value $\min\{|u_0 - l_1|, |u_0 - r_1|\}$ for the process with τ_1. This then follows for sufficiently large w, by applying Lemma 3.3 to the events P_u, that u is τ_2-unhappy, and Q_u that u is τ_1-unhappy.

5. The case $\tau = \frac{1}{2}$

This case was dealt with in [5], for both the simple and standard models. Although we have talked about threshold behaviour occurring at both $\tau = \kappa$ and $\tau = \frac{1}{2}$, it is worth noting that these two thresholds are of essentially different types. For a fixed large w, as one increases the value of τ past the lower threshold κ, one will very suddenly observe entirely different results for the final configuration – from a situation in which almost no type changing occurs, one moves to a situation where u chosen uniformly at random can be expected to belong to a firewall of length exponential

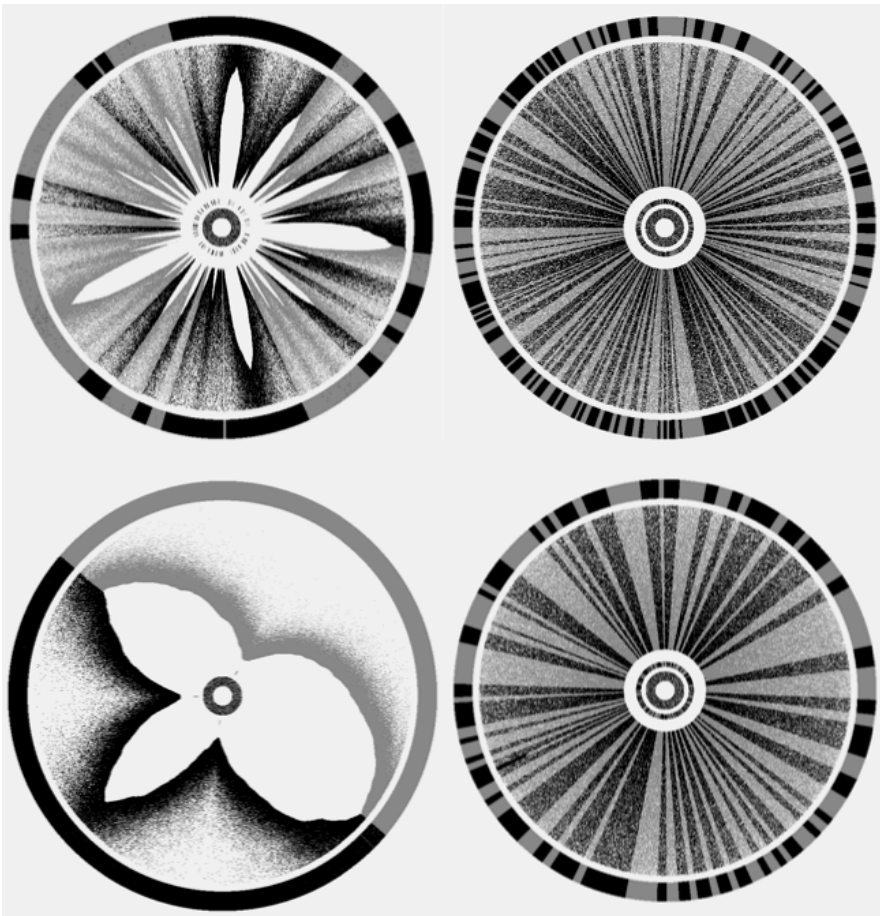

FIGURE 7. The second kind of threshold behaviour: top left $w = 1500, \tau = 0.48$; bottom left $w = 3000, \tau = 0.48$; top right $w = 1500, \tau = 0.5$; bottom right $w = 3000, \tau = 0.5$.

in w in the final configuration. As one increases τ gradually up to $\frac{1}{2}$, however, what one observes is a smooth decrease in the expected length of the firewall to which u belongs, with the situation only reversing very suddenly when $\tau > \frac{1}{2}$. In order to observe the difference in behaviour for when τ is just less than $\frac{1}{2}$ or when $\tau = \frac{1}{2}$, one should consider instead these two fixed τ, and gradually increase w. Then, when $\tau < \frac{1}{2}$, the expected length of the firewall to which u belongs grows exponentially, while for $\tau = \frac{1}{2}$ it grows only polynomially in w. Figure 7 displays this difference in behaviour, and also illustrates the distinct mechanisms involved for the two cases. These diagrams illustrate the process exactly as in Figure 1, the difference being that now $n = 1000000$ rather than 100000. It is also interesting to observe the difference in the way that the firewall spreads as a 'wave' emanating from unhappy nodes in the initial configuration for different $\tau < \frac{1}{2}$. When τ is just above κ the wave is sharply defined, (as illustrated in Figure 1), much less so when τ is just below $\frac{1}{2}$.

6. THE CASE $\tau > \frac{1}{2}$

We consider the standard model first. Note that if $\frac{1}{2} < \tau \leq \frac{w+1}{2w+1}$ then the process is identical to that for $\tau = \frac{1}{2}$. We therefore assume in what follows that w is sufficiently large to ensure $\tau > \frac{w+1}{2w+1}$, which is equivalent to the condition

that adjacent nodes of opposite types cannot both be happy. We show that with probability $1-\epsilon$ the starting configuration is such that complete segregation results with probability 1, where $\epsilon \to 0$ as $n \to \infty$. Recall that complete segregation refers to any configuration in which all α nodes belong to a single run, and that, as observed in Section 2, once a completely segregated configuration is reached all future configurations must be completely segregated.

Before giving a formal analysis it is instructive to informally explain the difference between the cases $\tau = \frac{1}{2}$ and $\tau > \frac{1}{2}$. In both scenarios the process quickly reaches a configuration in which all nodes belong to short firewalls. The essential difference between the two processes, is that once such a configuration has been reached in the case $\tau = 0.5$, all nodes will then be happy, meaning that the process immediately terminates. For larger τ, however, nodes at the ends of firewalls will be unhappy and will be able to swap with nodes of the opposite type which are at the ends of firewalls of the opposite type. A lengthy process then unfolds, during which all of the various firewalls increase and decrease in length, joining together until ultimately complete segregation is reached. As a consequence, the time to termination when $\tau = \frac{1}{2}$ is $O(n)$, while the time to termination for larger τ is much greater.

Let x be the number of α nodes in the initial configuration, and let $x_p = x/n$. Our task is to show that for sufficiently large n it is possible to reach a completely segregated configuration from any other, so long as x_p is close to $\frac{1}{2}$. To prove this, however, one must be able to ensure the existence of unhappy individuals of both types at each step along the way. Surprisingly, the following lemma turns out to be the most tricky aspect of proving Theorem 1.4. The proof is given in Section 9. The mention of intervals of length $4w+1$ in the statement of the lemma is a just a technicality which will be convenient later.

Lemma 6.1. *With probability $1-\epsilon$ the initial configuration on n nodes will have a value x sufficient to ensure unhappy nodes of both types outside any interval of length $4w+1$ in **any** configuration with x-many α nodes and $(n-x)$-many β nodes (regardless of ordering), where $\epsilon \to 0$ as $n \to \infty$.*

From now on, then, we assume that all configurations considered have unhappy nodes of both types outside any given interval of length $4w+1$. We build a list of configurations from which it is possible to reach complete segregation. Let $\Theta_s(u)$ denote the bias of u at the end of stage s. Note that in order for a swap between an unhappy α node u and an unhappy β node v to be legal at stage $s+1$, it suffices either that $v \notin \mathcal{N}(v)$ and $\Theta_s(v) \geq \Theta_s(u) - 2$ (since then v changing type will increase the bias at that position by 2 to become at least $\Theta_s(u)$, and similarly u changing type will decrease the bias by 2 to become at most $\Theta_s(v)$), or else that $v \in \mathcal{N}(u)$ and $\Theta_s(v) \geq \Theta_s(u)$. One consequence of this, is that an unhappy α node can always legally swap into the position next to a happy α node if occupied by a β node.

Consider first any configuration which is not completely segregated, but which has a run of length at least $2w$. Without loss of generality, suppose that this is a run of α nodes occupying the interval $[a,b]$, where this interval is chosen to be of maximum possible length. Our aim is to show that from this configuration, one may legally reach another with greater maximal run length. Now if the nodes a and b are both happy then the length of the interval ensures that all nodes in the run are happy – this follows by induction on the distance from the edge of the interval by considering the difference between successive neighbourhoods. In this case let u be an unhappy α node (whose existence we have assumed is guaranteed) and let $c \in \{a,b\}$ be distance at least $w+1$ from u. Then u and the β neighbour of c may legally be swapped, increasing the length of the run by at least 1.

So suppose instead that at least one of the individuals a and b is not happy, and without loss of generality suppose that a has bias less than or equal to b. Then a and $b+1$ may legally be swapped. Performing this swap causes position $b+1$ to have at least the same bias as b did before the swap, and causes $a+1$ to have at most the same bias as a did before the swap. Thus, the swap has the effect of shifting the run one position to the right and may be repeated until the length of the run is increased by at least 1, i.e. for successive $i \geq 0$ we can swap the nodes $a+i$ and $b+i+1$, so long as the latter is of type β. The first stage at which the latter is of type α the length of the run has been increased. Putting these observations together, we conclude that from any configuration which has a run of length at least $2w$ it is possible to reach full segregation.

Next consider a configuration in which the longest run $[a,b]$ is of length at least w, but strictly less than $2w$. We shall suppose that $[a,b]$ contains α nodes, the case for β nodes is similar. Let c be the first α node strictly to the left of a.

If c is unhappy, then we may legally swap c and $a-1$, strictly increasing the length of the longest run. In order to see this, compare the neighbourhood of c before the swap with the neighbourhood of $a-1$ after the swap. If $|c-a| > w$, then prior to the swap there are at most $w+1$ many α nodes in $\mathcal{N}(c)$, while $\mathcal{N}(a-1)$ has $w+1$ many after the swap. If $|c-a| \leq w$, then in forming $\mathcal{N}(a-1)$ from $\mathcal{N}(c)$, we lose some nodes to the left of c, some of which may be β nodes, and gain some to the right of $a-1$, all of which are α nodes. If c is happy then the distance between c and a is at most w (otherwise c would only have α nodes in the left half of its neighbourhood) and we may successively swap unhappy α nodes from outside the interval $[c-w, a+2w]$ with the nodes $c+i$ for $1 \leq i < a-c$ (starting with $i=1$ and proceeding in order), in order to strictly increase the length of the longest run. This follows because with the ith swap $c+i$ becomes a happy α node.

