Electronic Journal of Probability

Electron. J. Probab. **23** (2018), no. 126, 1–42.
ISSN: 1083-6489 https://doi.org/10.1214/18-EJP250

The naming game on the complete graph

Eric Foxall*

Abstract

We consider a model of language development, known as the naming game, in which agents invent, share and then select descriptive words for a single object, in such a way as to promote local consensus. When formulated on a finite and connected graph, a global consensus eventually emerges in which all agents use a common unique word. Previous numerical studies of the model on the complete graph with n agents suggest that when no words initially exist, the time to consensus is of order $n^{1/2}$, assuming each agent speaks at a constant rate. We show rigorously that the time to consensus is at least $n^{1/2-o(1)}$, and that it is at most constant times $\log n$ when only two words remain. In order to do so we develop some useful estimates for semimartingales with bounded jumps.

Keywords: interacting particle system; naming game.
AMS MSC 2010: 60K35.
Submitted to EJP on March 1, 2017, final version accepted on November 23, 2018.

1 Introduction

The study of social dynamics from the standpoint of statistical physics is an area which has seen increased attention in recent years [5]. Historically, interacting particle system models of opinion dynamics, such as the voter model, have been of interest to mathematicians and studied in detail. However, new models emerging in the physics literature have yet to be given a fully rigorous mathematical treatment. One of these is a model of language development known as the naming game. This is a simple model of invention, sharing, and selection of words that displays eventual consensus towards a common vocabulary. It has been studied, using numerical simulations and heuristic computations, on lattices [1], the complete graph [3] and some random graphs [6]. As a first effort from the standpoint of probability theory, we study the naming game on the complete graph and give rigorous proof of some scaling relations that were observed numerically in [3].

We first recall the definition of the naming game on a general locally finite undirected graph $G = (V, E)$. Individuals correspond to vertices of the graph, and each individual

*University of Alberta, Edmonton, AB, Canada, E-mail: efoxall@ualberta.ca

speaks to its neighbours at a certain rate. The idea is that individuals are attempting to agree on a word to describe a certain object, for which initially, no descriptive words exist. The interaction rules are as follows.

- Speaker:
  - If the speaker does not know a word to describe the object then she invents a word and speaks it to the listener.
  - On the other hand, if the speaker does know at least one word to describe the object then she selects a word uniformly at random from her vocabulary and speaks it to the listener.
- Listener:
  - If the listener already knows the chosen word, then both speaker and listener delete the remainder of their vocabulary and remember only that word. This is called *agreement*.
  - Otherwise, the listener adds the chosen word to their vocabulary.

Thus there is a mechanism both for the creation of new words, and for deletion and eventual agreement upon a single word. We now make this description rigorous. The process is denoted $(W_t)_{t\geq 0}$ with $W_t : V \to \mathcal{P}_o(V)$ for each $t \geq 0$, where $\mathcal{P}_o(V)$ is the collection of finite subsets of V. Thus, for each vertex $v \in V$, we have a process $W_t(v)$, the *vocabulary* of v, whose state space consists of all finite subsets of the vertex set V and which is defined as

$$W_t(v) = \{w \in V : v \text{ knows the word invented by } w\}.$$

The process evolves as follows: For each $v \in V$, at the times of an independent Poisson process with rate one, v chooses a listener w uniformly at random from the set $\{u : uv \in E\}$; say this occurs at time t.

- If $W_{t^-}(v)$ is empty then v speaks word v to w, so that $W_t(v) = \{v\}$ and $W_t(w) = W_{t^-}(w) \cup \{v\}$.
- If $W_{t^-}(v)$ is non-empty then v chooses a uniform random word u from $W_{t^-}(v)$ and speaks it to w.
  - If $u \in W_{t^-}(w)$ then $W_t(v) = W_t(w) = \{u\}$.
  - If $u \notin W_{t^-}(w)$ then $W_t(v)$ is unchanged and $W_t(w) = W_{t^-}(w) \cup \{u\}$.

If G is connected and finite, then with probability one, the system eventually settles into one of the set of absorbing states

$$\{W_t(v) = \{w\} \text{ for all } v \in V : w \in V\}$$

and we would like to know what happens on the way to this consensus. Let

$$V_t = \bigcup_v W_t(v)$$

denote the set of words in existence at time t. If G is the complete graph on n vertices, i.e.,

$$V = \{1, \dots, n\} \quad \text{and} \quad E = \{\{v, w\} : v, w \in V,\ v \neq w\},$$

numerical studies and heuristic computations [2] indicate three distinct phases.

1. Early phase: V_t rises from 0 to about $n/2$ in about $\frac{1}{2}\log n$ time.

2. Middle phase: V_t remains fairly constant up till about $n^{1/2}$ time.
3. Late phase: V_t falls sharply to 1 within about $n^{1/4}$ time.

Some heuristics give a sense of the early and middle phases:

1. *Early phase:* a vertex creates a new word if it speaks before listening, which has probability $1/2$, so an average of $n/2$ words are created. This phase ends when every vertex has either spoken or listened. Since for each vertex this occurs after exponential time with rate 2, if these times were independent then the early phase would end after the maximum of n exponential(2) random variables, which has expectation $\sum_{i=1}^n 1/(2i) \approx \frac{1}{2}\log n$.

2. *Middle phase:* suppose that, at first, vertices tend to learn only new words. Then, the vocabulary of each vertex grows at rate 1, so $\sum_v |W_t(v)| \approx nt$ and assuming vocabularies are evenly distributed, each of the roughly $n/2$ words in the population is known by about $2t/n$ individuals, so each time a word is spoken, the probability its listener already knows that word is about $2t/n$. Thus, agreements occur at a total rate of about $2t$, so about t^2 agreements occur on the time interval $[0,t]$. In order to achieve a global consensus, at least one agreement must occur at each of the roughly $n/2$ vertices that invent a word, so consensus requires at least $\sqrt{n/2}$ amount of time.

In this article we consider the early and middle phases, and the tail end of the late phase, that we call the final phase, in which V_t goes from 2 to 1. We note that it is possible that V_t jumps directly to 1 from a value larger than 2, although we think this is an unlikely event, for large n. The bulk of the late phase, during which the diversity of language collapses from a large number to a small number of different words, is more difficult to assess and is not considered here.

In the next section we construct the model as a stochastic process, then describe the main results and give the layout for the rest of the article.

2 Construction and main results

We first note a useful "graphical construction" of the process, on a general locally finite graph G, from arbitrary initial data. We assume the vertices are totally ordered according to some fixed order. Given $\mu > 0$, let $\{(s_i, u_i) : i \ge 1\}$ be an independent and identically distributed sequence, with each s_i exponentially distributed with mean one and each u_i independent of s_i and uniform on $[0,1]$, and for $i \ge 1$, let $t_i = \mu^{-1}\sum_{j=1}^i s_j$. Then, the set of points

$$U := \{(t_i, u_i) : i \ge 1\} \subset \mathbb{R}_+ \times [0,1]$$

defines what we call an augmented Poisson point process with intensity μ, since (t_i) are the jump times of a Poisson process with intensity μ and each point t_i comes equipped with an independent uniform random variable u_i to help with the decision-making process.

Let F denote the set of directed edges $\{(v,w) : vw \in E\}$, and associate to each directed edge $(v,w) \in F$ an independent augmented Poisson point process $U(v,w)$ with intensity $(\deg v)^{-1}$. Suppose that $(t,u) \in U(v,w)$ and $|W_{t^-}(v)| = k$, with $W_{t^-}(v) = \{w_1, \dots, w_k\}$ labelled in increasing order.

- If $k = 0$ then v speaks word v to w at time t.
- If $k \ge 1$, then v speaks word w_i to w at time t if and only if

$$(i-1)/k \le u < i/k.$$

We then follow the rules as described above to determine W_t. If G is a finite graph, then since the intensity of the union $\bigcup_{(v,w)\in F} U(v,w)$ is finite, its points are well-ordered in time with probability 1, and so W_t can be determined from the initial state and the points $U(v,w)$ by updating sequentially in time. If G is an infinite graph, one needs to ensure that for each spacetime point (v,t), a finite number of events suffices to determine $W_t(v)$. Although this is not hard to do when there is some control on the degree, we will ignore it since from here on we focus on the case where G is the complete graph on n vertices and thus finite for any n.

Recall that $V_t = \bigcup_v W_t(v)$ denotes the set of words in existence at time t. The following result gives estimates of $|V_t|$, the cardinality of V_t, in the middle phase of the process.

Theorem 2.1. *For small enough $\epsilon > 0$, let $t_0 = (\frac{1}{2}+\epsilon)\log n$ and $t_1 = n^{1/2-\epsilon}$. Then*

$$P(\sup_{t_0\le t\le t_1} ||V_t| - \frac{n}{2}| \le n^{1-\epsilon}) \to 1 \quad \textit{as} \quad n\to\infty.$$

The result is proved in two main steps.

1. First, we show that $n/2 + n^{1/2+o(1)}$ are ever created, and within $(\frac{1}{2}+o(1))\log n$ time.
2. Then, we show that $o(n)$ words are deleted in $n^{1/2-o(1)}$ time.

The proof relies on approximating the size of the *cluster* $\mathcal{C}_t(w)$ corresponding to a given word w by a sort of branching process evolving in a non-stationary random environment. The cluster is defined by

$$\mathcal{C}_t(w) = \{v : w \in W_t(v)\}$$

and is the set of individuals that know word w at time t. We also need to control the correlation between distinct clusters $\mathcal{C}_t(w_1,), \mathcal{C}_t(w_2)$. To achieve both tasks we will use a slightly modified graphical construction which is better tailored to tracking the evolution of one or more distinguished clusters.

Suppose the process is started from an initial configuration in which, for some distinct pair of words A, B and each $v\in V$, $W_0(v) \in \{\{A\},\{B\},\{A,B\}\}$. Then, the same is true of $W_t(v)$, for all $v \in V$ and $t>0$. Since vertices in the complete graph are indistinguishable, if we let

$$x_t = n^{-1}|\{v\in V\colon W_t(v) = \{A\}\}|, \tag{2.1}$$
$$y_t = n^{-1}|\{v\in V\colon W_t(v) = \{B\}\}| \quad \text{and} \tag{2.2}$$
$$z_t = n^{-1}|\{v\in V\colon W_t(v) = \{A,B\}\}| \tag{2.3}$$

denote the proportion of each type, then (x,y,z) is a continuous-time Markov chain with state space $\Lambda_n = \{(x,y,z)\in \mathbb{N}^3/n\colon x+y+z=1\}$. The set of possible interactions between pairs of individuals of the three possible types A, B and AB, and the rate of each, is recorded in Table 1. Counting the number of edges connecting such pairs, we then easily obtain Table 2, that records the jump sizes Δ and transition rates q for the set of possible transitions of (x,y,z).

As mentioned earlier, the process eventually reaches an absorbing state.
In this context, that means that with probability one,

$$\lim_{t\to\infty}(x_t,y_t,z_t) \in \{(1,0,0),(0,1,0)\}.$$

The following result gives a sharp upper bound, over initial values in Λ_n, on how long this will take as a function of n, up to a $1+o(1)$ multiple of precision as $n\to\infty$.

reactants		product	$(n-1)\cdot$(rate)
$A + AB$	$\to$	$2AB$	$1/2$
$A + AB$	$\to$	$2A$	$3/2$
$B + AB$	$\to$	$2AB$	$1/2$
$B + AB$	$\to$	$2B$	$3/2$
$AB + AB$	$\to$	$2A$	1
$AB + AB$	$\to$	$2B$	1
$A + B$	$\to$	$B + AB$	1
$A + B$	$\to$	$A + AB$	1

Figure 1: List of possible interactions between two individuals

$n\Delta x$	$n\Delta y$	$n\Delta z$	$(n-1)q/n^2$
-1	0	1	$xz/2$
1	0	-1	$3xz/2$
0	-1	1	$yz/2$
0	1	-1	$3yz/2$
2	0	-2	$z(z-n^{-1})/2$
0	2	-2	$z(z-n^{-1})/2$
-1	0	1	xy
0	-1	1	xy

Figure 2: List of transitions with jumps Δ and rates q

Theorem 2.2. *Let z^* be the positive solution to $z^*(4+z^*) = 1$ and define the time to consensus*

$$T_c = \inf\{t : (x_t, y_t, z_t) \in \{(1,0,0), (0,1,0)\}\}.$$

(i) For any sequence of initial distributions and any $\alpha > 0$,

$$P(T_c/\log(n) \le 1 + 1/2z^* + \alpha) \to 1 \text{ as } n \to \infty,$$

(ii) and if $|x_0 - y_0| = O(1/\sqrt{n})$ and $|z_0 - z^*| = o(1)$ then

$$T_c/\log n \to 1 + 1/2z^* \text{ in probability as } n \to \infty.$$

Notice that, if individuals only remember the last word they heard, then starting from a configuration with two words, we obtain the voter model on the complete graph, for which the time to consensus is of order n. The reason it is much faster here is because, once a majority of type A or B develops, it is maintained. An easy way to see this is using the following informal argument. The average rate of change of (x, y, z) is given by summing Δ times q over the transitions in Table 2, which leads to the system of ODEs with $z = 1 - (x + y)$ and

$$\begin{aligned} x' &= xz + z^2 - xy, \\ y' &= yz + z^2 - xy. \end{aligned} \tag{2.4}$$

As depicted in Figure 3, on the set $\{(x,y) \in \mathbb{R}^2_+ : x + y \le 1\}$, there is a saddle point on the blue line $x = y$ and two stable equilibria, $(1, 0)$ and $(0, 1)$, that attract all points except those on the blue line. Since ours is a stochastic process, trajectories beginning on the blue line stray from it due to fluctuations, and are then swept away to one of the two stable equilibria. Quantitative arguments are given in Section 5.