It remains to show that we can always move to a configuration with a run of length at least w. Given any configuration (satisfying the condition that any other configuration reached by swapping unhappy nodes has unhappy nodes of both types outside any given interval of length $4w+1$), we first of all perform a procedure which selects a number of individuals with relatively unfavourable bias. Let u_0 be an α node with least possible bias. Given u_k for $k < w^2$ choose u_{k+1} outside $\bigcup_{j \leq k} \mathcal{N}(u_j)$ which has least possible bias amongst all the α nodes at such sites (we can suppose that n is large enough that such a choice is possible). Once u_k is defined for each $k \leq w^2$, choose v_0 outside $\bigcup_{j \leq w^2} \mathcal{N}(u_j)$, which has the greatest possible bias amongst all β nodes at such sites. Given v_k for $k < w^2$ choose v_{k+1} outside $\bigcup_{j \leq w^2} \mathcal{N}(u_j)$ and outside $\bigcup_{j \leq k} \mathcal{N}(v_j)$ which has greatest possible bias amongst all the β nodes at such sites. Now choose an interval $[a,b]$ of length w such that $[a-w, b+w]$ has no intersection with any of the neighbourhoods $\mathcal{N}(u_j)$ or $\mathcal{N}(v_j)$ for $j \leq w^2$ (again assuming n is sufficiently large).

The point of this procedure was to provide a pool of individuals which we can use for legal swaps. Now we perform another iteration, which produces a run in the interval $[a,b]$. At any point during the iteration we say that a neighbourhood $\mathcal{N}(u_j)$ or $\mathcal{N}(v_j)$ is *tarnished* if any node in that neighbourhood has been involved in a swap since the iteration began.

Step 0. Let $\gamma = \alpha$ if a is of type α, otherwise let $\gamma = \beta$. At any given point in the iteration the value of γ specifies the type of run that we are looking to produce, which may change as the iteration progresses. Set $i := 0$.

Step $s > 0$. We are given a configuration in which all nodes in $[a, a+i]$ are of type γ, and no nodes in $[a-w, b+w]$ have yet been involved in swaps, except possibly those in $[a, a+i]$. Also, at most i^2 of the neighbourhoods $\mathcal{N}(u_k)$ or $\mathcal{N}(v_k)$ are tarnished. Let j be the least $> i$ such that $a+j$ is not of type γ. If $j \geq w$ then the iteration is complete, and we carry out no further instructions. Otherwise, if $a+j$ is unhappy, we divide into two subcases. If $a+j-1$ is happy then we can legally swap any unhappy γ node from outside the interval $[a-w, b+w]$ with $a+j$. If $a+j-1$ is unhappy, we can select any u_k or v_k of type γ whose neighbourhood is not tarnished and legally swap that node with $a+j$. In order to see why this is the case suppose that $\gamma = \alpha$, the case for β is similar. Firstly, the fact that at most i^2 of these neighbourhoods are tarnished means that we certainly have untarnished neighbourhoods of the appropriate type to choose from. Also, before the iteration began any u_k had at most bias $\Theta(\mathcal{N}(a+j-1))$ (as that value was defined then). At the present point in the iteration, $\mathcal{N}(u_k)$ being untarnished means that the neighbourhood $\mathcal{N}(u_k)$ is unchanged, while any node in the interval $[a-w, b+w]$ which is now of different type than before, is presently of type α. Swapping u_k with $a+j$ will cause it to have bias at least the same as that which $a+j-1$ had before the swap. In either of these two subcases, once the swap is performed, redefine $i := j$ and proceed to step $s+1$. The final case to consider is that the individual at $a+j$ is happy. In this case, if $\gamma = \alpha$ then redefine $\gamma = \beta$, or if instead $\gamma = \beta$ then redefine $\gamma = \alpha$. For successive values of k, $1 \leq k \leq j$, swap an unhappy γ individual from outside the interval $[a-w, b+w]$ with $a+j-k$, causing that node to become happy. Once this sequence of swaps is complete, redefine $i := j$ and proceed to step $s+1$.

The simple model. We wish to show that, whatever the initial configuration, with probability 1 a configuration in which all nodes are of the same type is reached. For $\gamma \in \{\alpha, \beta\}$, if $\gamma = \beta$ then let $\gamma^* = \alpha$, and otherwise let $\gamma^* = \beta$. For the purposes of this discussion, we shall say that γ is a *minority type* if there are at most as many nodes of type γ as of type γ^*. The proof of Lemma 6.1 actually suffices to show that if γ is a minority type, then there exists at least one γ node which is unhappy, and which would have at least as many neighbours of its own (new) type if it changed

type. Given any configuration, one can then select γ of minority type and successively select nodes of type γ which can legally have their type swapped, until all nodes are of type γ^*.

7. Proofs deferred from Section 3

Lemma 3.2. *Suppose $h : \mathbb{N} \to \mathbb{N}$ and $p \in (0,1)$ are such that there exists $k \in (0,1)$ so that for all large enough N, we have $\left(1 + \left(\frac{1}{p} - 1\right)k\right)h(N) > N \geq h(N) > pN > 0$. Then for all large enough N, if $X_N \sim b(N,p)$, we have*

$$P(X_N = h(N)) \;\leq\; \boldsymbol{P}(X_N \geq h(N)) \;\leq\; \left(\frac{1}{1-k}\right) \cdot \boldsymbol{P}(X_N = h(N)).$$

That is to say in asymptotic notation, $\boldsymbol{P}(X_N \geq h(N)) = \Theta(\boldsymbol{P}(X_N = h(N)))$.

Proof. Fix N, and for the duration of this proof let h denote $h(N)$. The first inequality is self-evident. For the second, $\mathbf{P}(X_N \geq h) = \sum_{j=h}^{N} p^j (1-p)^{N-j} \binom{N}{j}$. Looking at the ratio between successive terms of this sum, we find

$$p^{j+1}(1-p)^{N-j-1}\binom{N}{j+1} \Big/ \left(p^j(1-p)^{N-j}\binom{N}{j}\right) = \frac{p}{1-p} \cdot \frac{N-j}{j+1}.$$

What is more, this ratio decreases at each step, since $\frac{N-j}{j+1} > \frac{N-(j+1)}{j+2}$. Thus for all $j = h, \ldots, N$, we have $\frac{p}{1-p} \cdot \frac{N-j}{j+1} \leq \frac{p}{1-p} \cdot \frac{N-h}{h+1} < k$. Thus

$$\mathbf{P}(X_N \geq h(N)) < p^h(1-p)^{N-h}\binom{N}{h}\sum_{j=0}^{N-h} k^j < p^h(1-p)^{N-h}\binom{N}{h}\left(\frac{1}{1-k}\right).$$

□

Lemma 3.3. *Let P_u and Q_u be events which only depend on the neighbourhood of u in the initial configuration, meaning that if the neighbourhood of v in the initial configuration is identical that of u (i.e. for all $i \in [-w, w]$, $u + i$ is of the same type as $v + i$), then P_u holds iff P_v holds and Q_u holds iff Q_v holds. Suppose also that:*

(i) $\boldsymbol{P}(P_u) \neq 0$ and $\boldsymbol{P}(Q_u) \neq 0$.

(ii) For all k, for all sufficiently large w, $\boldsymbol{P}(P_u)/\boldsymbol{P}(Q_u) > kw$.

For any u, let x_u be the first node to the left[13] of u such that either P_{x_u} or Q_{x_u} holds. For any $\epsilon > 0$, if $0 \ll w \ll n$ then the following occurs with probability $> 1 - \epsilon$ for u chosen uniformly at random: x_u is defined and for no node v in $[x_u - 2w, x_u]$ does Q_v hold.

An analogous result holds when 'left' is replaced by 'right'.

Proof. For a general node u, we'll say it is of *type 1* if P_u holds and no node $u' \in [u - 2w, u]$ satisfies $Q_{u'}$, and of *type 2* if Q_u holds. Let π be the probability, for u chosen uniformly at random, that x_u is defined and of type 1. Our aim is to show that $\pi > 1 - \epsilon$, for all $0 \ll w \ll n$. We define an iteration which assigns colours to nodes as follows.

Step 0. Pick a node t_0 uniformly at random.

Step $s + 1$. Let v_s be the first node to the left of t_s, such that $v_s = x_{t_s}$ or such that $s > 0$ and $v_s = t_0$. Carry out the instructions for the first case below which applies:

(1) If there exists no such v_s then terminate the iteration.

(2) If $v_s = t_0$ and $s > 0$ then make t_s undefined and terminate the iteration.

[13] By the first node to the left of u satisfying a certain condition we mean the first in the sequence $u, u-1, u-2, \cdots$ which satisfies the condition.

(3) If there exist any nodes x in $[v_s - 2w, v_s]$ such that Q_x holds, then colour t_s black, otherwise it is of type 1 and we colour it white. Define $t_{s+1} = v_s - (2w+1)$, unless t_0 lies in the interval $[v_s - (2w+1), v_s)$, in which case terminate the iteration.

This completes the description of the iteration. Let S be the maximum value of s for which t_s is defined and coloured.

First note that hypothesis (i) in the statement of the lemma guarantees that $S \to \infty$ as $n \to \infty$. Similarly, we may assume that at least one t_s is coloured black. Now let π' be the proportion of the t_s which are coloured white, and observe that with high probability, $\pi' \to \pi$ as $n \to \infty$, i.e. for all $\epsilon' > 0$, for all $0 \ll n$, we have $|\pi - \pi'| < \epsilon'$ with probability $> 1 - \epsilon'$. In order to see this, consider the situation at the beginning of step $s+1$ of the iteration, when $s > 0$. In order to define t_s, we moved left until finding $x_{t_{s-1}}$, and then moved a further $2w+1$ nodes to the left. So far, then, nothing that has happened in the iteration tells us anything about the neighbourhood of t_s and all those nodes to the left of t_s and strictly to the right of t_0. So long as $S > s$, the way in which t_s is coloured depends only on these nodes.

So, we now wish to show that for all sufficiently large w, with probability tending to 1 as $n \to \infty$, we have $\pi' > 1 - \epsilon$. Let ρ be the ratio of type 1 nodes to type 2 nodes (in $[0, n-1]$), and consider the interval $(t_{s+1}, t_s]$. However t_s is coloured, we have at most $(2w+1)$ many type 1 nodes in here, namely some subset of $[v_s - 2w, v_s]$. Similarly, if t_s is coloured black then we get at least one type 2 node in here. Thus, summing over all intervals $(t_{s+1}, t_s]$, we find that

$$\rho \leq \frac{(2w+1)S}{|\{s \leq S \ : \ t_s \text{ is black }\}|} \quad \text{which is to say} \quad \rho \leq \frac{2w+1}{1-\pi'}.$$

So if we can show that $\rho \gg 2w+1$, it will follow that π' is close to 1 as required. Let p_1 and p_2 be the probabilities that uniformly randomly selected u itself is of type 1 and 2 respectively. Then, by the weak law of large numbers, we have $\left|\rho - \frac{p_1}{p_2}\right| < \epsilon'$ with probability approaching 1 as $n \to \infty$. Now property (ii) from the statement of the lemma means that for each k, for all sufficiently large w, $\frac{p_1}{p_2} > kw$, giving the result. □

8. Proofs deferred from Section 4

8.1. Proving Lemma 4.10.

First, let us restate the result:

Lemma 4.10. *For fixed k_0 and $\epsilon' > 0$, $\boldsymbol{P}$(**Stable clear**) $> 1 - \epsilon'$ for all w sufficiently large (and all n sufficiently large compared to w).*

Proof. For any u, let x_u be the first node to the left of u which, in the initial configuration, either has high bias or else belongs to an interval which is τ_0-stable. Recall that $\kappa < \tau_0 < \tau$ and that k_0 is fixed while we take w large. Note also that, at step i of the iteration which defines the sequence $l_1, l_2, ..$, the fact that l_i has borderline bias tells us nothing about the neighbourhood of $l_i - (2w+1)$ or the nodes to the left of this neighbourhood (but at a distance small compared to n). It therefore suffices to show that for any $\epsilon' > 0$, if $0 \ll w \ll n$ then the following occurs with probability $> 1 - \epsilon'$ for u chosen uniformly at random: x_u is defined and no node in $[x_u - 2w, x_u]$ belongs to a τ_0-stable interval (and that an analogous result holds when 'left' is replaced by 'right').