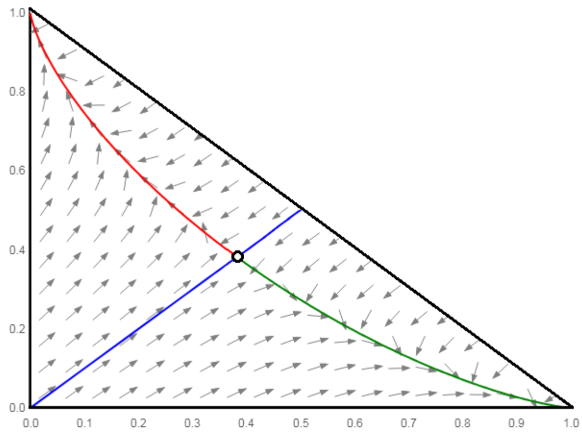

Figure 3: Phase portrait for the ODE system (2.4) in the (x, y) plane. Image generated using Darryl Nester's applet on `https://bluffton.edu/homepages/facstaff/nesterd/java/slopefields.html`

The paper is laid out as follows. In Section 3 we derive some useful estimates for semimartingales with bounded jumps, and give some formulas that help with computations later on. This section can be read independently of the rest of the paper, and may be of use in other applications. In Section 4 we prove Theorem 2.1 in several steps. In Section 4.1 we show that about $n/2$ words are created in about $\frac{1}{2}\log n$ time, using Chebyshev's inequality and a coupon-collecting argument, respectively. In Section 4.2 we show that at most $n^{1-\epsilon}/2$ words are deleted in $n^{1/2-\epsilon}$ time, which as noted above is achieved by controlling the number of individuals that know a given word, and which requires the estimates of Section 3. In Section 5 we use the approximating ODEs and the estimates of Section 3 to prove Theorem 2.2. Some additional results are collected in an Appendix, including the results of Section 3 and a general pathwise estimate for Poisson processes.

3 Sample path estimation

We use the theory of semimartingales, that is summarized in [8, Chapter I]. For our applications, we define the class of quasi-absolutely continuous semimartingales, which are (possibly discontinuous) processes for which we have drift and diffusion coefficients, generalizing the usual definition in the context of stochastic differential equations. We assume the reader is familiar with the notions of càdlàg, stopping time, predictable time and process, localization and martingale, which can be found in I.1 and I.2 of [8].

Below, we assume that processes are defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbf{F}, P)$ satisfying the usual conditions as described in [8, I.1.3], are optional as defined in [8, I.1.20], and take values in a complete metric space. If X is a càdlàg Feller process equipped with the completion of its natural filtration then since it is cádlág and adapted it is optional, and as shown in [10, I.5] it satisfies the usual conditions. Since the naming game is a continuous-time Markov chain, it is Feller (see for example [9, Proposition 17.2].

Given a càdlàg process X we denote by $X_- = (X_{t^-})_{t\geq 0}$ the left-continuous process (with $X_{0^-} = X_0$) and by $\Delta X = X - X_-$ the process of jumps. We say that X has *bounded jumps* if $|\Delta X| \leq c$ a.s. for some constant $c > 0$, and let $\Delta_\star(X)$ denote the infimum of such values of c. X is *quasi-left continuous* (qlc) if $\Delta X_\tau = 0$ a.s. on $\{\tau < \infty\}$ for any predictable time τ.

Given a process A, define the process $Var(A)$ by setting $Var(A)_t(\omega)$ equal to the total variation of the function $s \mapsto A_s(\omega)$ on the interval $[0,t]$. A process A has *finite variation* if $Var(A)_t(\omega) < \infty$ for each t, ω, and is *locally integrable* if it has a localizing sequence (τ_n) such that $E[Var(A)_{\tau_n}] < \infty$ for each n. The *compensator* of a locally integrable process A, denoted A^p, is the unique predictable and locally integrable process such that $A - A^p$ is a local martingale (see [8, I.3.18]).

A *semimartingale* (s-m) X is an $\mathbb{R}$-valued process that can be written as $X = X_0 + M + A$, where X_0 is an $\mathcal{F}_0$-measurable random variable, M is a local martingale and A has finite variation. We call a semimartingale *special* if it can be written in the above manner with a process A that is also predictable. If X is special, then as noted in [8, I.4.22], the decomposition with A predictable is unique, so we write

$$X = X_0 + X^m + X^p \tag{3.1}$$

where X^m is a local martingale with $X^m_0 = 0$ and X^p is predictable, both uniquely defined. By [8, I.4.24], if X has bounded jumps then it is special and $|\Delta X^m| \le 2\Delta_\star(X)$, and if it also qlc then using [8, I.2.35] in the proof of [8, I.4.24], we have the more convenient estimate $\Delta_\star(X^m) \le \Delta_\star(X)$.

Any $\mathbb{R}$-valued Markov chain is a semimartingale, since it is right-continuous and has locally finite variation, and is also quasi-left continuous, effectively because the jump times of a Poisson process are totally inaccessible; if this explanation is insufficient use Proposition 22.20 in [9] and note that Markov chains are Feller processes. As shown in [8, I.4.28], a deterministic function $f : \mathbb{R}_+ \to \mathbb{R}$ is a semimartingale iff it is right-continuous with finite variation over each compact interval, and is quasi-left continuous iff it is continuous, since any fixed time is predictable.

We will occasionally assume X is defined only up to some predictable time ζ that may be finite; in this case, information about X can be recovered from the stopped processes X^{τ_n} defined by $X^{\tau_n}_t = X_{t\wedge\tau_n}$, where τ_n is an announcing sequence for ζ, i.e., an increasing sequence of stopping times with limit ζ.

If local martingales M, N are locally square-integrable then as shown in [8, I.4.2], MN has a compensator, denoted $\langle M, N\rangle$ and called the predictable quadratic covariation. If $M = N$ we denote it $\langle M\rangle$ and call it predictable quadratic variation (pqv). Any local martingale M with $M_0 = 0$ and bounded jumps is locally square integrable (see [8, I.4.1]). If X is a special s-m and X^m is locally square-integrable we use $\langle X\rangle$ to denote $\langle X^m\rangle$.

We begin with a simple result that leads to an exponential estimate. It can probably be deduced from Theorem 2.3 in [7], but since the proof is not long, we include it in the Appendix.

Lemma 3.1. *Let M be a local martingale with $M_0 = 0$ and $|\Delta M| \le c$ for some $c > 0$. Then,*

$$\exp(M - (e^c/2)\langle M\rangle)$$

is a supermartingale with initial value 1.

The next result characterizes quasi-left continuity for semimartingales whose martingale part is locally square-integrable, and also motivates the definition of quasi-absolute continuity below.

Lemma 3.2. *Let X be a special semimartingale with locally square-integrable martingale part X^m.*
The following are equivalent:

1. *X is quasi-left continuous.*
2. *X^p is continuous and X^m is quasi-left continuous.*
3. *X^p and $\langle X\rangle$ are continuous.*

The following result is the exponential sample path estimate that we use throughout the paper. It resembles the estimates given in [7], but is different in the following sense: instead of bounding the running maximum of $|X^m|$ subject to a fixed constraint on $\langle X\rangle$, it bounds $|X^m|_t$ by a multiple of $\langle X\rangle_t$, plus a fixed error, uniformly over $t>0$.

Lemma 3.3. *Let X be a semimartingale with $|\Delta X|\le c$ for some $c>0$. Then for $\lambda, a>0$,*

$$\textit{if}\quad 0<\lambda c\le 1/2 \quad \textit{then}\quad P(\sup_{t\ge 0}|X_t^m|-\lambda\langle X\rangle_t\ge a\,)\le 2e^{-\lambda a}. \tag{3.2}$$

Using Lemma 3.2 as inspiration, say that a special semimartingale X with locally square-integrable martingale part X^m is *quasi-absolutely continuous* (qac) if both X^p and $\langle X^m\rangle$ are absolutely continuous. In this case define the *drift* $\mu(X)=(\mu_t(X))_t$ and the *diffusivity* $\sigma^2(X)=(\sigma_t(X))_t$ for Lebesgue-a.e. t by

$$\mu_t(X)=\frac{d}{dt}X_t^p,\quad \sigma_t^2(X)=\frac{d}{dt}\langle X\rangle_t. \tag{3.3}$$

Note that since X is assumed optional, we can take a single set of Lebesgue-a.e. t for which $\mu_t(X),\sigma_t^2(X)$ are defined a.s. This is necessary when performing calculations, as in the proof of Lemma 3.4.

For deterministic processes, qac is equivalent to absolute continuity, since $\mu_t(f)=f(t)$, $\sigma_t^2(f)=0$ and absolute continuity implies locally finite variation. If X is a pure jump Markov process on a space S with bounded rate kernel α as in [9, 17.2] and $f:S\to\mathbb{R}$ is bounded, then the process with $Y_t=f(X_t)$ is qac with

$$\mu_t(Y)=\int_S(f(x)-Y_t)\alpha(X_t,dx)\quad\text{and}\quad \sigma_t^2(Y)=\int_S(f(x)-Y_t)^2\alpha(X_t,dx). \tag{3.4}$$

Most of the processes dealt with in this paper will be of the above form, with X the naming game and f some observable. The next result gives a formula for the drift of the product of two qac semimartingales.

Lemma 3.4 (Product rule)**.** *Suppose X_t,Y_t are $\mathbb{R}$-valued qac s-m on a common filtered probability space. Then both $(XY)^p$ and $\langle X^m,Y^m\rangle$ exist and are absolutely continuous, and $\mu_t(XY)=\frac{d}{dt}(XY)_t^p$ is given by*

$$\mu(XY)=\sigma(X,Y)+X_-\mu(Y)+Y_-\mu(X),$$

where $\sigma_t(X,Y)=\frac{d}{dt}\langle X^m,Y^m\rangle_t$.

The following result helps to estimate the drift of functions of qac processes. It is Lemma 3 in [4].

Lemma 3.5 (Taylor approximation)**.** *Let X be a qac s-m with bounded jumps and let $f\in C^2(\mathbb{R})$. Then, $f(X)$ is a qac s-m and satisfies the following inequality for Lebesgue-a.e. t:*

$$|\mu_t(f(X))-f'(X_t)\mu_t(X)|\le\frac{1}{2}\sigma_t^2(X)\sup_{|x-X_t|\le\Delta_\star(X)}|f''(x)|.$$

The next result, which is Corollary 1 in [4], shows it is difficult for a process to surmount a "drift barrier", i.e., an interval $(0,x)$ in which there is at least a fixed amount $\mu_\star>0$ of negative drift, and diffusivity at most $\sigma_\star^2$. The strength of the estimate is exponential in $\mu_\star x/\sigma_\star^2$, both in time and probability.

Lemma 3.6 (Drift barrier)**.** *Fix $x>0$ and let X be a qac s-m on $\mathbb{R}$ with bounded jumps, such that $\Delta_\star(X)\le x/2$. Let τ be a stopping time. Suppose there are positive reals $\mu_\star,\sigma_\star^2,C_\mu,C_\Delta$ with $\max\{\Delta_\star(X)\mu_\star/\sigma_\star^2,1/2\}\le C_\Delta$ so that if $0<X_t<x$ and $t<\tau$ then*

$$\mu_t(X)\le-\mu_\star,\quad |\mu_t(X)|\le C_\mu\quad\textit{and}\quad\sigma_t^2(X)\le\sigma_\star^2. \tag{3.5}$$

Let $\Gamma = \exp(\mu_\star x/(32C_\Delta \sigma_\star^2))$. *Then we have*

$$P\left(\sup_{t\le \tau\wedge\lfloor\Gamma\rfloor x/16C_\mu} X_t \ge x \;\mid\; X_0 \le x/2\right) \le 4/\Gamma. \tag{3.6}$$

This result gives an upper bound on a non-decreasing qac s-m with bounded jumps, whose drift is sublinear with respect to some deterministic functions.

Lemma 3.7 (Sublinear drift)**.** *Let X be a qac s-m on $\mathbb{R}_+$ with jumps bounded by $c > 0$, defined for all $t < \zeta = \sup_{r>0} \inf\{t : X_t \ge r\}$. Suppose moreover that X is either*

(i) non-decreasing or
(ii) satisfies $\sigma^2(X) \le c\mu(X)$,

and also that $\mu(X)$ satisfies the inequality

$$\mu_t(X) \le b(t) + \ell(t)X_t \tag{3.7}$$

for some locally integrable non-nonegative deterministic functions $b(t)$, $\ell(t)$. Let $m(t) = \exp(\int_0^t \ell(s)ds)$ and let $Y_t = X_t/(X_0 m(t)) - \int_0^t b(s)/m(s)ds$ denote the rescaled process. Let $\zeta' = \zeta \wedge \inf\{t : m(t) = \infty\}$ and $\beta = \int_0^\infty b(t)/m(t)^2 dt$, and assume $\beta < \infty$. Then, $\zeta \ge \zeta'$ and for $y \ge 2$,

$$P(\sup_{t<\zeta'} Y_t \ge y) \le 2\mathbb{E}[e^{-(y-2)X_0/4c(1+\beta)}].$$

4 Early and middle phases

In this section we consider the behaviour of $|V_t|$ for $t \le n^{1/2-o(1)}$. Define

$$V_t^o = \bigcup_{(v,s):s\le t} W_s(v) \quad \text{and} \quad V_t^\times = V_t \setminus V_t^o,$$

respectively the set of words created up to time t, and the set of words created and then deleted by time t. Theorem 2.1 is implied by the following two propositions, whose proof is the objective of this section.

Proposition 4.1. *For each $\epsilon > 0$, $\lim_{n\to\infty} P(\sup_{t\ge(\frac{1}{2}+\epsilon)\log n} |\,|V_t^o| - \frac{n}{2}| \ge n^{1/2+\epsilon}) = 0$.*

Proposition 4.2. *For each $\epsilon > 0$, $\lim_{n\to\infty} P(\sup_{t\le n^{1/2-\epsilon}} |V_t^\times| \le n^{1-\epsilon}/2) = 1$.*

In words, in order to estimate $|V_t|$ we obtain good control on $|V_t^o|$, then show that $|V_t^\times|$ is not too big. We begin with V_t^o.

4.1 Creation of vocabulary

Our first task is to prove Proposition 4.1, and to do so we show that $|V_t^o|$ rises from 0 to $n/2 + O(n^{1/2+o(1)})$ within $\frac{1}{2}\log n$ time, then remains constant. For a vertex v let $N_t(v) = |W_t(v)|$ denote the size of the vocabulary of individual v, and let

$$T_o = \inf\{t : \min_v N_t(v) \ge 1\}$$

be the first time that every individual knows at least one word. Clearly V_t^o is non-decreasing as a set, so $V_\infty^o = \lim_{t\to\infty} V_t^o$ exists and $|V_\infty^o| \le n$. Once everyone knows a word, no new words are created, so $V_t = V_{T_o}^o = V_\infty^o$ for $t \ge T_o$. Proposition 4.1 is implied by the following two lemmas, in which we estimate T_o and $V_{T_o}^o$.

Lemma 4.3. *For $c \ge 0$,*

$$P(|T_o - \frac{1}{2}\log n| \ge c) \le 2e^{-c} + o(1) \text{ as } n \to \infty.$$

Proof of Lemma 4.3. Let $M_t = \{v : N_t(v) = 0\}$ denote *mute* vertices, those not yet knowing a word, and observe that $T_o \le t$ is equivalent to $|M_t| = 0$. For each distinct ordered pair of vertices (v, w), at rate $(n-1)^{-1}$, the directed edge (v, w) has an event, and both v and w are removed from M_t, if either or both still belongs. If we let $Z_t = |M_t|$ denote the number of mute vertices at time t, it follows that Z_t is a Markov chain with $Z_0 = n$ and transitions

$$Z_t \to \begin{cases} Z_t - 1 & \text{at rate } 2(n-1)^{-1} Z_t (n - Z_t), \text{ and} \\ Z_t - 2 & \text{at rate } (n-1)^{-1} Z_t (Z_t - 1). \end{cases}$$

We find that

$$\begin{aligned} \lim_{h \to 0^+} h^{-1} \mathbb{E}[Z_{t+h} - Z_t \mid Z_t = z] &= -2(n-1)^{-1} z(n-z) - 2(n-1)^{-1} z(z-1) \\ &= -2(n-1)^{-1}(nz - z^2 + z^2 - z) \\ &= -2(n-1)^{-1}(n-1)z = -2z. \end{aligned}$$

Letting $m(t) = \mathbb{E}[Z_t]$, $m(0) = n$ and taking expectations in the above, $m'(t) = -2m(t)$, which has the unique solution $m(t) = ne^{-2t}$. Fix $c \in \mathbb{R}$ and let $t_c = \frac{1}{2} \log n + c$. Using Markov's inequality,

$$P(T_o > t_c) = P(Z_{t_c} \ge 1) \le \mathbb{E}[Z_{t_c}] = e^{-2c}.$$

To get a lower bound we turn to Z_t^2, which has transitions

$$Z_t^2 \to \begin{cases} Z_t^2 - Z_t + 1 & \text{at rate } 2(n-1)^{-1} Z_t (n - Z_t), \text{ and} \\ Z_t^2 - 4Z_t + 4 & \text{at rate } (n-1)^{-1} Z_t (Z_t - 1), \end{cases}$$

so

$$\begin{aligned} \lim_{h \to 0^+} h^{-1} \mathbb{E}[Z_{t+h}^2 - Z_t^2 \mid Z_t = z] &= -(2z-1)2(n-1)^{-1} z(n-z) - (4z-4)(n-1)^{-1} z(z-1) \\ &= -4z(n-1)^{-1}((z - \frac{1}{2})(n-z) + (z-1)^2) \\ &= -4z(n-1)^{-1}(nz - z^2 - \frac{n}{2} + \frac{z}{2} + z^2 - 2z + 1) \\ &= -4z(n-1)^{-1}((n - \frac{3}{2})z + 1 - \frac{n}{2}) \\ &= \frac{2(n-2)}{n-1} z - \frac{4(n-3/2)}{n-1} z^2. \end{aligned}$$

Letting $\nu(t) = \mathbb{E}[Z_t^2]$, $\nu(0) = n^2$ and taking expectations above,

$$\nu'(t) = -4(1 - (2(n-1))^{-1})\nu(t)^2 + 2(1 - (n-1)^{-1})m(t),$$

so letting $\gamma = 4 - 2/(n-1)$, using $m(t) = ne^{-2t}$ and solving the above DE, we find

$$\nu(t) \quad = \quad n^2 e^{-\gamma t} + 2(1 - 1/(n-1)) n e^{-\gamma t} (e^{(\gamma - 2)t} - 1)/(\gamma - 2).$$

As above let $t_c = \frac{1}{2} \log n + c$, then $m(t_c) = e^{-2c}$ and for fixed c,

$$\nu(t_c) = e^{-4c} + e^{-2c} + o(1) \text{ as } n \to \infty,$$

so $Var(Z_{t_c}) = \nu(t_c) - m(t_c)^2 = e^{-2c} + o(1)$. Using Chebyshev's inequality,

$$\begin{aligned} P(T_o \le t_c) = P(Z_t = 0) \le P(|Z_t - \mathbb{E}[Z_t]| \ge \mathbb{E}[Z_t]) \le Var(Z_{t_c})/\mathbb{E}[Z_{t_c}]^2 \le \frac{e^{-2c} + o(1)}{e^{-4c}} \\ = e^{2c} + o(1). \end{aligned}$$

The result follows by taking a union bound of both estimates. ☐

We note in passing that $|V_0^o| = 0$ and $|V_t^o|$ increases by 1 at rate Z_t. Heuristically, $Z_t \approx ne^{-2t}$, so $|V_t^o| \approx (n/2)(1 - e^{-2t})$, for $t \leq \frac{1}{2}\log n$. This can be made precise using stochastic calculus, although we do not pursue it here.

Lemma 4.4. *Let $X = |V_{T_o}^o|$ be the number of words ever created. Then,*

$$\lim_{n\to\infty} P(|X - n/2| \geq n^\alpha) = 0 \quad \textit{for all} \quad \alpha > 1/2.$$

Proof of Lemma 4.4. Letting X_v for each vertex $v \in V$ be the Bernoulli random variable equal to one if and only if v speaks before listening, by construction and obvious symmetry, we have

$$X = \textstyle\sum_{v\in V} X_v \quad \text{and} \quad P(X_v = 0) = P(X_v = 1) = 1/2.$$

It follows that the expected number of words is given by

$$E(X) = \textstyle\sum_{v\in V} E(X_v) = \sum_{v\in V} P(X_v = 1) = n/2. \tag{4.1}$$

To also compute the variance, fix $v, w \in V$ and let B be the event that the first edge becoming active starting from v or w is edge vw. Since there are $n-1$ edges starting from each vertex,

$$P(B) = \frac{1}{2(n-1)-1} = \frac{1}{2n-3}. \tag{4.2}$$

In addition, the two vertices cannot both speak before listening when B occurs whereas the two events are independent on the event B^c therefore

$$\begin{array}{rcl} P(X_v = X_w = 1 \,|\, B) & = & 0 \\ P\,(X_v = X_w = 1 \,|\, B^c) & = & P(X_v = 1 \,|\, B^c)\, P(X_w = 1 \,|\, B^c) = 1/4. \end{array} \tag{4.3}$$

Combining (4.2)–(4.3), we deduce that

$$\begin{array}{rcl} E(X^2) & = & \displaystyle\sum_{v\in V} P(X_v^2 = 1) + \sum_{v\neq w} P(X_v = X_w = 1) \\ & = & \displaystyle\sum_{v\in V} \frac{1}{2} + \sum_{v\neq w} \frac{1}{4}\,\frac{2n-4}{2n-3} = \frac{n}{2}\left(1 + \frac{(n-1)(n-2)}{2n-3}\right) \end{array}$$

which, together with some basic algebra, gives the variance

$$Var(X) = \frac{n}{2}\left(1 + \frac{(n-1)(n-2)}{2n-3} - \frac{n}{2}\right) = \frac{n}{4}\left(\frac{n-2}{2n-3}\right) = O(n). \tag{4.4}$$

From (4.1) and (4.4) and Chebyshev's inequality, we conclude that

$$\lim_{n\to\infty} P(|X - n/2| \geq n^\alpha) \leq \lim_{n\to\infty} n^{-2\alpha} Var(X) = 0$$

for all $\alpha > 1/2$. This completes the proof. □

4.2 Maintenance of vocabulary

Next, we prove Proposition 4.2, that says that with probability tending to 1 as $n \to \infty$,

$$\sup_{t \leq n^{1/2 - o(1)}} |V_t^\times| = o(n).$$

Clearly $V_t^\times$, like V_t^o, is non-decreasing, since once a word vanishes from the population, it does not come back. We first bound $|V_t^\times|$ by a simpler quantity. Say that *agreement* upon word y occurs at (v, w, t) if

$$y \in W_{t^-}(w) \text{ and } v \text{ speaks word } y \text{ to } w \text{ at time } t.$$

If word w is created at some time $s \le t$, then $w \in W_s(w)$, and remains in individual w's vocabulary at least until the first time $t > s$ that agreement occurs at $(\cdot, w, t)$ or $(w, \cdot, t)\}$. This implies

$$V_t^{\times} \subseteq H_t = \{w : \text{agreement occurs at } (\cdot, w, s) \text{ or } (w, \cdot, s) \text{ for some } s \le t\}.$$

In words, in order to delete a word w from the population, it must at least be deleted from its source. Since each agreement contributes at most 2 to H_t, it follows that

$$|V_t^{\times}| \le 2A_t \quad \text{where} \quad A_t = \begin{array}{l} |\{s \le t : \text{ agreement occurs at } (\cdot, \cdot, s)\}| \\ \text{(number of agreements up to time } t). \end{array}$$

In order to control A_t we first define some useful observable quantities. For $w \in V$ we recall the *cluster* $\mathcal{C}_t(w)$ of w, that is, the set of individuals that know word w at time t:

$$\mathcal{C}_t(w) = \{v : w \in W_t(v)\}.$$

Recall that $N_t(v) = |W_t(v)|$ denotes the size of the vocabulary of individual v, and let

$$R_t(w) = \mathbf{1}(N_t(w) = 0) + \sum_{v \in \mathcal{C}_t(w)} 1/N_t(v) \tag{4.5}$$

denote the rate at which word w is spoken. Let $J(w, v)$ denote the times at which w speaks to v, and let

$$N_t^{\ell}(v) = \sum_w |J(w, v) \cap [0, t]| = \text{number of listening events for } v \text{ up to time } t,$$

noting that $N_t(v) \le N_t^{\ell}(v)$ and $\{(N_t^{\ell}(v)) : v \in V\}$ is a collection of independent Poisson processes with intensity 1. If we let

$$\begin{aligned} \tau_a(v) &= \inf\{t : v \in H_t\} \quad \text{and} \\ \tau_a(v, t) &= 0 \vee \sup\{s \le t : \text{agreement occurs at } (v, \cdot, s) \text{ or } (\cdot, v, s)\}, \end{aligned}$$

then $N_t(v) = N_t^{\ell}(v) - N^{\ell}_{\tau_a(v,t)^-}(v)$, and in particular,

$$N_t(v) = N_t^{\ell}(v) \text{ for } t < \tau_a(v).$$

Let $S_t(w) = |\mathcal{C}_t(w)|$ and $P_t(w) = (S_t(w) - 1)/(n - 1)$, and let $S_t = \max_w S_t(w)$. Each site v that knows word w speaks it at rate $N_t(v)^{-1}/(n-1)$ to each of the other $S_t(w) - 1$ sites in $\mathcal{C}_t(w)$. Letting

$$A_t(w) = |\{s \le t : \text{agreement occurs upon word } w \text{ at time } s\}|,$$

so that $A_t = \sum_w A_t(w)$, it follows that $A_t(w)$ increases by 1 at rate

$$(S_t(w) - 1) \sum_{v \in \mathcal{C}_t(w)} \frac{N_t(v)^{-1}}{n - 1} = R_t(w) P_t(w).$$

Since $\sum_{w \in V} R_t(w) = n$ is the total speaking rate and $P_t(w) \le (S_t - 1)/(n - 1) \le S_t/n$, summing the above display over $w \in V$ we find

$$A_t \text{ increases by } 1 \text{ at rate at most } S_t. \tag{4.6}$$

We have reduced the problem of controlling $|V_t^{\times}|$ to that of controlling S_t. The following becomes the goal of this subsection. Since its proof has a few parts, we call it a theorem.

Theorem 4.5. *For small $\epsilon > 0$,*

$$\lim_{n\to\infty} P\Big(\sup_{t\le n^{1/2-\epsilon}} \frac{S_t}{(1+t)^{1+\epsilon}} \ge (\log n)^9\Big) = 0.$$

Before moving onto the proof of Theorem 4.5 we first use it to obtain Proposition 4.2.