In order to see this, consider for a moment working with τ_0 rather than τ. In Section 3 we showed that for any k and for sufficiently large w, the probability that a randomly chosen node is unhappy (with respect to τ_0 rather than τ) is more than kw times the probability that a randomly chosen node belongs to a τ_0-stable interval. Now, since $\tau > \tau_0$, this only makes unhappy nodes *more* likely in the initial configuration. Thus, working with τ, for any k and for all sufficiently large w, the probability that a randomly chosen node is unhappy is more than kw times the probability that a randomly chosen node belongs to a τ_0-stable interval. The result then follows from Lemma 3.3. □

8.2. **Proving Lemma 4.17.** We shall need to prove an intermediate lemma first. It turns out that in order to get around the fact that we can't immediately guarantee the required version of condition (ii) of Lemma 3.3, what we need is a bound on the number of nodes of borderline bias which can be expected in the neighbourhood of u such that $\mathtt{Bb}^*(u)$ holds. The following lemma is a step in this direction:

Definition 8.1. *Given $k \geq 1$, let $\mathcal{N}_k(u)$ be the interval $[u - \lceil w/k \rceil, u + \lceil w/k \rceil]$. $\mathtt{Bb}^*(u, k, z)$ holds if $\mathtt{Bb}^*(u)$ holds and there are at most z many nodes with borderline bias in $\mathcal{N}_k(u)$ in the initial configuration.*

Lemma 8.2. *Suppose we are given only that $\mathtt{Bb}^*(u)$ holds. For any $\epsilon' > 0$ there exists z such that for $0 \ll k \ll w$, $\mathtt{Bb}^*(u, k, z)$ holds with probability $> 1 - \epsilon'$.*

Proof. We consider the case that u has positive bias ρ, and let θ be the proportion of the nodes in $\mathcal{N}(u)$ which are of type α (so $\rho = (2\theta - 1)(2w + 1)$). The case that u has negative bias is almost identical.

First of all, we want to show that for sufficiently large z, if we step $\lfloor z/2 \rfloor$ many nodes to the left (or right) of u, then we will very probably have bias which is well below ρ. The argument here is very similar to the proof of Corollary 4.16 – in forming the neighbourhood of $v = u - \lfloor z/2 \rfloor$ we lose $\lfloor z/2 \rfloor$ many nodes from $\mathcal{N}(u)$, with the proportion of α nodes being close to θ, and we gain the same number of new nodes, with the proportion of α nodes here being close to $\frac{1}{2}$. Arguing more precisely, from amongst the nodes that we lose from $\mathcal{N}(u)$, the expected number of α nodes, x_0 say, is $\theta\lfloor z/2 \rfloor$. By applying Chebyshev's Inequality just as in the proof of Lemma 4.14, we conclude that for any $\epsilon'' > 0$ and for all sufficiently large z, $\mathbf{P}(|(x_0/\lfloor z/2 \rfloor) - \theta| > \epsilon'') \ll \epsilon'$. Now consider x_1, the number of α nodes in $\mathcal{N}(v) \backslash \mathcal{N}(u)$. The weak law of large numbers tells us that for any $\epsilon'' > 0$ and for all sufficiently large z, $\mathbf{P}(|(x_1/\lfloor z/2 \rfloor) - \frac{1}{2}| > \epsilon'') \ll \epsilon'$. Combining these facts gives that for any $m > 0$ and for all sufficiently large z:

$$\mathbf{P}(\rho - \Theta(\mathcal{N}(v)) < m) \ll \epsilon'.$$

So far then, we have considered moving $\lfloor z/2 \rfloor$ many nodes to the left of u, and have concluded that the bias at this node v will very probably be well below ρ (a similar argument also applies, of course, moving to the right). Now we have to show that as we move left from v, so long as we remain within $\mathcal{N}_k(u)$ the bias will very probably remain below ρ. In order to do this, we approximate the bias as we move to the left, by a biased random walk. Recall that for a biased random walk starting at value $-m \leq -1$, with probability $p > \frac{1}{2}$ of going down at each step and probability $1 - p$ of going up, the probability of ever hitting 0 is:

$$\left(\frac{1-p}{p}\right)^m.$$

So let us briefly adopt the approximation that nodes in $\mathcal{N}(u)$ are i.i.d. random variables, each of which has probability θ of being of type α. Then, as we move left one position from a location in $\mathcal{N}_k(u)$ to the left of u, the probability that the bias increases by 2 is $\frac{1}{2}(1 - \theta)$, the probability that the bias remains the same is $\frac{1}{2}$, and the probability that the bias decreases by 2 is $\frac{1}{2}\theta$. Removing those steps at which the bias does not change, we get a biased random walk with probability θ of going down at each step and probability $1 - \theta$ of going up. Choose θ' with $\frac{1}{2} < \theta' < \theta$. Now, dropping the false assumption of independence, by taking k sufficiently large we ensure that as we take successive steps left from v inside the interval $\mathcal{N}_k(u)$, at each step, no matter what has occurred at previous steps, the probability of the bias increasing is less than $\frac{1}{2}(1 - \theta')$, the probability of the bias remaining the same is $\frac{1}{2}$, and the probability of the bias decreasing is greater than $\frac{1}{2}\theta'$. Thus, if the bias at $u - \lfloor z/2 \rfloor$ is $\leq \rho - m$, then the probability that any nodes in the interval $[u - \lceil w/k \rceil, u - \lfloor z/2 \rfloor)$ have high bias is less than $(\frac{1-\theta'}{\theta'})^m$.

Finally, let m be such that $(\frac{1-\theta'}{\theta'})^m \ll \epsilon'$, and let z be sufficiently large that, for $v = u - \lfloor z/2 \rfloor$, $\mathbf{P}(\rho - \Theta(\mathcal{N}(v)) < m) \ll \epsilon'$. □

The reader might observe that the proof of Lemma 8.2 actually establishes something stronger, and perhaps also more natural, than claimed. Suppose we are given that $\mathtt{Bb}^*(u)$ holds. The proof suffices to show that for any $\epsilon' > 0$ there exists z such that, for $0 \ll k \ll w$, the probability there are more than z nodes of high bias in $\mathcal{N}_k(u)$ is less than ϵ'. One might reasonably wonder why we did not define $\mathtt{Bb}^*(u, k, z)$ to reflect this stronger condition – that there are at

most z many nodes of *high* bias (rather than borderline bias) in $\mathcal{N}_k(u)$. The reason is that in later counting arguments we shall need failure of the condition (in specific circumstances) to guarantee that there are, in fact, at least z many nodes of borderline bias in the relevant neighbourhood for which the condition fails.

Of course, Lemma 8.2, only restricts the number of nodes of borderline bias that will normally occur in the interval $\mathcal{N}_k(u)$. We have to be able to deal with a larger interval:

Definition 8.3. *We say that* $\mathtt{Bb}^*(u,z)$ *holds if* $\mathtt{Bb}^*(u)$ *holds and there are at most* z *many nodes of borderline bias in the interval* $[u-(2w+1), u+(2w+1)]$ *in the initial configuration.*

Now note that if k_2 is sufficiently large compared to k_1, if ϵ' is sufficiently small, and if $\mathtt{Bb}^*(u,k_1,z)$ and $\mathtt{Smooth}^*_{k_2,\epsilon'}(u)$ both hold, then in the initial configuration there are at most z many nodes of borderline bias in the interval $[u-(2w+1), u+(2w+1)]$. Applying Lemmas 8.2 and 4.16 we therefore have:

Corollary 8.4 (Few nodes of borderline bias)**.** *Suppose we are given that* $\mathtt{Bb}^*(u)$ *holds. For any* $\epsilon' > 0$*,* z *may be chosen so that for all sufficiently large* w*,* $\mathtt{Bb}^*(u,z)$ *holds with probability* $> 1-\epsilon'$*.*

We are now finally ready to prove Lemma 4.17, which we now restate:

Lemma 4.17. *For any node* u*, let* x_u *be the first node to the left of* u *which has high bias in the initial configuration. For any* $\epsilon' > 0$ *and* $k \geq 1$*, if* $0 \ll w \ll n$ *and* u *is chosen uniformly at random, then* x_u *is defined and* $\mathtt{Smooth}^*_{,k,\epsilon'}(x_u)$ *holds with probability* $> 1-\epsilon'$*.*

An analogous result holds when 'left' is replaced by 'right'.

Proof. Applying Corollary 8.4, choose z such that, for sufficiently large w, if we are given that v has borderline bias, then $\mathtt{Bb}^*(v,z)$ fails to hold with probability $\ll \epsilon'$, i.e. putting $\epsilon_1 = \mathbf{P}(\neg\mathtt{Bb}^*(v,z)|\mathtt{Bb}^*(v))$, choose z so that $\epsilon_1 \ll \epsilon'$ for sufficiently large w. Let $\epsilon_2 = \mathbf{P}(\neg\mathtt{Smooth}^*_{k,\epsilon'}(v)|\mathtt{Bb}^*(v))$. Then Corollary 4.16 gives that for all sufficiently large w, $\epsilon_2 \ll \epsilon'/z$ – just apply the statement of the corollary to k and $\epsilon'' \ll \epsilon'/z$.

Now define

$$\mu := \frac{\mathbf{P}(\mathtt{Smooth}^*_{k,\epsilon'}(v)|\mathtt{Bb}^*(v,z))}{\mathbf{P}(\neg\mathtt{Smooth}^*_{k,\epsilon'}(v)|\mathtt{Bb}^*(v,z))}.$$

Then

$$\mu = \frac{\mathbf{P}(\mathtt{Smooth}^*_{k,\epsilon'}(v) \wedge \mathtt{Bb}^*(v,z)|\mathtt{Bb}^*(v))}{\mathbf{P}(\neg\mathtt{Smooth}^*_{k,\epsilon'}(v) \wedge \mathtt{Bb}^*(v,z)|\mathtt{Bb}^*(v)))} \geq \frac{1-\epsilon_1-\epsilon_2}{\epsilon_2} \gg \frac{z}{\epsilon'}.$$

Thus for sufficiently large w, with probability approaching 1 as $n \to \infty$, the ratio amongst the nodes v such that $\mathtt{Bb}^*(v,z)$ holds, between the number such that $\mathtt{Smooth}^*_{k,\epsilon'}(v)$ holds and the number such that it does not, is much greater than z/ϵ'.