Proof of Proposition 4.2. From (4.6), for any $T > 0$, $\sup_{t\le T} A_t \le \text{Poisson}(\int_0^T S_u du)$. Using Theorem 4.5, with probability $1 - o(1)$ as $n \to \infty$

$$\begin{array}{rcl} \int_0^{n^{1/2-\epsilon}} S_u du & \le & (\log n)^9 \int_0^{n^{1/2-\epsilon}} (1+u)^{1+\epsilon} du \le (\log n)^9 (2+\epsilon)^{-1} (1+n^{1/2-\epsilon})^{2+\epsilon} \\ & \le & (\log n)^9 n^{1-3\epsilon/2-\epsilon^2} \end{array}$$

with the last inequality holding for large n. Since $P(\text{Poisson}(\lambda) \le 2\lambda) \to 1$ as $\lambda \to \infty$ it follows that for large n, $\sup_{t\le n^{1/2-\epsilon}} A_t \le n^{1-\epsilon}/4$ with probability $1 - o(1)$, and since $|V_t^\times| \le 2A_t$, this gives $\sup_{t\le n^{1/2-\epsilon}} |V_t^\times| \le n^{1-\epsilon}/2$ as desired. $\square$

To begin the proof of Theorem 4.5 we introduce a modified construction to help us make a coupling. First, for each ordered triple (y, z, v) let $R_t(y, z, v)$ be the rate at which word y is spoken by site z to v, let $R_t(y, v) = \sum_z R_t(y, z, v)$ be the rate at which site v hears word y, and as above let $R_t(y) = \sum_v R_t(y, v)$ be the rate at which word y is spoken. We calculate

$$\begin{array}{rcl} R_t(y,z,v) & = & (N_t(z))^{-1}\mathbf{1}(z \in \mathcal{C}_t(y),\ z \ne v) + \mathbf{1}(y = z \ne v,\ N_t(y) = 0))/(n-1) \text{ and} \\ R_t(y,v) & = & (\mathbf{1}(N_t(y) = 0, v \ne y) + \sum_{z\in\mathcal{C}_t(y)\setminus\{v\}} N_t(z)^{-1})/(n-1). \end{array} \tag{4.7}$$

Clearly $\sum_y R_t(y, v) = 1$ for each v, w and $t \ge 0$. Fix an ordering $v_1 < \cdots < v_n$ of V and define an independent family $\{U_v : v \in V\}$ of augmented Poisson point processes with intensity 1, that will correspond to listening events. For $v \in V$, $1 \le i, j \le n$ and $t \ge 0$ let

$$I_t(v,i,j) = \left[\sum_{k=1}^{i-1} R_t(v_k, v) + \sum_{m=1}^{j-1} R_t(v_i, v_m, v), \sum_{k=1}^{i-1} R_t(v_k, v) + \sum_{m=1}^{j} R_t(v_i, v_m, v)\right),$$

noting that $\{I_t(v, i, j) : 1 \le i, j \le n\}$ partitions $[0, 1)$. Then, if $(t, u) \in U_v$ and $u \in I_{t^-}(v, i, j)$, word v_i is spoken by v_j to v, which defines the process. Using this construction and given $C, R > 0$ we obtain upper bounds $\mathbf{C}_t(w), \mathbf{R}_t(w)$ on $\mathcal{C}_t(w), R_t(w)$ for all $w \in V$, valid up to the time

$$\begin{array}{rcl} T_{C,R} & = & \min_{w\in V} T_w(C,R) \quad \text{where} \\ T_w(C,R) & = & \inf\{t : \sum_{v\in\mathbf{C}_t(w)\cap H_t} N_t(v)^{-1} \ge C \text{ or } \mathbf{R}_t(w) \ge R\}. \end{array}$$

That is, we obtain for each $w \in V$ a pair of processes $\mathbf{C}_t(w), \mathbf{R}_t(w)$ with nice properties, such that $\mathcal{C}_t(w) \subseteq \mathbf{C}_t(w)$ and $R_t(w) \le \mathbf{R}_t(w)$ for $t \le T_{C,R}$ pointwise on realizations of the process. The definitions will look a bit strange but should be easier to understand after reading the proof of the upcoming Lemma 4.6. Given $w \in V$, $\mathbf{C}_t(w), \mathbf{R}_t(w)$ are non-decreasing and defined as follows. For $i \in \{1, \ldots, n\}$ let

$$\begin{array}{rcl} b_t(v,i) & = & \sum_{k=1}^{i-1} R_t(v_k, v), \text{ and for } x \in [0,1) \text{ let} \\ I_t(v,i,x) & = & [b_t(v,i),\ b_t(v,i) + x) \mod 1. \end{array}$$

Define

$$N_t^\ell(v,i,R) = |\{(s,u) \in U_v : s \le t,\ u \notin I_t(v,i,R)\}| \le N_t^\ell(v),$$

which increases at constant rate $1-R$, and ignores listening events during which word v_i is spoken, so long as v_i is spoken at rate at most R. Let $\mathbf{C}_0(w) = \{w\}$ and $\mathbf{R}_0(w) = 1 + C$ for each w. $\mathbf{R}_t(v_i)$ is defined for each i and $t > 0$ as follows:

$$\mathbf{R}_t(v_i) = 1 + C + \sum_{v \in \mathbf{C}_t(v_i)\setminus\{v_i\}} 1/(1 + N_t^\ell(v, i, R)).$$

In words, $\mathbf{R}_t(v_i)$ assigns a basic speaking rate of 1 (for v_i) plus C (to account for sites in $\mathbf{C}_t(v_i)$ at which agreement has occurred – see the definition of $T_w(C, R)$), plus an additional (smaller and more accurate) speaking rate for all sites in $\mathbf{C}_t(v_i)$ aside from v_i. Then, $\mathbf{C}_t(v_i)$ is defined as follows.

$$\begin{array}{ll} \text{if} & (t,u) \in U_v \text{ and } u \in I_{t^-}(v, i, \mathbf{R}_{t^-}(v_i)/(n-1)), \\ \text{then} & \mathbf{C}_t(v_i) = \mathbf{C}_{t^-}(v_i) \cup \{v\}. \end{array}$$

This is such that, if word v_i is spoken at rate at most $\mathbf{R}_t(v_i)$, then any site that is added to $\mathcal{C}_t(v_i)$ is also added to $\mathbf{C}_t(v_i)$. We now demonstrate the claimed comparison.

Lemma 4.6. *For each $w \in V$ and all $t < T_{C,R}$, $\mathcal{C}_t(w) \subseteq \mathbf{C}_t(w)$ and $R_t(w) \le \mathbf{R}_t(w)$.*

Proof. Let

$$\tau_c(w) = \inf\{t : \mathcal{C}_t(w) \neq \varnothing\},$$

then $\mathcal{C}_t(w) \subset \{w\} \subseteq \mathbf{C}_t(w)$ and $R_t(w) \le 1 \le \mathbf{R}_t$ for $t < \tau_c(w)$. For the remainder, assume $t \ge \tau_c(w)$ and let i be such that $w = v_i$. By construction, $v \in V$ is added to $\mathcal{C}_t(w)$ if

$$v \notin \mathcal{C}_{t^-}(w),\ (t,u) \in U_v \text{ and } u \in I_{t^-}(v, i, R_t(w, v)) \tag{4.8}$$

and otherwise, $\mathcal{C}_t(w)$ does not increase. If $t \ge \tau_c(w)$ then $N_t(w) \ge 1$, and if $z \notin H_t$ then $N_t^\ell(z) = N_t(z)$. So, from the second line of (4.7),

$$\begin{array}{rcl} (n-1)R_t(w,v) & = & \sum_{z \in \mathcal{C}_t(w)\setminus\{v\}} 1/N_t(z) \\ & \le & \sum_{z \in \mathcal{C}_t(w)} 1/N_t^\ell(z) + \sum_{z \in \mathcal{C}_t(w) \cap H_t} 1/N_t(z). \end{array}$$

If $w \in \mathcal{C}_t(w)$ then $N_t(w)^{-1} \le 1$. By definition of $T_{C,R}$, if $\mathcal{C}_t(w) \subseteq \mathbf{C}_t(w)$ and $t < T_{C,R}$ then

$$(n-1)R_t(w,v) \le (1 + C + \sum_{z \in \mathcal{C}_t(w)\setminus\{w\}} 1/N_t^\ell(z)).$$

If $v \in \mathbf{C}_t(w)$ and $t < T_{C,R}$ then $N_t^\ell(v) \ge N_t^\ell(v, R) + 1$, since this implies existence of a point in

$$U_v \cap \{(s,u) : s \le t \text{ and } u \in I_s(v, i, \mathbf{R}_t(w)/(n-1))\},$$

that is counted in $N_t^\ell(v)$ but not in $N_t^\ell(v, i, R)$. If $\mathcal{C}_t(w) \subseteq \mathbf{C}_t(w)$ it follows that $R_t(w,v) \le \mathbf{R}_t(w)/(n-1)$ for each v which implies the containment $\mathcal{C}_t(w) \subseteq \mathbf{C}_t(w)$ is preserved across transitions (4.8) that cause $\mathcal{C}_t(w)$ to increase. Since $\mathbf{C}_t(w)$ is non-decreasing and transitions are well-ordered this implies $\mathcal{C}_t(w) \subseteq \mathbf{C}_t(w)$ for $t < T_{C,R}$. It remains to check $R_t(w) \le \mathbf{R}_t(w)$ for $\tau_c(w) \le t < T_{C,R}$. But in this case, (4.5) and the previous argument give

$$R_t(w) = \sum_{v \in \mathcal{C}_t(w)} 1/N_t(v) \le 1 + C + \sum_{v \in \mathcal{C}_t(w)\setminus\{w\}} 1/N_t^\ell(v) \le \mathbf{R}_t(w). \qquad \square$$

Next we fix w and examine $\mathbf{C}_t(w), \mathbf{R}_t(w)$ assuming $t < T_{C,R}$, and dropping the (w) for neatness. Notice that $|\mathbf{C}_t|$ is non-decreasing and increases by 1 at rate at least $(1+C)(n - |\mathbf{C}_t|)/(n-1)$, which implies $\lim_{t\to\infty} |\mathbf{C}_t| = n$. Since $|\mathbf{C}_t|$ increases by one at a time, let $y_1, \dots, y_n$ be the order in which vertices are added to $\mathbf{C}_t$, with $w = y_1$, and condition on $(y_1,\dots,y_n)$. We track $Z_t = |\mathbf{C}_t|$ and $N_t^i = N_t^\ell(y_i,R),\ i=1,\dots,n$ which suffices to determine $\mathbf{C}_t,\mathbf{R}_t$. Let $t_i = \inf\{Z_t = i\}$ denote the time at which y_i is added to $\mathbf{C}_t$, and let k be such that $w = v_k$. For $i \in \{2,\dots,n\}$, t_i is the least value of t such that there is a point

$$(t,u) \in \bigcup_{j\ge i} U_{y_j} \cap \{(s,v) : s \in [t_{i-1},\infty),\ v \in I_s(y_j,k,\mathbf{R}_{t_{i-1}}/(n-1))\},$$

and in addition, this point belongs to U_{y_i}. Using this and basic properties of exponential random variables, together with the thinning property of the Poisson process, we find that conditioned on $(y_1,\dots,y_n)$,

$$(Z_t, N_t^1,\dots,N_t^n)_{t<T_{C,R}}$$

is a Markov chain with the following transitions:

$$\begin{aligned} Z_t \to Z_t+1 \quad &\text{at rate} \quad \mathbf{R}_t(n-Z_t)/(n-1), \text{ and} \\ \text{for } i=1,\dots,n,\ N_t^i \to N_t^i+1 \quad &\text{at rate} \quad 1-R/(n-1). \end{aligned}$$

In particular, $\{(N_t^i)_{t<T_{C,R}} : i=1,\dots,n\}$ is an i.i.d. collection of Poisson processes with intensity $1-R/(n-1)$. Since the above does not depend on the choice of values for $(y_1,\dots,y_n)$ the same holds unconditionally. Thus Z_t can be viewed as follows: initially $Z_0 = 1$, then subject to the random environment determined by the $\{(N_t^i)\}_{i=2}^n$, Z_t increases by 1 at rate $\mathbf{R}_t(n-Z_t)/(n-1)$. Define (Λ_t, X_t) by

$$\begin{aligned} &\Lambda_t(z) = 1 + C + \sum_{i=2}^{z} 1/(1+N_t^i) \quad \text{and} \\ &X_0 = 1,\ X_t \text{ increases by } 1 \text{ at rate } \Lambda_t(X_t). \end{aligned} \tag{4.9}$$

Since $(n-Z_t)/(n-1) \le 1$ and Λ_t is non-decreasing in z, it follows that

$$(Z_t,\mathbf{R}_t)_{t<T_{C,R}} \quad \text{is dominated by} \quad (X_t,\Lambda_t(X_t)). \tag{4.10}$$

We can think of (X_t) as a branching process with immigration rate $1+C$, in which individual i produces offspring at the time-decreasing rate $1/(1+N_t^i)$. Two tasks lie ahead. The first is to estimate (X_t). The second is to estimate $T_{C,R}$. We then combine the results to obtain Theorem 4.5. This is outlined as follows.

Proposition 4.7. *Let $b = 1+C$. If $\epsilon > 0$, $b \le (27\log n)^4$ and $R = o(n)$ then*

$$P(\sup_{t\le T_{C,R}} S_t/(1+t)^{1+\epsilon} > (\log n)^9) = o(1/n).$$

Proposition 4.8. *If $\epsilon > 0$ is small, $b = (27\log n)^4$ and $R = b + (\log n)^{11}$ then*

$$\lim_{n\to\infty} P(T_{C,R} \le n^{1/2-\epsilon}) = 0.$$

Proof of Theorem 4.5. Use Propositions 4.7 and 4.8 with $b = (27\log n)^4$ and $R = b + (\log n)^{11}$. ☐

4.2.1 Estimation of (X_t)

Since n does not appear in the definition of (X_t) we may as well define it using an infinite sequence $\{(N_t^i)_{t\ge 0} : i = 1,2,\dots\}$ of Poisson processes with intensity $r = 1-R/(n-1)$. Clearly $r \le 1$. Since R will be chosen $o(n)$, we will have $r \to 1$ as $n \to \infty$, so throughout we assume $r \ge 1/2$.

We begin with a useful heuristic. Let $b = 1 + C$. Replacing N_t^i with its expectation rt, X_t increases by 1 at rate $b + X_t(1 + rt)^{-1}$, which we approximate with the differential equation

$$x' = b + x/(1 + rt).$$

Let $m(t) = \exp(\int_0^t (1 + rs)^{-1} ds) = (1 + rt)^{1/r}$. The above equation is linear and has solution

$$x(t) = m(t)x(0) + bm(t) \int_0^t ds/m(s).$$

If r is close to 1 then $x(t)$ grows just a bit faster than linearly in time. In order to analyze (X_t) we break it up into two steps:

1. Up to a fixed time T, when the N_t^i are fairly small.
2. From time T to ∞, when the N_t^i are fairly large.

The reason to do this is because the estimates that say $|N_t^i - rt| = o(rt)$ are only effective once rt has had time to increase. The following is the main result of this subsection.

Proposition 4.9. *There exist $M, x_0 \in [1, \infty)$ so that if $r \in [1/2, 1]$, $b \geq 2$ and $x \geq b \vee x_0$ then*

$$P(\sup_{t \geq 0} X_t - Mx(1 + t)^{1/r}(x + \log(1 + t)) > 0) \leq 3e^{-x^{1/4}/9}.$$

Recall $S_t = \max_w |\mathcal{C}_t(w)|$. Using this result we can prove Proposition 4.7.

Proof of Proposition 4.7. For each $w \in V$, using Lemma 4.6 and (4.10),

$$(|\mathcal{C}_t(w)|)_{t \leq T_{C,R}} \quad \text{is dominated by} \quad (X_t).$$

Since $R = o(n)$ by assumption and recalling $r = 1 - R/(n - 1)$, $r \geq 1/2$ and $1/r \leq 1 + \epsilon/2$ for large n. Letting $x = (27 \log n)^4$, $x \geq b$ by assumption and $x \geq x_0$ for large n, so applying the result of Proposition 4.9 and taking a union bound over w,

$$P(\sup_{t \leq T_{C,R}} S_t - \Phi(t, x) > 0) \leq 3ne^{-x^{1/4}/9} = 3n^{1-3} = o(1/n), \text{ where}$$
$$\Phi(t, x) = Mx(x + \log(1 + t))(1 + t)^{1/r}.$$

Since $1 + \log(1 + t) = O((1 + t)^{\epsilon/2})$ and using $1/r \leq 1 + \epsilon/2$,

$$\Phi(t, x) = O((\log n)^8 (1 + t)^{1+\epsilon}),$$

which is at most $(\log n)^9 (1 + t)^{1+\epsilon}$ for large n and all $t \geq 0$, completing the proof. $\square$

We tackle the proof of Proposition 4.9 in a couple of steps.

Step 1. We obtain a somewhat crude upper bound on (X_t) that has the virtue of being effective starting at time 0. For $i \geq 1$ let $t_i = \inf\{t : X_t = i\}$, define $N_i = N_{t_i}^i$ then define $Y_t,\ Q_t$ by

$$Y_0 = 1 \quad \text{and} \quad Y_t \to Y_t + 1 \quad \text{at rate} \quad Q_t = b + \sum_{i=2}^{Y_t} 1/(1 + N_{t_i}^i).$$

In words, at the moment t_i an individual i is added to the process, the corresponding counting process N_t^i is stopped, so that i always contributes $(1 + N_{t_i}^i)^{-1}$ to Q_t. Since $(1 + N_{t_i}^i)^{-1} \geq (1 + N_t^i)^{-1}$ for $t \geq t_i$, (X_t) is dominated by (Y_t). The next result controls (Y_t).

Lemma 4.10. *There is $M_1 \in [1, \infty)$ so that for $a, b \geq 2$ and $r \in [1/2, 1]$,*

$$P(\sup_{t \geq 0} Y_t/(1 + t)^{1+1/r} \geq abM_1) \leq 4e^{-(a-2)/2}$$

Proof. Begin by observing that (Q_t) has the concise description

$$Q_0 = b \quad \text{and} \quad Q_t \to Q_t + \Delta_t \quad \text{at rate} \quad Q_t$$

where the increment $\Delta_t \stackrel{d}{=} (1 + \text{Poisson}(rt))^{-1}$ is independently sampled every time there is a jump. Our first task is to control the size of Q_t. We compute the drift:

$$\mu_t(Q) = \ell(t) Q_t \quad \text{with} \quad \ell(t) := \mathbb{E}[\Delta_t]$$

Let $g(t) = \exp(\int_0^t \mathbb{E}[(1+\text{Poisson}(rs))^{-1}]ds)$. Using Lemma 3.7 with $b(t) = 0$ and $c = 1$, for $a \geq 2$ we find

$$P(\sup_t Q_t / g(t)) \geq ab) \leq 2e^{-(a-2)b/4} \tag{4.11}$$

This translates to a bound on $(Y_t)_{t \geq 0}$ as follows. Since $\mu_t(Y) = Q_t$,

$$Y_t^m = Y_t - Y_0 - \int_0^t Q_s ds$$

Since $(Y_t)_{t \geq 0}$ has transition rate Q_t and jump size exactly 1, $\sigma^2(Y_t) = Q_t$. Using Lemma 3.3 with $\lambda = 1/2$ and $c = 1$ while noting $Y_0 = 1$,

$$P(Y_t \geq 1 + a + \frac{3}{2} \int_0^t Q_s ds \quad \text{for some} \quad t \geq 0) \leq 2e^{-a/2}$$

Combining with (4.11), taking a union bound, and noting $e^{-(a-2)b/4}, e^{-a/2} \leq e^{-(a-2)/2}$ if $b \geq 2$,

$$P(Y_t \geq 1 + a \left(1 + \frac{3b}{2} \int_0^t g(s) ds\right) \quad \text{for some} \quad t \geq 0) \leq 4e^{-(a-2)/2}. \tag{4.12}$$

Intuitively, $g(t)$ grows roughly like $m(t)$. Let $\xi = \text{Poisson}(\lambda)$. Since $x \mapsto (1+x)^{-1}$ is convex, the inequality $\mathbb{E}[(1+\xi)^{-1}] \geq (1 + \mathbb{E}[\xi])^{-1}$ goes in the wrong direction for an upper bound on $g(t)$. Anticipating our needs, we let $x = \lambda^\alpha / 2$ in Lemma 5.11 to find

$$P(\xi < \lambda - \lambda^{1/2+\alpha}/2) \leq e^{-\lambda^{2\alpha}/8} \quad \text{if} \quad 0 < \alpha \leq 1/2.$$

Using the fact that $(1+\xi)^{-1} \leq 1$ and that probabilities are at most 1, then using Lemma 5.10 with $c = 1/2$, if $\lambda \geq 1$ (which implies $c\lambda^{\alpha-1} \leq 1$) then

$$\begin{aligned}
\mathbb{E}[(1+\xi)^{-1}] &= \mathbb{E}[(1+\xi)^{-1} \, ; \, \xi \geq \lambda - \lambda^{1/2+\alpha}/2] + \mathbb{E}[(1+\xi)^{-1} \, ; \, \xi < \lambda - \lambda^{1/2+\alpha}/2] \\
&\leq (1 + \lambda - \lambda^{1/2+\alpha}/2)^{-1} P(\xi \geq \lambda - \lambda^{1/2+\alpha}/2) + P(\xi < \lambda - \lambda^{1/2+\alpha}/2) \\
&\leq (1 + \lambda - \lambda^{1/2+\alpha}/2)^{-1} + e^{-\lambda^{2\alpha}/8} \\
&\leq (1+\lambda)^{-1} + (1+\lambda)^{-3/2+\alpha} + e^{-\lambda^{2\alpha}/8}.
\end{aligned}$$

If $\lambda < 1$ we will use the trivial estimate $\mathbb{E}[(1+\xi)^{-1}] \leq 1$. If $0 < \alpha < 1/2$ and $0 < r \leq 1$ then

$$c(r, \alpha) := 2 + \int_0^\infty ((1+rs)^{-3/2+\alpha} + e^{-(rs)^{2\alpha}/8}) ds < \infty.$$

Let $c(r) = \inf\{c(r,\alpha) : \alpha \in (0, 1/2)\}$ and let $c = c(1/2)$. Since $c(r,\alpha)$ decreases with r, it follows that $c(r) \leq c$ for $r \geq 1/2$. Recalling $m(t) = \exp(\int_0^t ds/(1+rs))$ defined earlier, and noting $rs \geq 1$ if $s \geq 2$, it follows that

$$\begin{aligned}
g(t) &\leq \inf_{\alpha \in (0,1/2)} \exp\left(\int_0^2 1 ds + \int_2^t ((1+rs)^{-1} + (1+rs)^{-3/2+\alpha} + e^{-(rs)^{2\alpha}/8}) ds\right) \\
&\leq \inf_{\alpha \in (0,1/2)} \exp\left(2 + \int_0^t ((1+rs)^{-1} + (1+rs)^{-3/2+\alpha} + e^{-(rs)^{2\alpha}/8}) ds\right) \\
&\leq e^c m(t)
\end{aligned}$$

Since $m(t) = (1+rt)^{1/r}$ and $1/(r(1+1/r)) = 1/(r+1)$, it follows that

$$\begin{aligned} \textstyle\int_0^t g(s)ds \le e^c \int_0^t (1+rt)^{1/r} &= e^c((1+rt)^{1/r+1}-1)/(r+1) \\ &\le e^c((1+t)^{1/r+1}-1), \end{aligned}$$

using $r > 0$ and $r \le 1$ in the last step. If $a \ge 2$ and $b \ge 1$ then since $c > 0$, $1+a(1-3be^c/2) \le 0$ and

$$1 + a\left(1 + \frac{3b}{2}\int_0^t g(s)ds\right) \le \frac{3abe^c}{2}(1+t)^{1/r+1}.$$

To conclude, take $M_1 = \frac{3}{2}e^c$ and use (4.12). □

Step 2. Next, we do two things.

1. *Lemma 4.11*. We control the environment $\{(N_t^i)\}_{i\ge1}$ for $t \in [T,\infty)$.
2. *Lemma 4.12*. We use this to get an upper bound on (X_t) for $t \in [T,\infty)$.

Let

$$\tau_{lp}(i) = \sup\{t : N_t^i - rt + 2(rt)^{3/4} < 0\} \quad \text{for} \quad i \ge 1$$

denote the last passage time of N_t^i below the curve $v(t) = rt - 2(rt)^{3/4}$, and for $t \ge 0$ let

$$I_t = \max\{i : \tau_{lp}(j) \le t \text{ for all } j \le i\}.$$

For later use, we note that

$$\Lambda_t(x) \le b + x/(1+v(t)) \text{ for } x \le I_t. \tag{4.13}$$

Lemma 4.11. *If $T > 0$ is large enough, then for any $r \in [1, 2/1]$,*

$$P(\inf_{t>T} I_t - e^{t^{1/2}/9} < 0) \le e^{-T^{1/2}/9}.$$

Proof. For each i, using Lemma 5.12 with $\alpha = 1/4$ and $\tau_2 = \tau_{lp}(i)$,

$$P(\tau_{lp}(i) \ge t) \le 6t^{1/2}e^{-(rt)^{1/2}/3} \quad \text{if} \quad t \ge 4 \;\; \text{and} \;\; t^{1/2} \ge 24.$$

Let $f(t) = t^{-1/4}e^{(rt)^{1/2}/6}/\sqrt{6}$, so the right-hand side above is $1/f(t)^2$.
For t large enough that $f(t) \ge 1$, a union bound gives

$$\begin{aligned} P(I_t < f(t)) = P(\max\nolimits_{j\le\lceil f(t)\rceil} \tau_{lp}(j) > t) &\le \lceil f(t)\rceil/f(t)^2 \\ &\le f(t)^{-1}(1+f(t)^{-1}) \le 2f(t)^{-1}. \end{aligned}$$

For $T > 0$ let $c_1(T) = \sup_{t\ge T} f(t)/f(t+1)$ and note that $c_1(T) \to 1$ as $T \to \infty$.
Since I_t is non-decreasing, if $I_t \ge f(t)$ and $t > T$ then

$$I_{t+h} \ge f(t) \ge c_1(T)f(t+h) \quad \text{for} \quad h \in [0,1).$$

Taking a union bound over the estimate at times $T+k$, $k \ge 0$ gives

$$\begin{aligned} p(T) := P(\inf_{t>T} I_t - c_1(T)f(t) < 0) &\le \sum_{k\ge0} P(I_{T+k} < f(T+k)) \\ &\le \sum_{k\ge0} 2/f(T+k) \end{aligned}$$

If T is large enough that $f(t)$ is increasing, it follows that

$$p(T) \le \int_{T-1}^{\infty} 2\sqrt{6}t^{1/4}e^{-(rt)^{1/2}/6}dt.$$

With $\alpha = 1/4$ let $c_2(T) = (\alpha r^{2\alpha}/3 - (3/2 - 3\alpha)T^{-2\alpha})^{-1}$. Using Lemma 5.9 with $a = 1/2 - \alpha, \beta = 2\alpha$ and $c = r^{2\alpha}/6$ while noting $1 + a - \beta = 3/2 - 3\alpha > 0$, if $1/c_2(T) > 0$ then we obtain

$$p(T) \le 2\sqrt{6}c_2(T)(T-1)^{3/4}e^{-r^{1/2}(T-1)^{1/2}/6}.$$

For large T, $c_1(T) \ge 1/2$ and $c_2(T) \le 6r^{-2\alpha}/\alpha = 24r^{-1/2} \le 48$, since $r \ge 1/2$. Since $r^{1/2} \ge 1/\sqrt{2} > 2/3$, by giving up a bit in the exponents, for large T we have

$$p(T) \le e^{-(1+T)^{1/2}/9} \quad \text{and} \quad c_1(T)f(t) \ge e^{t^{1/2}/9}$$

and the result follows. □

Lemma 4.12. *There is $M_2 \in [1,\infty)$ so that if $T > 0$, $\tau = \inf\{t > T : X_t > I_t\}$, $r \in [1/2, 1)$, $a \ge 2$ and $x_1, b > 0$ then*

$$P(\sup_{T \le t < \tau} X_t - M_2(1+t)^{1/r}\left(\frac{ax_1}{(1+T)^{1/r}} + b\log(1+t)\right) > 0 \mid X_T \le x_1) \\ \le 2e^{-(a-2)x_1/4(1+b(1+T))}$$

Proof. Recall Λ_t defined in (4.9). Using (4.9) and (4.13), it follows that for $T \le t < \tau$ and conditioned on $X_T \le x_1$, (X_t) is dominated by the process $(\tilde{X}_t)$ with $\tilde{X}_T = x_1$ that increases by 1 at rate $b + \tilde{X}_t/(1 + v(t))$, where $v(t) = rt - 2(rt)^{3/4}$. We proceed as in the proof of Lemma 4.10. We have

$$\mu_t(\tilde{X}) = b + \ell(t)\tilde{X}_t \quad \text{with} \quad \ell(t) := 1/(1 + v(t)).$$