Consider the initial configuration. We define an iteration which assigns colours to nodes as follows.

Step 0. Pick a node t_0 uniformly at random.

Step $s+1$. Let v_s be the first node v to the left of t_s such that $v = x_{t_s}$ or such that $s > 0$ and $v = t_0$. Carry out the instructions for the first case below which applies ($\neg$ denotes negation):

(1) If there exists no such v then terminate the iteration, and declare that it has 'ended prematurely'.

(2) If $v_s = t_0$ and $s > 0$ then make t_s undefined and terminate the iteration.

(3) If $|v_s - t_s| < 2w+1$ then colour t_s pink.

(4) If $\mathtt{Bb}^*(v_s,z)$ and $\mathtt{Smooth}^*_{k,\epsilon'}(v_s)$ both occur, then give t_s the colours white and blue.

(5) If $\mathtt{Bb}^*(v_s,z)$ and $\neg\mathtt{Smooth}^*_{k,\epsilon'}(v_s)$ both occur, then give t_s the colours white and red.

(6) If v_s satisfies $\neg\mathtt{Bb}^*(v_s,z)$, then give t_s the colour black.

In cases (3)–(6), define $t_{s+1} = v_s - (2w+1)$, unless t_0 lies in the interval $[v_s - (2w+1), v_s)$, in which case terminate the iteration.

This completes the description of the iteration.

First note that the probability that the iteration terminates prematurely can be made arbitrarily small by taking n large, and similarly that we may assume there are t_s of all colours. Now let S be the greatest s such that t_s is defined when the iteration terminates, and let π be the proportion of the t_s, $s \leq S$, such that t_s is coloured black.[14] Amongst the nodes u which have borderline bias, let ρ be the ratio between the number for which $\mathtt{Bb}^*(u,z)$ holds and the number for which it does not – so that for large n, ρ can be expected to be close to $(1-\epsilon_1)/\epsilon_1$. In order to find an upper bound for ρ, first let us find an upper bound for the number of nodes u such that $\mathtt{Bb}^*(u,z)$ holds. Let v_s be as defined in the iteration. However we colour t_s, there can be at most z many nodes u in $I_s := (v_s - (2w+1), v_s]$ which satisfy $\mathtt{Bb}^*(u,z)$, giving an upper bound of $(S+1)z$. Now, let us find a lower bound for the number of nodes with borderline bias for which $\mathtt{Bb}^*(u,z)$ fails. If t_s is coloured pink, then we are not guaranteed any nodes in I_s of borderline bias for which $\mathtt{Bb}^*(u,z)$ fails. If t_s is coloured white, then the same applies. If t_s is coloured black, however, we are guaranteed at least z many nodes of borderline bias in I_s for which $\mathtt{Bb}^*(u,z)$ fails. We therefore get a lower bound of $(S+1)\pi z$. Thus:

$$\rho \leq \frac{z(S+1)}{\pi(S+1)z} = \frac{1}{\pi}.$$

We chose z previously, so that for all sufficiently large w, $\epsilon_1 \ll \epsilon'$. Since ρ is close to $(1-\epsilon_1)/\epsilon_1$ for large n, we infer that for all sufficiently large w, with probability tending to 1 as $n \to \infty$, we have $\rho \gg 1/\epsilon'$, so that $\pi \ll \epsilon'$. For large n, the probability t_s is coloured pink is less than $(2w+1)e^{-w/d}$ (with d as in Definition 4.2). So, for sufficiently large w, with probability tending to 1 as $n \to \infty$, the proportion of the t_s which are not coloured white is $\ll \epsilon'$.

Now let β be the proportion of the t_s which are coloured red. Amongst the nodes such that $\mathtt{Bb}^*(u,z)$ holds, let γ be the ratio between the number for which $\mathtt{Smooth}^*_{k,\epsilon'}(u)$ holds, and the number for which it does not – so that γ can be expected to be close to μ (as defined earlier) for large n. Again, let us form an upper bound for γ. However t_s is coloured, we get at most z many nodes in the interval I_s for which $\mathtt{Bb}^*(u,z) \wedge \mathtt{Smooth}^*_{k,\epsilon'}(u)$ holds. If t_s is coloured red then we get at least one node in the interval I_s for which $\mathtt{Bb}^*(u,z) \wedge \neg\mathtt{Smooth}^*_{k,\epsilon'}(u)$ holds. Thus:

$$\gamma \leq \frac{z(S+1)}{\beta(S+1)} = \frac{z}{\beta}.$$

We previously observed that, for all sufficiently large w, $\mu \gg z/\epsilon'$. Thus, for all sufficiently large w, with probability tending to 1 as $n \to \infty$, we have that $\gamma \gg z/\epsilon'$, so that $\beta \ll \epsilon'$. Since we have already established that, for $0 \ll w \ll n$, the probability t_s is not coloured white is $\ll \epsilon'$, we conclude that for $0 \ll w \ll n$ the probability t_s is coloured blue is $> 1-\epsilon'$. Now for large n, the probability that t_s is coloured blue is less than the probability, for u chosen uniformly at random, that v_u is defined and $\mathtt{Bb}^*(v_u,z) \wedge \mathtt{Smooth}^*_{k,\epsilon'}(v_u)$ holds, which concludes the proof. □

8.3. **The proofs of Lemmas 4.19 and 4.24.**

Lemma 4.19. *For any $\epsilon' > 0$, if $0 \ll w \ll n$ then for all $i \in [1, k_0)$, l_i and r_i will (be defined and will) complete with probability $> 1-\epsilon'$.*

Proof. We form an upper bound for the probability that l_i will fail to complete (the proof for r_i is essentially identical). Suppose that **Good spacing** holds, and that w is large. As before, it is convenient to define $l_0 = u_0$. Fix $i < k_0$ and let $I_1 = [l_{i+1}, l_i]$ and $I_2 = [l_i, l_{i-1}]$. Let k be the greatest such that, when $1 \leq j \leq k$, $I_1(j:k)$ and $I_2(j:k)$ are of length $\geq w+1$. For $1 \leq j \leq \lfloor k/2 \rfloor$ define:

$$J_j := I_1(j:k) \cup I_1(j:k)^- \cup I_2(j:k) \cup I_2(j:k)^-.$$

[14] Note that π is defined differently here than in the proof of Lemma 3.3.

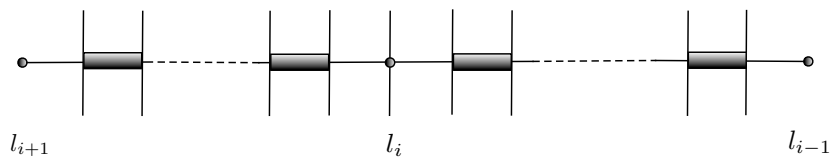

FIGURE 8. $J_2 = I_1(2:k) \cup I_1(2:k)^- \cup I_2(2:k) \cup I_2(2:k)^-$.

Figure 8 below shows, as an example, what J_2 looks like:

For $1 \leq j \leq \lfloor k/2 \rfloor$, let P_j be the event that a node in J_j changes type, and note that P_3 cannot occur until P_2 has occurred, P_4 cannot occur until P_3 has occurred, and so on. This follows simply because nodes in $[l_{i+1}, l_{i-1}]$ but outside the neighbourhoods $\mathcal{N}(l_{i+1})$, $\mathcal{N}(l_i)$ and $\mathcal{N}(l_{i-1})$ are initially happy, and cannot become unhappy until another node in their neighbourhood changes type. Now the basic idea is that if completion fails to occur, then the sequence of events $P_2, ..., P_{\lfloor k/2 \rfloor}$ must occur before any stage at which there are no unhappy nodes in the neighbourhood of l_i.

We label certain stages as being a 'step towards completion', and certain others as being a 'step towards failure of completion' (while some stages are labelled as neither). These labels do not fully represent all aspects of the process, but contain enough information for our purposes.

Steps towards completion. So long as there are unhappy nodes in the neighbourhood of l_i, we label any stage at which a node in this neighbourhood swaps type as a *step towards completion*. Subsequent to any stage at which l_i completes, we label every stage as a step towards completion.

Steps towards failure of completion. If $1 \leq j < \lfloor k/2 \rfloor$ is the greatest such that P_j has occurred prior to stage s or no P_j has occurred and $j = 1$, and if P_{j+1} occurs at stage s, then we label s a *step towards failure of completion*.

Now at any stage s at which some P_j for $j \leq \lfloor k/2 \rfloor$ is yet to occur, and at which there are unhappy nodes in the neighbourhood of l_i, the probability of s being a step towards failure of completion, is at most $4(w+1)$ times the probability of it being a step towards completion (since there are at most $4(w+1)$ times as many nodes which, if chosen to swap, will cause a step towards failure of completion, as those which will cause a step towards completion). Choosing $d' > d$ we get that, since **Good spacing** holds, for all sufficiently large w, $\lfloor k/2 \rfloor > e^{w/d'}$. We may therefore consider the first $e^{w/d'}$ many stages which are steps either towards completion or failure of completion[15] and, for large w, consider the probability that at most $2w$ of these are steps towards completion. This probability is less than the probability that a random variable with binomial distribution $\mathrm{Bin}(e^{w/d'}, 1/(4w+1))$ takes a value of at most $2w$. By the weak law of large numbers, this probability tends to 0 as $w \to \infty$. Now if there do not exist unhappy nodes of both types in the interval $[l_i - 2w, l_i]$ in the initial configuration (which is the case with probability tending to 1 as $w \to \infty$, since it holds if $\texttt{Smooth}^*(l_i, k', \epsilon'')$ holds for large k' and small ϵ''), then $2w+1$ many steps towards completion prior to $P_{\lfloor k/2 \rfloor}$ occurring, suffices to ensure completion for l_i. □

Lemma 4.24. *Suppose that all events in Π hold. Let i be the least such that l_i is defined and originates a firewall, and let j be least such that r_j is defined and originates a firewall. For any $\epsilon' > 0$, for all sufficiently large w, with probability $> 1 - \epsilon'$, u_0 will eventually be contained in one of the two firewalls originated at l_i and r_j.*

Proof. If the firewalls originated at l_i and r_j are of the same type, α say, then the result is immediate – by Lemma 4.21 all $l_{i'}$ for $1 \leq i' < i$ and all $r_{j'}$ for $1 \leq j' < j$ subside, meaning that no β-stable intervals can ever be created in $[l_i, r_j]$. So suppose that the firewall originated at l_i is of type α and the firewall originated at r_j is of type β.