For $a > 0$ let $E_a = \{\sup_{t \ge T} \tilde{X}_t/g(t) - (ax_1 + b\int_T^t ds/g(s)) > 0\}$, where $g(t) = \exp(\int_T^t \ell(s)ds)$, and let $\beta = b\int_T^\infty ds/g(s)^2$. Using Lemma 3.7 with $c = 1$, for $a \ge 2$ we find

$$P(E_a) \le 2e^{-(a-2)x_1/4(1+\beta)}. \tag{4.14}$$

Using Lemma 5.10 with $\lambda = rt$, $\alpha = 1/4$ and $c = 2$, if $4(rt)^{-1/2} \le 1$, i.e., $rt \ge 16$, then

$$\ell(t) = 1/(1 + v(t)) \le (1 + rt)^{-1} + (1 + rt)^{-5/4}. \tag{4.15}$$

Let $c(r) = \int_0^\infty (1 + rs)^{-5/4}ds$, which is finite for $r \in (0, 1]$ and decreases with r. Let $c = c(1/2)$, so that $c(r) \le c$ for $r \in [1/2, 1]$. Combining and noting that $u \mapsto (1+ut)/(1+uT)$ is non-decreasing in u if $t \ge T$,

$$g(t) \le e^c \exp\left(\int_T^t ds/(1 + rs)\right) = e^c\left(\frac{1 + rt}{1 + rT}\right)^{1/r} \le \left(\frac{1+t}{1+T}\right)^{1/r}.$$

Therefore

$$E_a \supseteq \{\sup_{T \le t} \tilde{X}_t - e^c(1+t)^{1/r}\left(\frac{ax_1}{(1+T)^{1/r}} + \frac{b}{(1+T)^{1/r}}\int_T^t ds/g(s)\right) > 0.$$

Using $\ell(t) \ge 1/(1 + t)$ and integrating, $g(t) \ge (1 + t)/(1 + T)$. Since $1/r \ge 1$ and $\log(1 + T) \ge 0$,

$$\frac{b}{(1+T)^{1/r}}\int_T^t ds/g(s) \le b\int_T^t ds/(1 + s) \le b\log(1 + t),$$

so with $M_2 = e^c$,

$$E_a \supseteq \{\sup_{T \le t} \tilde{X}_t - M_2(1+t)^{1/r}\left(\frac{ax_1}{(1+T)^{1/r}} + b\log(1+t)\right) > 0. \tag{4.16}$$

Using again $g(t) \geq (1+t)/(1+T)$,

$$\begin{aligned} \beta = b\int_T^\infty ds/g(s)^2 &\leq b(1+T)^2\int_T^\infty (1+s)^{-2}ds \\ &= b(1+T). \end{aligned}$$

Using the above in (4.14) and combining with (4.16), we obtain the result. □

Proof of Proposition 4.9. We note that Lemma 4.10 is true with X_t in place of Y_t since (X_t) is dominated by (Y_t). Let $L(t) = e^{t^{1/2}/9}$. Recall x from the statement of the Proposition, and let $T = x^{1/2} - 1$ so that $x = (1+T)^2$. Let $x_1 = abM_1(1+T)^{1+1/r}$. Let

$$\begin{aligned} E &= \{\sup_{t\leq T} X_t/(1+t)^{1+1/r} \leq abM_1\}, \\ F &= \{\inf_{t\geq T} I_t - L(t) \geq 0\} \quad \text{and} \\ G &= \{\sup_{T\leq t<\tau} X_t - (1+rt)^{1/r}(aM_2x_1/(1+rT)^{1/r} + bM_2\log(1+t)) \leq 0\} \end{aligned}$$

be the complements of the events from, respectively, Lemma 4.10, 4.11 and 4.12. On E,

$$\sup_{t\leq T} X_t/(1+t)^{1/r} \leq abM_1(1+T).$$

In particular, $X_T \leq x_1$, so using Lemma 4.12, for $b \geq 2$ and large T,

$$\begin{aligned} P(G^c\cap E) &\leq P(G^c \cap \{X_T \leq x_1\}) \leq P(G^c \mid X_t \leq x_1) \\ &\leq 2\exp(-(a-2)x_1/4(1+b(1+T))) \\ &\leq 2\exp(-(a-2)aM_1(1+T)^{1/r}/5). \end{aligned} \tag{4.17}$$

On G, using the definition of x_1 and the assumption $b \leq x$,

$$\sup_{T\leq t<\tau} X_t - (1+t)^{1/r}(M(a,T) + xM_2\log(1+t)) \leq 0, \quad \text{where} \quad M(a,T) = a^2xM_1M_2(1+T).$$

Since $a, M_2 \geq 1$, on $E\cap G$ the above inequality holds for all $t < \tau$. Let $a = T^{1/2}$ and let $M = M_1M_2$. Since $x = (1+T)^2$ by definition, $M(a,T) \leq TxM_1M_2(1+T) \leq xM(1+T)^2 = x^2M$, so on $E \cap G$, since $M_1 \geq 1$,

$$\sup_{t<\tau} X_t - Mx(1+t)^{1/r}(x + \log(1+t)) \leq 0.$$

Writing x as $(1+T)^2$, on $E\cap F\cap G$, for $T \leq t < \tau$

$$I_t \geq L(t) = e^{t^{1/2}/9} \quad \text{and} \quad X_t \leq M(1+T)^2(1+t)^{1/r}((1+T)^2 + \log(1+t)).$$

Since $\tau = \inf\{t > T : X_t > I_t\}$, if T is large enough then $L(t) \geq M(1+t)^{2+1/r}((1+T)^2 + \log(1+t) \geq M(1+T)^2(1+t)^{1/r}((1+T)^2 + \log(1+t)$ for all $t \geq T$ and so $\tau = \infty$ on $E\cap F\cap G$.

Using Lemmas 4.10 and 4.11, for large T,

$$\begin{aligned} P(E^c) &\leq 4e^{-T^{1/2}/2+1} \quad \text{and} \\ P(F^c) &\leq e^{-(1+T)^{1/2}/9}. \end{aligned}$$

Comparing to (4.17), the weakest bound is on F^c. So, for large T,

$$P((E\cap F\cap G)^c) \leq P(F^c) + P(E^c) + P(G^c\cap E) \quad \leq \quad 3e^{-(1+T)^{1/2}/9}$$

Let T_0 be large enough that above estimates hold for $T > T_0$ and let $x_0 = (1+T_0)^2$. The result is proved. □

4.2.2 Estimation of $T_{C,R}$

Write $T_{C,R} = T_C \wedge T_R$, where

$$\begin{array}{rcl} T_C & = & \inf\{t : \max_w \sum_{v \in \mathbf{C}_t(w)\cap H_t} N_t(v)^{-1} \geq C\} \quad \text{and} \\ T_R & = & \inf\{t : \max_w \mathbf{R}_t(w) \geq R\}. \end{array}$$

Proposition 4.13. *Let* $b = 1 + C$*. If* $b + (\log n)^{11} \leq R = o(n/\log n)$ *then*

$$\lim_{n\to\infty} P(T_R \leq n^{1/2} \wedge T_C) = 0.$$

Proposition 4.14. *For each* $\epsilon > 0$ *and with* $b = (27\log n)^4$*, if* $R \leq (\log n)^{12}$ *then*

$$\lim_{n\to\infty} P(T_C \leq n^{1/2-\epsilon} \wedge T_R) = 0.$$

Proof of Proposition 4.8. Notice that

$$T_{C,R} \leq t \quad \Leftrightarrow \quad T_C \leq t \wedge T_R \ \text{ or } \ T_R \leq t \wedge T_C,$$

then use Propositions 4.13 and 4.14 and take a union bound. □

Next we prove Proposition 4.13, which is the simpler of the two.

Proof of Proposition 4.13. For any w, $\mathbf{R}_t(w)$ is dominated by $\Lambda_t(X_t)$ on the time interval $[0, T_{C,R}]$.
Thus for any $t > 0$, a union bound gives

$$P(T_R \leq t \wedge T_C) \leq nP(\sup_{s\leq t} \Lambda_s(X_s) \geq R).$$

For any function $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$,

$$\{\sup_{s\leq t} \Lambda_s(X_s) \geq R\} \subset \{\sup_{s\geq 0} X_s - \Phi(s) > 0\} \cup \{\sup_{s\leq t} \Lambda_s(\Phi_s) \geq R\}. \tag{4.18}$$

Let $\Phi(s) = M(27\log n)^4((27\log n)^4 + \log(1+s))(1+s)^{1/r}$. Taking $x = (27\log n)^4$ in Proposition 4.9,

$$P(\sup_{s\geq 0} X_s - \Phi(s) > 0) = o(1/n). \tag{4.19}$$

We have the trivial bound $\Lambda_s(x) \leq b + x$, so since $s \mapsto \Phi(s)$ is non-decreasing, for any $T > 0$, $\sup_{s\leq T} \Lambda_s(\Phi(s)) \leq b + \Phi(T)$. Using (4.15) and (4.13),

$$\Lambda_s(x) \leq b + 2x(1+rs)^{-1} \text{ for } x \leq I_t.$$

Let $L(s) = e^{s^{1/2}/9}$. Taking $T = (18\log n)^2$, if n is large enough and $r \in [1/2, 1]$ then $\Phi(s) \leq L(s)$ for $s \geq T$. Using Lemma 4.11 for $s \geq T$ and the trivial bound for $s \leq T$,

$$P(\sup_{s\geq 0} \Lambda_s(\Phi(s)) - (b + \Phi((18\log n)^2) \vee (2\Phi(s)(1+rs)^{-1})) > 0) = o(1/n). \tag{4.20}$$

Since $R = o(n)$ by assumption, $r \geq 1/2$ for large n and so $\Phi((18\log n)^2) = O((\log n)^{10})$.
Since $R = o(n/\log n)$ by assumption, for large n, $r = 1 - R/(n-1) \geq 1 - o(1/\log n)$ so $1/r - 1 = o(1/\log n)$ and for $s \leq n^{1/2}$,

$$(1+s)^{1/r}/(1+rs) = O((1+s)^{1/r-1}) = O(n^{o(1/\log n)}) = O(1),$$

so

$$\sup_{s\leq n^{1/2}} 2\Phi(s)(1+rs)^{-1} = O((\log n)^8).$$

Since $R \geq b + (\log n)^{11}$ by assumption, which is at least $b + \Phi((18\log n)^2) \vee (2\Phi(s)(1+rs)^{-1}))$ for large n and $s \leq n^{1/2}$, the result follows from (4.18),(4.19) and (4.20) with $t = n^{1/2}$. □

It remains to prove Proposition 4.14. Define the non-decreasing spacetime set of points

$$\mathcal{A}_t(w) = \left\{(v,s)\colon s\le t \text{ and either } \begin{array}{l} v\in \mathbf{C}_{s^-}(w) \text{ and agreement occurs at } (v,\cdot,s) \text{ or } (\cdot,v,s), \text{ or} \\ v\in H_s(w)\cap \mathbf{C}_s(w)\setminus \mathbf{C}_{s^-}(w). \end{array}\right\}$$

To get a more workable quantity we will use the fact that

$$\sum_{v\in \mathbf{C}_t(w)\cap H_t} N_t(v)^{-1} \le C_t(w) = \sum_{(v,s)\in\mathcal{A}_t(w)} 1/(1+N^{\ell}_{t-s}(v)).$$

This way,

$$\text{if } \sup_{s\le t}\max_w C_s(w) < C \text{ then } T_C > t. \tag{4.21}$$

So, to estimate T_C we control contributions to $C_t(w)$. Let $Q_t(w)$ denote the rate at which $\mathcal{A}_t(w)$ increases. Let $\{N^i_t : t\ge 0,\ i\ge 1\}$ be an independent collection of Poisson processes with intensity 1, let $Q,T>0$ and let $N(t)$ be an independent Poisson process with intensity Q. Let $t_i=\inf\{t : N(t)=i\}$ and let

$$B_t = \sum_{i\le N(t)} 1/(1+N^i_{t-t_i}).$$

Let $T_Q(w)=\inf\{t : Q_t(w)>Q\}$ and $T_Q=\min_w T_Q(w)$. Then for any w,

$$(C_t(w))_{t\le T_Q} \text{ is stochastically dominated by } (B_t). \tag{4.22}$$

In the next lemma we control B_t.