As we describe the argument we initially state two facts ($\dagger_1$) and ($\dagger_2$) without proof. Once the outline of the argument is complete, we then provide proofs for these facts. First, note that there are certain type changes within the interval $[l_i, r_j]$ which we are not presently concerned with. If u is in the neighbourhood of some x which is $l_{i'}$ for $1 \leq i' < i$ or

[15] In order to ensure the existence of $e^{w/d'}$ many such stages it is momentarily convenient to adopt the convention that the process continues after there are no unhappy nodes $u \in [1, n]$, but with nothing occurring at such stages, and that such stages are also labelled steps towards completion.

$r_{j'}$ for $1 \leq j' < j$, and changes type at a stage which is less than or equal to that at which x completes, then we say that this change of type is *previous*. Ignoring changes which are previous, it is then easy to formalise the first stage $s(I)$ at which either of the two firewalls originated at l_i or r_j 'have influence' on any given subinterval I of $[l_i, r_j]$: $s(I)$ is the first stage at which any node in I has a change of type which is not previous. The fact that all events in Π are satisfied means that $s(I)$ must be defined for any subinterval I of $[l_i, r_j]$.

For $\gamma \in \{\alpha, \beta\}$ we also define $s^\gamma(I)$ to be the first stage at which a node in I has a change of type to γ which is not previous – note that, unlike $s(I)$, these values may be undefined (we write $s^\gamma(I) \downarrow$ to indicate that $s^\gamma(I)$ is defined). Now let $u_1 = u_0 - (2w+1)$ and let $u_2 = u_0 + (2w+1)$. Let $I_0 = \mathcal{N}(u_0)$, let I_1 be the left part of $\mathcal{N}(u_1)$, i.e. $I_1 = [u_1 - w, u_1]$ and let I_2 be the right part of $\mathcal{N}(u_2)$.

($\dagger_1$) For any $\epsilon' > 0$, for all sufficiently large w, the probability that there does not exist $\gamma \in \{\alpha, \beta\}$ such that both $s^\gamma(I_1) \downarrow = s(I_1)$ and $s^\gamma(I_2) \downarrow = s(I_2)$, is $\ll \epsilon'$.

So suppose that there does exist such γ, and suppose $\gamma = \beta$ (the case $\gamma = \alpha$ is similar).

($\dagger_2$) Let $v \in [l_i, r_j]$ be the rightmost node such that either $v = l_i$ or $s^\alpha(\mathcal{N}(v)) \downarrow < s(I_1)$. For any $\epsilon' > 0$, for all sufficiently large w, the probability that $|u_0 - v| < e^{w/d}$ is $\ll \epsilon'$.

Let us suppose that $|u_0 - v| \geq e^{w/d}$. Now consider any node $u \in I_0$. Let I_3 be the leftmost w many nodes in $\mathcal{N}(u)$, and let I_4 be the rightmost w many nodes. Let $\delta = 1 - 2\tau$, so that if any node x has bias 0 in the initial configuration then $w\delta$ many type changes in $\mathcal{N}(x)$, all to type β suffice to give x almost exactly borderline bias (less than borderline bias plus 3 to be more precise). By the weak law of large numbers, for any $\epsilon' > 0$, for all sufficiently large w, the probability that any node in $[u_1 - w, u_2 + w]$ has bias in the initial configuration which is of modulus $> w\frac{1}{3}\delta$ is $\ll \epsilon'$. Suppose that this is not the case. Now no node in I_3 can become unhappy until there have been $> \frac{2}{3}\delta w$ many changes to β type in I_4 (with room to spare). Similarly no node in I_1 can become unhappy until there have been at least $\frac{2}{3}\delta w$ many changes to β type in I_3. So at stage $s(I_1)$ we conclude that there have been at least $\frac{4}{3}\delta w$ many changes to β type in the neighbourhood of u. This means that u now has bias at most $w\frac{1}{3}\delta - w\frac{8}{3}\delta$. We conclude that *all* α nodes u in I_0 must be unhappy at stage $s(I_1)$. We can now define a notion of completion for u_0. We say that u_0 completes at stage $s > s(I_1)$ if:

(1) No node in I_0 is unhappy at stage s, and this is not true for any $s' < s$ with $s' > s(I_1)$.

(2) Letting v be defined as in ($\dagger_2$), there exists x_0 with $v + 2w < x_0 < u_0 - 2w$, such that by the end of stage s, no node in $[x_0 - w, x_0]$ has had a change of type which is not previous.

We can then argue, precisely as in Lemma 4.19, that u_0 almost certainly completes. In this case it then follows that u_0 ultimately belongs to a firewall, which includes the interval $[u_0, r_j]$. To complete the proof, we are therefore left to verify ($\dagger_1$) and ($\dagger_2$).

We verify ($\dagger_1$) in such a way that almost precisely the same proof suffices to verify ($\dagger_2$) also. The very basic idea is that, given the large distance between l_i and r_j, it is very unlikely that the influence of these two firewalls will first meet sufficiently close to u_0 as to cause us trouble. Given $\epsilon' > 0$, choose $k \gg \frac{1}{\epsilon'}$. For all sufficiently large w, it follows from our earlier analysis that, in fact, the probability that $|u_0 - l_1| < (k+1)\lceil e^{w/d} \rceil$ or $|u_0 - r_1| < (k+1)\lceil e^{w/d} \rceil$ is $\ll \epsilon'$. So suppose that neither of these possibilities hold. Now, for each instance $\mathcal{Z}$ of the process on the circle of n nodes for which there does not exist $\gamma \in \{\alpha, \beta\}$ such that both $s^\gamma(I_1) \downarrow = s(I_1)$ and $s^\gamma(I_2) \downarrow = s(I_2)$, we wish to show that there are k-many distinct others which we can label as *corresponding* to $\mathcal{Z}$, each of which occurs with the same probability as $\mathcal{Z}$ and for which there *does* exist γ of the kind described (and for which all events in Π hold for the same u_0). We also require that, for distinct $\mathcal{Z}$ and $\mathcal{Z}'$ for which γ as required does not exist, the two corresponding sets of processes have no intersection. The k-many distinct processes corresponding to $\mathcal{Z}$ are easily defined. Given that γ does not exist, we must have that $s(I_1) = s^\alpha(I_1)$ and that $s(I_2) = s^\beta(I_2)$. Note also that (since the 'influence' of each of the firewalls originated at l_i and r_j has to spread by one interval of length at most w at a time) we then have that:

($*$) For all u to the left of I_1 in $[l_1, r_1]$, $s^\alpha(\mathcal{N}(u)) \downarrow = s(\mathcal{N}(u))$. Similarly, for all u to the right of I_2 in $[l_1, r_1]$, $s^\beta(\mathcal{N}(u)) \downarrow = s(\mathcal{N}(u))$.

The processes corresponding to $\mathcal{Z}$ are the 'rotations' of the entire process $\mathcal{Z}$ by $m\lceil e^{w/d}\rceil$ many nodes to the right, for $1 \leq m \leq k$. If $s(I_1) = s^{\alpha}(I_1)$ and $s(I_2) = s^{\beta}(I_2)$, then it follows from $(*)$ that this does not hold for the specified rotations. We therefore conclude that the probability that γ does not exist as required, is $< \frac{1}{k} \ll \epsilon'$.

As remarked above, an almost identical proof then suffices to establish $(\dagger_2)$. □

8.4. **Dealing with the standard model.** As noted in Section 4, our aim is to establish the following:

($\diamond$) Suppose w, τ (with $\kappa < \tau < \frac{1}{2}$) and $\epsilon > 0$ are fixed (where ϵ is the value we have fixed throughout this section, and which played a role in the definition of k_0). Let l_{k_0} and r_{k_0} be as defined previously. For any $\epsilon' > 0$, when n is sufficiently large the following holds with probability $> 1 - \epsilon'$: there exists a first stage at which there are no unhappy nodes in the interval $[l_{k_0}, r_{k_0}]$ and at all stages up to this the total number of unhappy α nodes divided by the total number of unhappy β nodes lies in the interval $[1 - \epsilon', 1 + \epsilon']$.

The very basic idea is as follows. For the remainder of the section we suppose that w, τ and ϵ are fixed. With probability close to 1, for large n we will have in the initial configuration that the proportion of nodes which are unhappy and of type α is roughly equal to the proportion of nodes which are unhappy and of type β. Briefly, however, let us make the simplification that these proportions will be *exactly* equal. We then want to show that the process can be sufficiently accurately modelled, for large n, by a system of differential equations, which are entirely symmetric in α and β. This symmetry means that the solution to the system of differential equations must describe an evolution in which there are always precisely equal numbers of unhappy α and β nodes. Of course, we then have to deal with the fact that the numbers of unhappy α and β nodes in the initial configuration need not actually be exactly equal, but this turns out not to present too many problems.

As discussed in [5], there are, however, some further complications which arise immediately as one looks to apply the Wormald machinery. The method applies to a process in which the state of the system at any given moment in time is described by an ℓ-dimensional vector of real numbers, where ℓ remains fixed, and we then look to approximate the discrete process by a continuous one as $n \to \infty$. How can we describe the configuration at any given stage by an ℓ-dimensional vector, where ℓ is independent from n? Let C_n be the graph which is a cycle of size n. Up until now, then, we have been considering processes unfolding on each C_n. From now on we consider also a value L which depends on w, but not on n. Generally we shall work under the assumption that $w \ll L \ll n$. For the sake of simplicity we assume that L divides n (but everything that follows is easily modified to deal with the possibility that this is not the case). As we consider the process unfolding on C_n, we consider also a parallel process on G_n, which is a disjoint union of cycles of length L. More precisely, nodes u and v are connected in G_n iff $\lfloor u/L \rfloor = \lfloor v/L \rfloor$ and $u \equiv v \pm 1 \bmod L$. In order to consider a parallel process on G_n, it is also convenient to modify the way in which we count the stages of the process. In the process as previously described, an unhappy pair of nodes of opposite type are selected at each stage, which may then swap (in fact will swap for the values of τ considered here). Since we shall now have a situation in which the same node u may be unhappy in C_n but happy in G_n, or vice versa, it becomes convenient to consider a process in which two nodes are selected uniformly at random at each stage, which will only then swap if they are of opposite type and both are unhappy. Of course this makes no real difference to the evolution of the system, except for the way in which we count the stages. The parallel process on G_n then unfolds as follows: when u and v are selected for a potential swap in C_n, they swap in G_n if they are both of opposite type and are unhappy in G_n. Now let us see how the configuration of G_n at any stage can be described by a 2^L-dimensional vector. We let 2^L denote the set of binary strings of length L. For each node u and each stage s we define a string $\tau_{u,s} \in 2^L$: $\tau_{u,s}(0) = 1$ if u is of type α at stage s, otherwise $\tau_{u,s}(0) = 0$, and then $\tau_{u,s}(1) = 1$ if the node to the right of u in G_n is of type α at stage s, and so on. For each $\sigma \in 2^L$ we define $\zeta_\sigma(s)$ to be the number of nodes u such that $\tau_{u,s} = \sigma$. Then the 2^L-dimensional vector $\boldsymbol{\zeta}(s)$ in which the components are the values $\zeta_\sigma(s)$ for $\sigma \in 2^L$ describes the configuration at stage s up to isomorphism.

Our first task now is to show that for large L and large n compared to L, the processes on G_n and C_n do not diverge too quickly. To this end we inductively define a set of *tainted* nodes for each stage s, denoted $T(s)$. These are nodes whose neighbourhoods might possibly look different in G_n and C_n. $T(0)$ is the set of u such that $-w \leq u \bmod L < w$ (we assume that the swapping process begins at stage $s = 1$). If u and v are chosen for a potential swap at stage $s > 0$ and are both untainted, then $T(s) = T(s-1)$. Otherwise $T(s)$ is the union of $T(s-1)$ with the set of all nodes which

are in the neighbourhood of u or v in either G_n or C_n. Immediately it is clear that $|T(s) - T(s-1)| < 4(2w+1)$. In fact, by counting more carefully those nodes which belong to the neighbourhoods of u and v in *both* graphs, we get $|T(s) - T|(s-1)| < 2(3w+1)$, so that, since w is large:

$$|T(s) - T(s-1)| < 7w. \tag{8.4.1}$$

The next lemma provides precisely the kind of probabilistic bound on the number of tainted nodes at any given stage which we will need later.

Lemma 8.5. *The following conditions hold at every stage $s \geq 0$:*

(1) The expected number of tainted nodes at stage s is bounded above by $e^{14ws/n}\left(\frac{2w-1}{L}\right)n$.

(2) The probability that $\frac{|T(s)|}{n} > 2e^{14ws/n}\left(\frac{2w-1}{L}\right)$ is at most $e^{-wn/(21L^2)}$.

Proof. Let $\rho(s) = |T(s)|$. In order to prove (1) first, let u and v be the nodes chosen for a swap at stage $s+1$. Let γ be the probability that either of these nodes are tainted (at the end of stage s). The probability that u is tainted is $\rho(s)/n$, and similarly for v, so $\gamma \leq 2\rho(s)/n$. Now if neither of u or v is tainted then $\rho(s+1) = \rho(s)$, and otherwise, by (8.4.1), $\rho(s+1) < \rho(s) + 7w$. We therefore have:

$$\mathbf{E}[\rho(s+1) \mid \rho(s)] < (1-\gamma)\rho(s) + \gamma(\rho(s) + 7w) \leq \rho(s) + 2\frac{\rho(s)}{n}7w = \rho(s) \cdot \left(1 + \frac{14w}{n}\right).$$

For $x > 0$, $1 + x < e^x$, giving:

$$\mathbf{E}[\rho(s+1) \mid \rho(s)] < \rho(s) \cdot e^{14w/n}.$$

So far, then, we have established that the sequence of random variables $Y_s = \rho(s)e^{-14ws/n}$ is a supermartingale. This suffices to give (1) since $Y_0 = \left(\frac{2w-1}{L}\right)n$.

In order to establish (2) we make use of the following version of Azuma's inequality: If $X_1, ..., X_s$ is a supermartingale with $X_0 = 0$ and $|X_i - X_{i-1}| \leq c_i$ for all i and constants c_i, then for all $\Omega > 0$,

$$\mathbf{P}(X_s \geq \Omega) \leq \exp\left(-\frac{\Omega^2}{2\sum_{i=1}^{s} c_i^2}\right). \tag{8.4.2}$$

If we put $X_s = \frac{Y_s}{n} - \frac{2w-1}{L}$ then the sequence $X_0, X_1, ...$ is a supermartingle and $X_0 = 0$. In order to apply (8.4.2) we need a bound on the values $|X_i - X_{i-1}|$. To this end we consider the sum:

$$|Y_s - e^{-14w/n}Y_s| + |Y_{s+1} - e^{-14w/n}Y_s|.$$

Now

$$Y_s - e^{-14w/n}Y_s = Y_s(1 - e^{-14w/n}) < Y_s 14w/n,$$

since for $x > 0$, $1 - e^{-x} < x$. Also,

$$Y_{s+1} - e^{-14w/n}Y_s = e^{-14(s+1)w/n}(\rho(s+1) - \rho(s)) \leq 7we^{-14(s+1)w/n},$$

by (8.4.1). Thus,

$$|Y_s - e^{-14w/n}Y_s| + |Y_{s+1} - e^{-14w/n}Y_s| < 14wY_s/n + 7we^{-14w(s+1)/n} < 21we^{-14ws/n}.$$

We may therefore put $c_s = \frac{21}{n}we^{-14ws/n}$. Now since for $0 < x < 1$, $\frac{1}{1-x} = 1 + x + x^2 + x^3 \cdots$ we conclude that $\sum_{s=0}^{\infty} c_s^2 = (21w/n)^2(1 - e^{-28w/n})^{-1} < 21w/n$, the latter inequality following from the fact that for small positive x, $1 - e^{-x} > 3x/4$.

Finally, applying (8.4.2), we get:

$$\begin{aligned}
\mathbf{P}\left(\frac{\rho(s)}{n} > 2e^{14ws/n}\left(\frac{2w-1}{L}\right)\right) &= \mathbf{P}\left(\frac{Y_s e^{14ws/n}}{n} > 2e^{14ws/n}\left(\frac{2w-1}{L}\right)\right) \\
&= \mathbf{P}\left(X_s + \left(\frac{2w-1}{L}\right) > 2\left(\frac{2w-1}{L}\right)\right) \\
&= \mathbf{P}\left(X_s > \left(\frac{2w-1}{L}\right)\right) \\
&\leq \exp\left(-\frac{n(2w-1)^2}{42wL^2}\right) \\
&< \exp\left(-\frac{nw}{21L^2}\right).
\end{aligned}$$

This establishes (2) as required. □

So Lemma 8.5 tells us that for fixed x, we can ensure with probability close to 1 that the proportion of nodes which are tainted by stage xn is very small, so long as we take L large and n large compared to L. With this in place, we now concentrate for a while on the processes on the graphs G_n. In order to approximate these processes by the solution to a system of differential equations, we first of all have to draw up reasonable candidates for the differential equations to be used. To this end we begin by considering the conditional expectations $\mathbf{E}[\zeta_\sigma(s+1) \mid \boldsymbol{\zeta}(s)]$. One can form an expression for this expectation, simply by listing all of the different ways in which the number of nodes u with $\tau_u = \sigma$ can increase or decrease at a given stage and establishing their probabilities (we formally defined $\tau_{u,s}$ rather than τ_u, but shall omit s where no ambiguity results). Consider a proposed swap between u and v which belong to different cycles, such that $\tau_u = \sigma'$ and $\tau_v = \sigma''$. Define $a(\sigma, \sigma', \sigma'')$ to be the net change resulting from this proposed swap, in the number of nodes u' for which $\tau_{u'} = \sigma$. Note that the net change will be zero unless u and v are of opposite type and are both unhappy, and that in any case $|a(\sigma, \sigma', \sigma'')| < 2L$. Then:

$$\mathbf{E}[\zeta_\sigma(s+1) - \zeta_\sigma(s) \mid \boldsymbol{\zeta}(s)] = \sum_{\sigma',\sigma''} a(\sigma, \sigma', \sigma'')\frac{\zeta_{\sigma'}(s)\zeta_{\sigma''}(s)}{n^2} + O\left(\frac{L^2}{n}\right). \tag{8.4.3}$$

The $O\left(\frac{L^2}{n}\right)$ correction term accounts for the possibility that the nodes selected for a potential swap might belong to the same cycle, the probability of this occurring being L/n and the net change then being bounded by L. We next perform some scaling, so that the variables reach fixed functions in the limit as $n \to \infty$: for each σ we want a real function $z_\sigma(s)$ to model the behaviour of $\frac{1}{n}\zeta_\sigma(sn)$. Now (8.4.3) *suggests* the following system of differential equations for the functions z_σ:

$$z'_\sigma(s) = \sum_{\sigma',\sigma''} a(\sigma, \sigma', \sigma'')z_{\sigma'}(s)z_{\sigma''}(s). \tag{8.4.4}$$

$$z_\sigma(0) = \frac{1}{n}\zeta_\sigma(0). \tag{8.4.5}$$

Note that for any solution to (8.4.4) and (8.4.5), if we put $y(s) = \sum_\sigma z_\sigma(s)$, then:

$$y'(s) = \sum_\sigma z'_\sigma(s) = \sum_{\sigma',\sigma''}\sum_\sigma a(\sigma, \sigma', \sigma'')z_{\sigma'}(s)z_{\sigma''}(s) = \sum_{\sigma',\sigma''} 0 = 0.$$

Thus, since $y(0) = 1$, we also have $y(s) = 1$ for all s in the domain. This in turn means that :

$$|z'_\sigma(s)| = |\sum_{\sigma'}\sum_{\sigma''} a(\sigma, \sigma', \sigma'')z_{\sigma'}(s)z_{\sigma''}(s)| \leq |\sum_{\sigma'}\sum_{\sigma''} 2Lz_{\sigma'}(s)z_{\sigma''}(s)| = |\sum_{\sigma'} 2Lz_{\sigma'}(s)| = 2L \tag{8.4.6}$$

Suppose for now that we are interested in the first xn steps of the process on G_n. Later we shall choose a value for x which suits our needs. For now though, this means that we are interested in solutions to (8.4.4) and (8.4.5) for $s \in [0, x]$. Let Π be the set of points $(s, z_1, ..., z_{2^L})$ such that $s \in [0, x]$ and, for each $i \leq 2^L$, $|z_i| \leq 2Lx + 1$. Let D be a bounded connected open subset of $\mathbb{R}^{2^L+1}$ containing Π and such that the minimum ℓ^∞ distance between the boundary of D and any point in Π is bounded below (and where the ℓ^∞ distance between two vectors is the maximum difference between corresponding components). For the remainder of this section, let all the strings in 2^L be enumerated as $\sigma_1, \sigma_2, ..., \sigma_{2^L}$. Then by standard results in the theory of differential equations (see [14] for example), the fact that each of the functions

$$f_i(s, z_1, ..., z_{2^L}) := \sum_{j,k} a(\sigma_i, \sigma_j, \sigma_k) z_j z_k$$

satisfies a Lipschitz condition for each argument z_r in D, together with (8.4.6), means that there exists a unique set of functions $\{z_{\sigma_i}(s) \mid \sigma_i \in 2^L\}$ defined on the interval $[0, x]$ and which satisfy (8.4.4) and (8.4.5) for all $s \in [0, x]$.

The following lemma is just a slightly stripped down version of Theorem 5.1 from [29].