Lemma 4.15. *Fix $T>T_0\ge 1$ and $Q\ge 1$. Then,*

$$P(\sup_{t\le T} B_t > 2QT_0 + 4Q\log(2\vee T)) \le Q(2+T)^2 e^{-T_0/16}.$$

Proof. We first bound B_t over intervals $[k-1,k]$, $k\in\mathbb{Z}\cap[0,1+T]$, then take a union bound to control the value over the interval $[0,T]$. Fix $k\le T+1$ and let $\tilde N(t)=N(k)-N(k-t)$. Define

$$\begin{aligned} E_1 &= \{\tilde N(t)\ge 2Q(t\vee T_0) \text{ for some } t\le k\} \quad \text{and} \\ E_2 &= \{N^i_t < t/2 \text{ for some } t\ge T_0 \text{ and } 2QT_0 < i \le 2Qk\}. \end{aligned}$$

Note that $P(E_2=0)$ if $k\le T_0$, since the range of i values is empty. Using Lemma 3.3 with $X_t=\tilde N(t)$, $X^p_t=\langle X^m\rangle_t = Qt$, $c=1$, $a=QT_0/2$ and $\lambda=1/2$,

$$P(E_1)\le 2e^{-QT_0/4}\le 2e^{-T_0/4}.$$

Using Lemma 3.3 with $X_t=N^i_t$, $X^p_t=\langle X^m\rangle_t=t$, $a=T_0/4$ and $\lambda=1/4$ and taking a union bound,

$$P(E_2)\le 2Q(0\vee(k-T_0))e^{-T_0/16}.$$

Order the jump times of $\tilde N(t)$ for $t\le k$ in increasing order as $\tilde t_1,\tilde t_2,\dots,\tilde t_{\tilde N(k)}$.
On E_1^c, $\tilde t_i > i/2Q$ for $2QT_0 < i \le 2Qk$, so on $E_1^c\cap E_2^c$, $N^i_{\tilde t_i -1}\ge (i/2Q-1)/2$ for $2QT_0 <$

$i \le 2Qk$ and

$$\begin{aligned}\sup_{t\in[k-1,k]} B_t &\le 2QT_0 + \sum_{i=2QT_0+1}^{2Qk} 1/(1+(i/2Q-1)/2) \\ &\le 2QT_0 + \int_{2QT_0}^{2QT} (1+(t/2Q-1)/2)^{-1}dt \\ &= 2QT_0 + 4Q(\log(1/2+k/2) - \log(1/2+T_0/2)) \\ &\le 2QT_0 + 4Q\log(k),\end{aligned}$$

using $k, T_0 \ge 1$ on the last step so that $1/2+k/2 \le k$ and $\log(1/2+T_0/2) \ge 0$. Summarizing,

$$\begin{aligned}&P(B_t > 2QT_0 + 4Q\log k \text{ for some } t \in [k-1,k]) \\ &\le 2Q(0 \vee (k-T_0))e^{-T_0/16} + 2e^{-T_0/4} \le 2Qke^{-T_0/16}\end{aligned}$$

since $2e^{-T_0/4} \le 2QT_0e^{-T_0/16}$ if $k > T_0$ and $2e^{-T_0/4} \le 2Qke^{-T_0/16}$ if $k \le T_0$, noting $k, Q \ge 1$ by assumption. Taking a union bound over $k = 1, 2, \dots, \lfloor T\rfloor, \lfloor T\rfloor + 1$ and noting $\sum_{k=1}^{\lfloor T\rfloor+1} k \le (2+T)^2/2$ gives the result. □

It remains to prove the following result.

Proposition 4.16. *If $\epsilon > 0$ is small, $k > 0$ is fixed, $Q = \log n$ and $R \le (\log n)^k$, then*

$$\lim_{n\to\infty} P(T_Q \le n^{1/2-\epsilon} \wedge T_{C,R}) = 0.$$

Before proving it, we show how it implies Proposition 4.14. Recall that whp (with high probability) refers to an event whose probability tends to 1 as $n \to \infty$. Note that if E_1, E_2 whp then $E_1 \cap E_2$ whp.

Proof of Proposition 4.14. We want to show that $T_C > n^{1/2-\epsilon} \wedge T_R$ whp. Since $T_C \ge T_C \wedge T_Q$, if $T_C \wedge T_Q > n^{1/2-\epsilon} \wedge T_R$ then $T_C > n^{1/2-\epsilon} \wedge T_R$. Moreover

$$T_C \wedge T_Q > n^{1/2-\epsilon} \wedge T_R \Leftrightarrow T_C > n^{1/2-\epsilon} \wedge T_R \wedge T_Q \quad \text{and} \quad T_Q > n^{1/2-\epsilon} \wedge T_R \wedge T_C.$$

Proposition 4.16 says that $T_Q > n^{1/2-\epsilon} \wedge T_R \wedge T_C$ whp, so it is enough to show that if $b = 1 + C = (27\log n)^4$ and $Q = \log n$ then $T_C > n^{1/2-\epsilon} \wedge T_R \wedge T_Q$ whp, or equivalently that

$$P(T_C \le n^{1/2-\epsilon} \wedge T_R \wedge T_Q) = o(1).$$

In Lemma 4.15 take $T = n^{1/2}$, $T_0 = 48\log n$ and $Q = \log n$ to find that

$$P(\sup_{t\le n^{1/2}} B_t > 98(\log n)^2) = O(n^{-2}\log n) = o(1/n).$$

Then, using (4.22) and Proposition 4.16 and taking a union bound over the n possible values of w,

$$P(\sup_{t\le n^{1/2-\epsilon}\wedge T_R\wedge T_Q} \max_w C_t(w) > 98(\log n)^2) = o(1).$$

The result then follows from (4.21) and the fact that $98(\log n)^2 < (27\log n)^4 - 1$ for large n. □

By taking a union bound over w and noting the probability does not depend on w, to obtain Proposition 4.16 it is sufficient to show that for any w and small $\epsilon > 0$,

$$\lim_{n\to\infty} P(\sup_{t\le n^{1/2-\epsilon}} Q_t(w) > \log n) = o(1/n), \tag{4.23}$$

noting that the probability is the same for any w. There are three ways that $\mathcal{A}_t(w)$ increases:

1. a site already in H_t is added to $\mathbf{C}_t$,
2. agreement occurs at a site already in $\mathbf{C}_t$, or
3. a site is added simultaneously to $\mathbf{C}_t$ and H_t.

Let $Q_t^i(w),\ i = 1, 2, 3$ denote the rate of each event, so that $Q_t(w) = \sum_{i=1}^3 Q_t^i(w)$. Since each site in $V \setminus \mathbf{C}_t(w)$ is added to $\mathbf{C}_t(w)$ at rate $\mathbf{R}_t(w)/(n-1) \le R/(n-1)$,

$$Q_t^1(w) \le |H_t|R/(n-1). \tag{4.24}$$

Since there are $|\mathbf{C}_t(w) \cap \mathcal{C}_t(v)|$ sites in $\mathbf{C}_t(w)$ that can agree on word v, and each word is spoken at rate at most $R/(n-1)$ to each site,

$$\begin{aligned} Q_t^2(w) &\le \textstyle\sum_v |\mathbf{C}_t(w) \cap \mathcal{C}_t(v)|R/(n-1) \\ &\le \textstyle\sum_{v \ne w} |\mathbf{C}_t(w) \cap \mathcal{C}_t(v)|R/(n-1) + S_t R/(n-1), \end{aligned} \tag{4.25}$$

recalling that $S_t = \max_w |\mathcal{C}_t(w)|$ is the size of the largest cluster. Each time a person speaks, the probability that agreement occurs is at most $S_t/(n-1)$. Since $\mathbf{C}_t(w)$ increases at rate $\le R$, it follows that

$$Q_t^3(w) \le S_t R/(n-1). \tag{4.26}$$

The reader may think that $Q_t^3(w)$ should be 0, since a new addition to a cluster does not yet know the word. However, the upper bound cluster $\mathbf{C}_t(w)$ can grow when in the process itself, a word other than w is being spoken. Using Proposition 4.7 we control $Q_t^1(w)$ and $Q_t^3(w)$, which is two thirds of Proposition 4.16.

Lemma 4.17. *For each w, small $\epsilon > 0$, $R \le n^\epsilon$, $b = 1 + C \le (27 \log n)^4$ and $i = 1, 3$,*

$$P(\sup_{t \le n^{1/2-\epsilon} \wedge T_{C,R}} Q_t^i(w) > 1) = o(1/n).$$

Proof. From (4.24) and (4.26) and the choice of R, it suffices to show that

$$P(\sup_{t \le n^{1/2-\epsilon} \wedge T_{C,R}} \max S_t, |H_t| > n^{1-\epsilon}) = o(1/n).$$

Using Proposition 4.7,

$$P(\sup_{t \le T_{C,R}} S_t - \Phi(t,x) > 0) = o(1/n),$$

where $\Phi(t,x) = (\log n)^9 (1+t)^{1+\epsilon}$. The desired result for $i = 3$ then follows from (4.26), since $\sup_{t \le n^{1/2-\epsilon}} (\log n)^9 (1+t)^{1+\epsilon} = (\log n)^9 (1 + n^{1/2-\epsilon})^{1+\epsilon} = o(n^{1-\epsilon})$. To get the result for $i = 1$ recall from the beginning of this section that $|H_t| \le 2A_t$, the number of agreements up to time t, and from (4.6) that $A_t \le \text{Poisson}(\int_0^u S_u du)$. Using the above bound on S_t, with probability $1 - o(1/n)$,

$$\begin{aligned} \int_0^{n^{1/2-\epsilon} \wedge T_{C,R}} S_u du &\le \int_0^{n^{1/2-\epsilon}} \Phi(u,x) du \\ &= (\log n)^9 \tfrac{1}{2+\epsilon}((1 + n^{1/2-\epsilon})^{2+\epsilon}) - 1) = o(n^{1-\epsilon}) \end{aligned}$$

for large n. From Lemma 5.11, $P(\text{Poisson}(\lambda) > 2\lambda) \le e^{-\lambda/3}$. The above implies that wp $1 - o(1/n)$, H_t is dominated by Poisson$(n^{1-\epsilon}/2)$ for large n. So, it follows that

$$P(\sup_{t \le n^{1/2-\epsilon} \wedge T_{C,R}} |H_t| > n^{1-\epsilon}) \le e^{-n^{1-\epsilon}/6} + o(1/n) = o(1/n). \qquad \square$$

It remains to control $|\mathbf{C}_t(w) \cap \mathcal{C}_t(v)|$. First we modify slightly the construction from the beginning of Section 4.2, using a randomization trick. The reason it needs modifying is to ensure the growth of $\mathbf{C}_t(w)$ and any $\mathcal{C}_t(v)$ are not strongly correlated. Since we only randomize the location of "exceptional" events that expand $\mathbf{C}_t(w)$, the reader may verify that up to a random permutation of certain vertices, the marginal distribution of each $\mathbf{C}_t(w)$, and its domination of $\mathcal{C}_t(w)$, are unchanged.

To carry out the modification, make the $\{U_v\}$ doubly-augmented, that is, each U_v is again a Poisson point process with intensity 1, but on $[0,\infty)\times[0,1]^2$ instead of $[0,\infty)\times[0,1]$. $\mathbf{R}_t(w)$ is defined in the same way as before, and $\mathbf{C}_t(w)$ is defined as follows.

$$\begin{array}{ll} \text{if} & (t,u_1,u_2)\in U_v \text{ and } u_1\in I_{t^-}(v,i,R_{t^-}(w)/(n-1)), \\ \text{or if} & (t,u_1,u_2)\in U_v, u_1\notin I_{t^-}(v,i,R_{t^-}(w)/(n-1)) \\ & \text{and } u_2\le (n-1)^{-1}(\mathbf{R}_{t^-}(w)-R_{t^-}(w))/(1-R_{t^-}(w)), \\ \text{then} & \mathbf{C}_t=\mathbf{C}_{t^-}\cup\{v\}. \end{array}$$

In other words,

- if $\mathcal{C}_t(w)$ was about to include v, then $\mathbf{C}_t(w)$ will too, and
- if $\mathbf{C}_t(w)$ increases when $\mathcal{C}_t(w)$ does not, then with respect to what other clusters are doing, it does so as randomly as possible.

We now control the size of $\mathbf{C}_t(w)\cap\mathcal{C}_t(v)$, for any $v\ne w$. "wp" is shorthand for "with probability".

Lemma 4.18. *For any $\epsilon,k>0$, if $R\le n^{\epsilon/4}$ then*

$$P\Big(\sup_{t\le n^{1/2-\epsilon}\wedge T_{C,R}}\sum_{v\ne w}|\mathbf{C}_t(w)\cap\mathcal{C}_t(v)|\ge n/(\log n)^k\Big)=o(1/n).$$

Proof. Let $K_t=\sum_{v\ne w}|\mathbf{C}_t(w)\cap\mathcal{C}_t(v)|$. There are three ways K_t can increase.

1. $\mathbf{C}_t(w)$ acquires a site that belongs to some (possibly many) $\mathcal{C}_t(v)$, $v\ne w$,
2. some $\mathcal{C}_t(v)$, $v\ne w$ acquires a site that belongs to $\mathbf{C}_t(w)$, and
3. $\mathbf{C}_t(w)$ and some $\mathcal{C}_t(v)$, $v\ne w$ simultaneously acquire the same site.

It suffices to show the contribution to $\sup_{t\le n^{1/2-\epsilon}\wedge T_{C,R}}K_t$ from each item is $o(n/(\log n)^k)$ wp $1-o(1/n)$. For item 1, the jump size is at most $\max_v N_t^\ell(v)$, while for items 2,3 the jump size is 1. Let $R_1(t),R_2(t),R_3(t)$ denote the rate of each event. Then,

$$\begin{array}{rcl} R_i(t) & \le & \mathbf{R}_t(w) \text{ for } i\in\{1,3\}, \text{ and} \\ R_2(t) & \le & |\mathbf{C}_t(w)|\sum_{v\ne w}R_t(v)/(n-1). \end{array}$$

Since $N_t^\ell(v)\le N_{n^{1/2}}^\ell(v)$ for $t\le n^{1/2}$ and each v, and since each $N_{n^{1/2}}^\ell(v)\sim\text{Poisson}(n^{1/2})$, using Lemma 5.11 with $x=n^{1/4}$ and a union bound,

$$P\Big(\sup_{t\le n^{1/2}}\max_v N_t^\ell(v)>2n^{1/2}\Big)\le ne^{-n^{1/2}/3}=o(1/n).$$

For $t<T_{C,R}$, $\mathbf{R}_t(w)\le R\le n^{\epsilon/4}$ by assumption, so wp $1-o(1/n)$, the contribution from item 1 is at most $2n^{1/2}\text{Poisson}(Rn^{1/2-\epsilon})\le 2n^{1/2}\text{Poisson}(n^{1/2-3\epsilon/4}$ which (using again Lemma 5.11 with $x=\lambda$) is wp $1-o(1/n)$ at most $4n^{1-3\epsilon/4}=o(n/(\log n)^k)$ for any fixed k. This also bounds the contribution from item 3 since the rate has the same bound and the jump size is 1.