Lemma 8.6. *Let D and x be as above. Suppose that for all n, $\lambda(n) > L^2/n$ and that as $n \to \infty$, $\lambda(n) \to 0$. Consider the unique set of functions $\{z_\sigma \mid \sigma \in 2^L\}$ defined on the interval $[0, x]$ which satisfy (8.4.5) and (8.4.4) for all $s \in [0, x]$. Then the following holds with probability $1 - O(\frac{2^{L+1}L}{\lambda(n)} \exp(-\frac{n\lambda(n)^3}{8L^3}))$:*

$$\zeta_\sigma(s) = n z_\sigma(s/n) + O(n\lambda(n))$$

uniformly for $0 \leq s \leq xn$.

Actually, in the statement Theorem 5.1 in [29], the conclusion is only guaranteed for those s for which the vector $(s, z_{\sigma_1}(s), \ldots, z_{\sigma_{2^L}}(s))$ is within ℓ^∞ distance $C\lambda(n)$ of the boundary of D for some sufficiently large constant C (where C is independent from n). In our case, however, (8.4.6) together with the way in which we defined D removes these complications. Since we required that the minimum ℓ^∞ distance between D and Π be bounded below, and since $\lambda(n) \to 0$ as $n \to \infty$, the required condition is automatically satisfied for all sufficiently large n and all $s \in [0, x]$ (and then the constant in the $O(\frac{2^{L+1}L}{\lambda(n)} \exp(-\frac{n\lambda(n)^3}{8L^3}))$ term can be modified to ensure the theorem holds for all n).

Consider now the symmetry properties which must be satisfied by the functions z_σ. For each $\sigma \in 2^L$ let $\bar{\sigma}$ be the string which results from changing every 1 in σ to 0, and changing every 0 to 1. For $\boldsymbol{z} \in \mathbb{R}^{2^L}$, let $\iota(\boldsymbol{z})$ be the vector which results from swapping the ith component with the jth component whenever $\bar{\sigma}_i = \sigma_j$ (recalling our enumeration of the strings of length 2^L from before). If we have $\iota(\boldsymbol{\zeta}(0)) = \boldsymbol{\zeta}(0)$ (which we probably will not, but for large n we shall have something close to it with good probability), then the symmetry properties of (8.4.4) guarantee that for all $s \in [0, x]$, $\iota(\boldsymbol{\zeta}(s)) = \boldsymbol{\zeta}(s)$. Define $u(\sigma) = 1$ if any node u such that $\tau_u = \sigma$ is an unhappy α node, $u(\sigma) = -1$ if any node u such that $\tau_u = \sigma$ is an unhappy β node, and $u(\sigma) = 0$ otherwise. For $\boldsymbol{z} = (z_1, ..., z_{2^L}) \in \mathbb{R}^{2^L}$ consider the linear function:

$$\Delta(\boldsymbol{z}) = \sum_i u(\sigma_i) z_i.$$

Then $\Delta(\boldsymbol{\zeta}(s)/n)$ is the difference at stage s, between the fraction of nodes which are unhappy and of type α, and the fraction which are unhappy and of type β. Let $\boldsymbol{z}(s)$ be the vector solution to (8.4.4) and (8.4.5), i.e. with ith component $z_{\sigma_i}(s)$. If $\iota(\boldsymbol{\zeta}(0)) = \boldsymbol{\zeta}(0)$ then we have that $\Delta(\boldsymbol{z}(s)) = 0$ for all $s \in [0, x]$.

In the above we considered the case that $\iota(\boldsymbol{\zeta}(0)) = \boldsymbol{\zeta}(0)$. We now have to deal with the fact that this will probably not hold exactly. Recall that we are interested in the processes on C_n and G_n up stage xn for some fixed x, which we are yet to specify, but which will not depend on n or L. For now suppose that we want to ensure the following modified version of our ultimate aim ($\diamond$) (one of the differences being that this modified statement refers to a difference in numbers rather than a ratio):

(†) For fixed $x > 0$ and any $\epsilon_1 > 0$ the following holds for the process on C_n, with probability $> 1 - \epsilon_1$ for all sufficiently large n: at all stages $\leq xn$ the difference between the fraction of nodes which are unhappy and of type α and the fraction which are unhappy and of type β, is less than ϵ_1.

To establish (†) take $\epsilon_2 < \epsilon_1$. It suffices to ensure that with probability $> 1 - \epsilon_2$, $|\Delta(\boldsymbol{z}(s))| < \epsilon_2$ for all $s \in [0, x]$, since then we can take L large and n sufficiently large compared to L and apply Lemmas 8.5 and 8.6 (putting $\lambda(n) = \max\{\frac{L^2+1}{n}, n^{-1/4}\}$, for example) to get the required result for C_n.

Let S denote the set of vectors $\boldsymbol{y} = (y_1, ..., y_{2^L})$ such that the solution to (8.4.4) with $\boldsymbol{z}(0) = \boldsymbol{y}$, satisfies $|\Delta(\boldsymbol{z}(s))| < \epsilon_2$ for all $s \in [0, x]$. Let $\Sigma = \{\boldsymbol{y} \mid \iota(\boldsymbol{y}) = \boldsymbol{y}, \forall i, y_i \geq 0, \sum_i y_i = 1\}$. By the continuity properties of ordinary differential equations (see for example [6]) it follows that S is an open set, so since Σ is compact there exists d_0 such that every point within distance d_0 of Σ belongs to S. Then it follows directly from the law of large numbers that when n is sufficiently large compared to L, $\boldsymbol{\zeta}(0)/n$ is within distance d_0 of Σ with probability $> 1-\epsilon_2$. This establishes (†) as required.

Finally, we have to specify the value x. Recall that our aim is to prove ($\diamond$), as specified previously. Let l_1 be the length of the interval $[l_{k_0}, r_{k_0}]$. Note that (for fixed w and ϵ, where ϵ is the value for which we are proving Theorem 1.2, and which plays a role in the value of k_0, and for ϵ' as in ($\diamond$)), we can take l_2 such that for all sufficiently large n, the probability that either $|l_{k_0} - u_0| > l_2/2$ or $|u_0 - r_{k_0}| > l_2/2$ is $\ll \epsilon'$. Taking $l_3 \gg l_2/\epsilon'$, consider now the first stage s_1 at which, for some $\gamma \in \{\alpha, \beta\}$, the number of unhappy γ nodes as a proportion of n is less than $1/l_3$. Taking $\epsilon'' \ll \epsilon'$ and putting $x > (2w+1)(l_3)^2/\epsilon''$, suppose towards a contradiction that $\mathbf{P}(s_1 > xn) > \epsilon''$. In that case, with probability $> \epsilon''$, the following holds at each stage $1 \leq s \leq xn$: given the configuration at the end of stage $s-1$, the probability at stage s of choosing two nodes which do swap is at least $\frac{1}{(l_3)^2}$. The expected number of stages $\leq xn$ at which swaps do occur is therefore $> (2w+1)n$, which gives the required contradiction since, according to the observation made in Section 2, there can be at most $(2w+1)n$ stages at which a swap occurs. Applying (†) to $\epsilon_1 \ll \epsilon'/l_3$ then establishes that with probability approaching 1 as $n \to \infty$, $s_1 < xn$ and at all stages $\leq xn$ the difference between the fraction of nodes which are unhappy and of type α and the fraction which are unhappy and of type β, is less than ϵ_1. When the latter conditions hold, this means that at all stages up to s_1 the ratio between the number of unhappy α nodes and the number of unhappy β nodes lies in the interval $[1-\epsilon', 1+\epsilon']$. Now choosing $l_3 \gg l_2/\epsilon'$ as we did previously, means that, since u_0 was chosen uniformly at random, and given that the total proportion of nodes which are unhappy at stage s_1 is at most $\frac{2+\epsilon'}{l_3}$, the probability that there are any unhappy nodes in the interval $[u_0 - \lceil l_2/2 \rceil, u_0 + \lceil l_2/2 \rceil]$ (or in the almost certainly smaller interval $[l_{k_0}, r_{k_0}]$) at stage s_1 is $\ll \epsilon'$. Thus we have established ($\diamond$), as required.

9. Proofs deferred from Section 6

We must prove Lemma 6.1. Initially, however, we ignore the required conditions concerning intervals of length $4w+1$, and we simply look to establish that when x_p is close to $\frac{1}{2}$ there are unhappy α nodes in any configuration. Rather than considering the nodes to be arranged in a circle, it is temporarily useful to suppose instead that they are arranged in a line which extends infinitely far to the right, with positions indexed by natural numbers (starting with 0). So we consider a fixed configuration in which each natural number is either of type α or β. It is convenient to suppose further that the node at position w is of type α, and that all α nodes $\geq w$ are happy. From these assumptions, we will deduce the following bound on the proportion of α nodes:

(†) For any $\epsilon > 0$ there exists ℓ such that for all $\ell' \geq \ell$, the proportion of α nodes in the interval $[0, \ell']$ is $\geq \frac{w+1}{2w} - \epsilon$.

This suffices because the existence of any circle in which all α nodes are happy and in which $0 < x_p < \frac{w+1}{2w}$ (the latter condition on x_p being satisfied almost certainly for large enough n, by the weak law of large numbers) would give a contradiction, since this circle could then be cut and infinitely many copies placed end to end, giving a counter example to (†).

In order to prove (†), it will be useful to consider the right and left parts of any neighbourhood $\mathcal{N}(u)$ for $u \geq w$. We define $\mathcal{L}(u)$ (respectively $\mathcal{R}(u)$) to be the leftmost (rightmost) w-many nodes in $\mathcal{N}(u)$. Note that any α node $u \geq w$ must have α nodes in both $\mathcal{L}(u)$ and $\mathcal{R}(u)$. Also, any β nodes $v > w$ must have α nodes in $\mathcal{L}(v)$ and in $\mathcal{R}(v)$, otherwise the next α node to the left or right (respectively) would not be happy.

We define a sequence of α nodes. Let $u_0 = w$. Given u_k define u_{k+1} to be the rightmost α node in $\mathcal{N}_k := \mathcal{N}(u_k)$. Also define $\mathcal{L}_k := \mathcal{L}(u_k)$ and $\mathcal{R}_k = \mathcal{R}(u_k)$. Now, for $m \geq 1$ define $I_m := \bigcup_{k=0}^{m} \mathcal{N}_k$ and $S_m := \Sigma_{k=0}^{m} \Theta(\mathcal{N}_k)$. Since each u_k is happy, $\Theta(\mathcal{N}_k) \geq 3$, giving $S_m \geq 3(m+1)$. This doesn't immediately tell us the bias on the interval I_m, however, because nodes may have been counted multiple times in forming the sum S_m. We therefore want to consider the way in which the neighbourhoods $\mathcal{N}_k$ overlap. For $k \geq 1$, define $\mathcal{N}'_k := \mathcal{R}_{k-1} \cap \mathcal{L}_{k+1}$ $(= \mathcal{N}_{k-1} \cap \mathcal{N}_{k+1})$. Notice that $u_k \in \mathcal{N}'_k$, but that $u_{k-1} \notin \mathcal{N}'_k$ and $u_{k+1} \notin \mathcal{N}'_k$. We similarly partition each $\mathcal{N}'_k$ into a left and right part. Define $\mathcal{L}'_k := \mathcal{N}'_k \cap \mathcal{L}_k$ and $\mathcal{R}'_k := \mathcal{N}'_k \cap \mathcal{R}_k$. It is immediate that if $\mathcal{R}'_k$ is non-empty then all nodes in it are of type β. Now we partition the nodes in I_m according to the number of times they are counted in forming the sum S_m, i.e. the number of $\mathcal{N}_k$ that they belong to for $k \leq m$. We define J_r^m to be the set of $u \in I_m$ which belong to exactly r distinct neighbourhoods $\mathcal{N}_k$ for $k \leq m$. Notice that $\bigcup_{r\geq 3} J_r^m \subseteq \bigcup_{k=1}^{m-1} \mathcal{N}'_k$. On the other hand, $u_k \in \bigcup_{r\geq 3} J_r^m$ for $1 \leq k \leq m-1$. We now want to assess the size and bias of each of the various J_r^m.

$\boldsymbol{J_1^m}$. We have $J_1^m \subseteq \mathcal{L}(u_0) \cup \mathcal{R}(u_m)$, so $|J_1^m| \leq 2w$.

$\boldsymbol{J_4^m}$. We show that only β nodes can belong to J_4^m. Suppose that $u \in J_4^m$ is of type α. Then $u \in \mathcal{N}'_k$ for some k with $1 \leq k \leq m-1$, and by the remark above $u \notin \mathcal{R}'_k$. So $u = u_k$ or $u \in \mathcal{L}'_k$. In either case, $u \in \mathcal{N}_{k-2}$ will contradict the definition of u_{k-1}, while $u \in \mathcal{N}_{k+2}$ will contradict the definition of u_{k+1}, giving the required contradiction.

$\boldsymbol{J_r^m}$, $r \geq 5$. Suppose $u \in \mathcal{N}_{k-2} \cap \mathcal{N}_{k-1} \cap \mathcal{N}_k \cap \mathcal{N}_{k+1} \cap \mathcal{N}_{k+2}$. Then either $u_k \in \mathcal{N}_{k-2}$ contradicting the definition of u_{k-1} or $u_k \in \mathcal{N}_{k+2}$ contradicting that of u_{k+1}. Thus J_r^m is empty for $r \geq 5$.

$\boldsymbol{J_3^m}$. We claim that $\Theta(J_3^m) < m + w$. From our analysis of J_r^m for $r > 3$ it follows that we must have $u_k \in J_3^m$ for $1 \leq k \leq m-1$. So $\Theta(J_3^m) \leq m - 1 + \Sigma_{k=1}^{m-1}|\mathcal{L}'_k| - \Sigma_{k=1}^{m-1}|\mathcal{R}'_k|$. It therefore suffices to show that $\Sigma_{k=1}^{m-1}(|\mathcal{L}'_k| - |\mathcal{R}'_k|)$ is bounded by w. For $k \geq 1$ define $d_k := u_k - u_{k-1}$. Then $|\mathcal{R}'_k| = w - d_k$ and $|\mathcal{L}'_k| = w - d_{k+1}$. Hence $|\mathcal{L}'_k| = |\mathcal{R}'_{k+1}|$, and $\Sigma_{k=1}^{m-1}(|\mathcal{L}'_k| - |\mathcal{R}'_k|) = |\mathcal{L}'_{m-1}| - |\mathcal{R}'_1| \leq w$, proving the claim.

Putting these facts together, we get that $2\Theta(I_m) > 3m - 2w - m - w = 2m - 3w$, so $\Theta(I_m) > m - \frac{3}{2}w$. For $m > \frac{3}{2}w$ choose a partition, $I_m = \Pi_0 \cup \Pi_1 \cup \Pi_2$, so that $|\Pi_0| = |\Pi_1|$, Π_0 contains only β nodes, and Π_1 and Π_2 contain only α nodes (meaning that $|\Pi_2| > m - \frac{3}{2}w$). Then since $|I_m| \leq 2w + 1 + mw$, the proportion of the nodes in I_m which belong to Π_2 is greater than

$$\frac{m - \frac{3}{2}w}{2w + 1 + mw} > \frac{m}{m+3} \cdot \frac{1}{w} - \frac{3}{2m+6}$$

which tends to $\frac{1}{w}$ as $m \to \infty$. Thus, for each $\epsilon > 0$, there exists m such that for all $m' \geq m$, the proportion of α nodes in $I_{m'}$ is $\geq \frac{w+1}{2w} - \epsilon$, giving (†), as required.

Finally, we have to deal with the fact that the statement of the lemma actually requires the existence of unhappy nodes outside any interval of length $4w+1$. We can get this at the expense of assuming n to be reasonably large compared to w. Given any circle configuration in which all α nodes outside an interval of length $4w+1$ are happy, cut the circle at the left end of the interval of length $4w+1$ and consider the individuals in the circle to lie in a line with sites indexed by natural numbers $< n$ (so those in the interval of length $4w+1$ occupy $[0, 4w]$). Performing an almost identical analysis we still conclude that, for any $\epsilon > 0$, so long as n is sufficiently large, there exists ℓ such that for all $\ell' \geq \ell$ (with $\ell' < n$), the proportion of α nodes in the interval $[0, \ell']$ is $\geq \frac{w+1}{2w} - \epsilon$. We conclude that with probability $1 - \epsilon$ the initial configuration on a circle of size n will have a value x sufficient to ensure unhappy nodes of both types outside any interval of length $4w+1$ in *any* configuration with the same number of α and β nodes, where $\epsilon \to 0$ as $n \to \infty$.

References

[1] G. Barmpalias, R. Elwes and A. Lewis-Pye, Tipping Points in 1-Dimensional Schelling Models with Switching Agents. *Journal of Statistical Physics* (2015) 158:806–852.

[2] G. Barmpalias, R. Elwes and A. Lewis-Pye, Minority population in the one-dimensional Schelling model of segregation, to appear.

[3] G. Barmpalias, R. Elwes and A. Lewis-Pye, Unperturbed Schelling segregation in two and three dimensions, *Journal of Statistical Physics*, 164 (4) (2016), 1460–1487.

[4] R.A. Baughman, Racial Segregation and the Effectiveness of Expanding Public Health Insurance for Children, mimeo, 2004.

[5] C. Brandt, N. Immorlica, G. Kamath, R. Kleinberg, An Analysis of One-Dimensional Schelling Segregation, *Proc. 44th Annual ACM Symposium on Theory of Computing (STOC 2012)*.

[6] F. Brauer and J. A. Nohel, The Qualitative Theory of Ordinary Differential Equations. Dover, 1989.
[7] W. Clark, Residential segregation in American cities: A review and interpretation, *Population Research and Policy Review*, 5(2):95-127, 1986.
[8] W. Clark, and M. Fossett, Understanding the social context of the Schelling segregation model, *Proceedings of the National Academy of Sciences*, col 105, 11, 4109–4114, 2008.
[9] L. Dall'Asta, C. Castellano, M. Marsili, Statistical physics of the Schelling model of segregation, *J. Stat. Mech*, 7, 2008.
[10] R. Durrett, J. Steif, Fixation results for threshold voter systems, *The Annals of Probability*, vol. 21, 232-247, 1993.
[11] R. Farley, C. Steeh, T. Jackson, M. Krysan, and K. Reeves, Continued racial residential segregation in Detroit: "Chocolate city, vanilla suburbs" revisited, *Journal of Housing Research*, 4(1):1-38, 1993.
[12] L. Gauvin, J. Vannemenus, J.-P. Nadal, Phase diagram of a Schelling segregation model, *European Physical Journal B*, 70, 293–304, 2009.
[13] A. D. Henry, P. Prałat, and C. Zhang, Emergence of segregation in evolving social networks. *Proceedings of the National Academy of Sciences*, 108(21):8605-8610, May 2011.
[14] W. Hurewicz, Lectures on Ordinary Differential Equations, M.I.T. Press, Cambridge Massachusetts, 1958.
[15] N. Immorlica, R. Kleinberg, B. Lucier, M. Zadomighaddam, Exponential segregation in a two-dimensional Schelling model with tolerant individuals, ACM-SIAM Symposium on Discrete Algorithms, SODA 2017.
[16] J. Kleinberg, Cascading Behavior in Networks: Algorithmic and Economic Issues, *Algorithmic Game Theory*, N. Nisan, T. Roughgarden, E. Tardos, V. Vazirani, eds., Cambridge University Press, 2007.
[17] B. D. McKay, On Littlewood's estimate for the binomial distribution, *Adv. Appl. Prob.*, 21 (1989) 475-478. Available at http://cs.anu.edu.au/ bdm/papers/littlewood2.pdf.
[18] T.J. Nechyba, School finance, spatial income segregation and the nature of communities, *Journal of Urban Economics*, 54 (1), 61–88, 2003.
[19] G. Ódor, Self-organising, two temperature Ising model describing human segregation, *International journal of modern physics C*, 3, 393–398, 2008.
[20] H. Omidvar, M. Franceschetti, Self-organized segregation on the grid, *PODC 2017 Proceedings of the ACM Symposium on Principles of Distributed Computing*, 401–410, 2017.
[21] R. Pancs and N. Vriend, Schelling's spatial proximity model of segregation revisited, *Journal of Public Economics*, 91(1-2):1-24, 2007.
[22] M. Pollicott, and H. Weiss, The dynamics of Schelling-type segregation models and a non-linear graph Laplacian variational problem, *Adv. Appl. Math.*, 27, 17-40, 2001.
[23] T. C. Schelling. Models of segregation. *The American Economic Review*, 59(2):488–493, 1969.
[24] T. Schelling, *Micromotives and Macrobehavior*, New York, Norton, 1978.
[25] D. Stauffer and S. Solomon, Ising, Schelling and self-organising segregation, *European Physical Journal B*, 57, 473–479, 2007.
[26] A.M. Turing, The Chemical Basis of Morphogenesis , *Phil. Trans. R. Soc. London B*, 237, pp 37-72, 1952.
[27] M. White, Segregation and diversity measures in population distribution, Population index, 52(2):198-221, 1986.
[28] D. Vinković and A. Kirman, A Physical Analogue of the Schelling Model, *Proceedings of the National Academy of Sciences*, no. 51, volume 103, 19261-19265, 2006.
[29] N.C. Wormald, Differential equations for random processes and random graphs, *Annals of Applied Probability*, 5:1217-1235, 1995.
[30] H.P. Young, Individual Strategy and Social Structure: An Evolutionary Theory of Institutions. Princeton, NJ: Princeton University Press, 1998.
[31] J. Zhang, A dynamic model of residential segregation, *Journal of Mathematical Sociology*, 28(3):147-170, 2004.
[32] J. Zhang, Residential segregation in an all-integrationist world, *Journal of Economic Behavior & Organization*, 54(4):533-550, 2004.
[33] J. Zhang, Tipping and residential segregation: A unified Schelling model, *Journal of Regional Science*, 51:167-193, Feb. 2011.