For item 2, note that $\sum_{v\ne w}R_t(v)\le\sum_v R_t(v)=n$ and that for $t<T_{C,R}$, $\mathbf{C}_t(w)$ is dominated by X_t. Applying Proposition 4.9, bounding $\log(1+t)$ by $\log(1+n^{1/2})$ for $t\le n^{1/2}$ and using the trivial but convenient $n/(n-1)\le 2$ for $n\ge 2$, we find that for $x\ge b=1+C$ large enough,

$$P\Big(\sup_{t<n^{1/2}\wedge T_{C,R}}R_2(t)\ge 2Mx(x+\log(1+n^{1/2}))(1+t)^{1/r}\Big)\le 3e^{-x^{1/4}/9}.$$

Taking $x = (27\log n)^4$ the probability above is $o(1/n^2)$. Thus the contribution from item 2 is at most Poisson$(f(n^{1/2-\epsilon}))$, where

$$\begin{aligned} f(t) &= \int_0^t 2M(27\log n)^4((27\log n)^4 + \log(1+n^{1/2}))(1+s)^{1/r}ds \\ &\leq 4M(27\log n)^8(1+t)^{1+1/r}, \end{aligned}$$

using $\log(1+n^{1/2}) \leq (27\log n)^4$ and $1+1/r \geq 1$. If $R = o(n)$ then for any $\epsilon > 0$, $1/r \leq 1+\epsilon$ for large n. Therefore

$$f(n^{1/2-\epsilon}) = O((\log n)^8 n^{(1/2-\epsilon)(2+\epsilon)}) = O((\log n)^8 n^{1-3\epsilon/2-\epsilon^2}) = o(n/(\log n)^k).$$

It follows as before that Poisson$(f(n^{1/2-\epsilon})) = o(n/(\log n)^k)$ wp $1-o(1/n)$, and the proof is complete. □

Combining this with the other term in (4.25) we control $Q_t^2(w)$.

Lemma 4.19. *For any $k > 0$ and small $\epsilon > 0$, each w and $R \leq (\log n)^k$,*

$$P(\sup_{t \leq n^{1/2-\epsilon}\wedge T_{C,R}} Q_t^2(w) > 2) = o(1/n).$$

Proof. From the proof of Lemma 4.17 we know that $P(\sup_{t\leq n^{1/2-\epsilon}\wedge T_{C,R}} S_t > n^{1-\epsilon}) = o(1/n)$. Using this, (4.25), $R \leq (\log n)^k$ and Lemma 4.18,

$$P(\sup_{t \leq n^{1/2-\epsilon}\wedge T_{C,R}} Q_t^2(w) > (n/(\log n)^k) + n^{1-\epsilon})(\log n)^k/(n-1)) = o(1/n).$$

If n is large then $(n/(\log n)^k + n^{1-\epsilon})(\log n)^k/(n-1) \leq 2$ and the result follows. □

Proof of Proposition 4.16. This follows from (4.23), and Lemmas 4.17 and 4.19. □

5 Final phase

Change of variables. Recall (x, y, z) defined in (2.1), and let $u_t = |x_t - y_t|$, then for T_c from Theorem 2.2 we have $T_c = \inf\{t : u_t = 1\}$. Conveniently, (u, z) has a closed system of approximating ODEs. Take $x' - y'$ in (2.4) to obtain

$$u' = \text{sgn}(x-y)(x-y)' = \text{sgn}(x-y)(x-y)z = uz.$$

The above assumes sgn$(x-y)$ does not change along trajectories of the ODEs, but this is confirmed by the resulting equation. Next, since $z = 1-(x+y)$,

$$z' = -(x'+y') = -(x+y)z - 2z^2 + 2xy = -(1-z)z - 2z^2 + 2xy = -z - z^2 + 2xy.$$

The $2xy$ part should be written in terms of u, z. To do so note

$$z^2 = 1 - 2(x+y) + (x+y)^2 = 2z - 1 + (x+y)^2,$$

so $z^2 - 4xy = 2z - 1 + (x-y)^2 = 2z - 1 + u^2$. Thus $4xy = 1 - u^2 - 2z + z^2$ and so

$$z' = \frac{1}{2}(-2z - 2z^2 + 4xy) = \frac{1}{2}(1 - u^2 - 4z - z^2).$$

Summarizing, we have the following system of approximating ODEs for (u, z):

$$\begin{aligned} u' &= uz \\ z' &= \frac{1}{2}(1 - u^2 - 4z - z^2). \end{aligned} \tag{5.1}$$

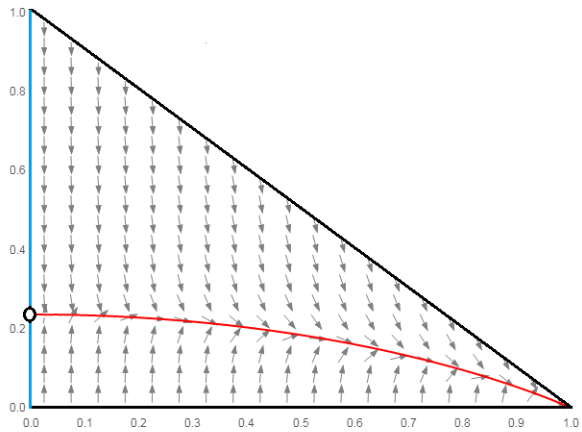

Figure 4: Phase portrait for the ODE system (5.1) in the (u, z) plane. Image generated using Darryl Nester's applet on `https://bluffton.edu/homepages/facstaff/nesterd/java/slopefields.html`

The direction field and a representative trajectory are depicted in Figure 4. The figure coincides with the portion of Figure 3 above the line $x = y$, rotated clockwise by an angle of $3\pi/4$. There are two relevant equilibria: the saddle point $(0, z^*)$ with z^* the positive solution of $z^*(4 + z^*) = 1$, and the stable equilibrium $(1, 0)$. Solutions approach the unstable manifold (red) of $(0, z^*)$ from above and below, then are attracted to the equilibrium $(1, 0)$. Approach to the red line is fast; the slow movemenet is near the equilibria.

We can attempt estimation of T_c is as follows. If $r > 0$ is a small fixed distance, then to reach an r-neighbourhood (nbhd) of the red line, and to go from within an r-nbhd of $(0, z^*)$ to within an r-nbhd of $(1, 0)$ takes $O(1)$ time. Linearizing around $(0, z^*)$ gives $u' = z^*u$ and $z' = 0$, so for u to go from a value of $1/n$ up to r takes $(1/z^*)\log(n/r) = (1/z^*)\log(n) - O(1)$ time. Linearizing around $(1, 0)$ gives

$$\begin{pmatrix} (1-u)' \\ z' \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} (1-u) \\ z \end{pmatrix},$$

which has eigenvalues $-1, -1$, so the approach to $(1, 0)$ is like e^{-t} or te^{-t}. Thus, to go from distance r of $(1, 0)$ to distance less than $1/n$ takes about $\log(n/r) = \log(n) - O(1)$ time. In total, this gives the estimate $T_c \sim (1+1/z^*)\log n$ when $(u_0, z_0) \approx (0, z^*)$. However, this disagrees with the statement of Theorem 2.2 by an amount $(1/2z^*)\log n$. To see why we need to include fluctuations.

Drift and diffusivity. Here we compute the drift and diffusivity of (u, z). The drift is the same as the right-hand side of (5.1) up to some $o(1)$ terms, but the diffusivity gives information which is unavailable from the deterministic approximation. We use the notation of Section 3. From (3.4), and indexing the rates q_i and jumps Δ_i in Table 2 by their row i, we have

$$\mu(x) = \sum_i q_i \Delta_i(x) \quad \text{and} \quad \sigma^2(x) = \sum_i q_i \Delta_i^2(x), \tag{5.2}$$

and similarly for y, z, u. It will be convenient for computations to slow time by a factor $(n-1)/n$, so that transition rates in the table are equal to q/n instead of $(n-1)q/n^2$.

Using Table 2 with q/n in place of $(n-1)q/n^2$ and using (5.2), we compute

$$\mu(x) = xz + z^2 - xy - z/n \quad \text{and}$$
$$\mu(y) = yz + z^2 - xy - z/n,$$

so following the approach used to obtain (5.1) we find $\mu(z) = \frac{1}{2}(1 - u^2 - 4(1-1/n)z - z^2)$. The computation of $\mu(u)$ is complicated by the absolute value. However, from Table 2 we have $\Delta_\star(u) \leq 2/n$, so if $u \geq 2/n$ then $\text{sgn}(u)$ does not change after a jump, which easily implies that

$$\text{if } \; u \geq 2/n \; \text{ then } \quad \mu(u) = \text{sgn}(x-y)\mu(x-y) = \text{sgn}(x-y)(x-y)z = uz.$$

On the other hand, since $\Delta u \geq |\Delta x - \Delta y|$, we always have the inequality $\mu(u) \geq |\mu(x) - \mu(y)| = |(x-y)z| = uz$. Altogether, this gives

$$\mu(z) = \frac{1}{2}(1 - u^2 - 4(1-1/n)z - z^2) \; \text{ and} \tag{5.3}$$
$$\mu(u) \begin{cases} = uz & \text{if } \; u \geq 2/n, \\ \geq uz & \text{if } \; u < 2/n. \end{cases}$$

We can obtain coarse upper bounds on the magnitude of the drift and diffusivity as follows. Note the total jump rate before time change is equal to n (each vertex speaks at rate 1), the jump size of each of x, y, z, u is at most $2/n$, and the time change only slows things down. Thus,

$$|\mu(x)| \leq n(2/n) = 2, \;\; \sigma^2(x) \leq n(2/n)^2 = 4/n \; \text{ and similarly for } \; y, z, u. \tag{5.4}$$

For u, z we can obtain a tighter bound on the diffusivity. The total speaking rate of $n-1$ (after time change) includes interactions of type $A+A$ and $B+B$, that have no effect on u, z. The rates of these interactions are respectively

$$nx \, \frac{nx-1}{n-1} \, \frac{n-1}{n} = x(nx-1) \quad \text{and} \quad y(ny-1).$$

Thus the rate of interactions affecting u, z is at most $n - 1 - n(x^2+y^2) + x + y = n(1 - (x^2+y^2)) - z$. Since $x^2 + y^2 = u^2 + 2xy \geq u^2$, an easy upper bound is $n(1-u^2) \leq 2n(1-u)$. Since $|\Delta u|, |\Delta z| \leq 2/n$ at each jump, it follows that $\sigma^2(u), \sigma^2(z) \leq 8(1-u)/n$.

To get a matching lower bound, note that interactions of type $A+!A$ and $B+!B$, with ! meaning "not", has $|\Delta u|, |\Delta z| \geq 1/n$, and the rate of such interactions is $nx(1-x) + ny(1-y)$. If $x \geq y$ then since $u = x - y$ and $y = 1 - (x+z) \leq 1 - x$, $1 - u = 1 - x + y \leq 2(1-x)$, so $1 - x \geq (1-u)/2$, and similarly $1 - y \geq (1-u)/2$ if $y \geq x$. Thus, $nx(1-x) + ny(1-y) \geq n(1-u)/2$, and we obtain $\sigma^2(u), \sigma^2(z) \geq (1-u)/2n$. We record the two estimates for later:

$$(1-u)/2 \leq n\sigma^2(z), n\sigma^2(u) \leq 8(1-u). \tag{5.5}$$

Determination of T_c. From (5.5) we see that when u is bounded below 1, (u, z) has fluctuations of order $1/\sqrt{n}$ on constant time scale. Including this in our previous estimation of T_c, we see that to escape from $(0, z^*)$, u only needs to go from $1/\sqrt{n}$ up to r, which takes time $(1/z^*)\log(\sqrt{n}/r) = (1/2z^*)\log(n) - O(1)$. Including the $\log(n)$ time needed to close in on $(1, 0)$ gives the correct estimate $T_c \sim (1 + 1/2z^*)\log n$.

To make these observations rigorous and to compute the growth of T_c as a function of n, we introduce two small parameters: $\delta > 0$ for small values of $z, |z - z^*|$ and $\epsilon > 0$ for small values of $u, 1-u$, and one large parameter $B > 0$ for the diffusion of $\sqrt{n}u$. We will focus on the following sequence of desirable events:

(i) wait until either $z \ge 2\delta$ or $u \ge 1-\epsilon$,
(ii) $z > \delta$ until either $u > 2\epsilon$ or $|z-z^*| \le \delta$,
(iii) $|z-z^*| \le 2\delta$ until $\sqrt{n}u > 2B$,
(iv) $|z-z^*| \le 2\delta$ and $\sqrt{n}u > B$ until $u > 2\epsilon$,
(v) $u > \epsilon$ and $z > \delta$ until $u \ge 1-\epsilon$,
(vi) $u \ge 1-4\epsilon$ until $u = 1$.

To encode these six events we use the following notation.
For an $\mathbb{R}$-valued process X, $x \in \mathbb{R}$ and $s \ge 0$ define

$$\tau_x^+(X,s) = \inf\{t \ge s\colon X_t > x\} \quad \text{and} \quad \tau_x^-(X,s) = \inf\{t \ge s\colon X_t \le x\},$$

and let $\tau_x^+(X) = \tau_x^+(X,0)$ and $\tau_x^-(X) = \tau_x^-(X,0)$. Also, to avoid crowding notation we'll use $X(t)$ interchangeably with X_t for any process X. We then define the following succession of five stopping times:

$$\begin{aligned}
T_1 &= \tau_{2\delta}^+(z) \wedge \tau_\epsilon^-(1-u), \\
T_2 &= \tau_\delta^-(z,T_1) \wedge \tau_\delta^-(|z-z^*|,T_1) \wedge \tau_{2\epsilon}^+(u,T_1), \\
T_3 &= \tau_{2\delta}^+(|z-z^*|,T_2) \wedge \tau_{2B}^+(\sqrt{n}u,T_2), \\
T_4 &= \tau_{2\delta}^+(|z-z^*|,T_3) \wedge \tau_B^-(\sqrt{n}u,T_3) \wedge \tau_{2\epsilon}^+(u,T_3), \\
T_5 &= \tau_\delta^-(z,T_4) \wedge \tau_\epsilon^-(u,T_4) \wedge \tau_\epsilon^-(1-u,T_4), \\
T_6 &= \tau_{4\epsilon}^+(1-u,T_5) \wedge \tau_0^-(1-u,T_5).
\end{aligned}$$

Clearly $T_1 \le \cdots \le T_6$. The next result says the above sequence of desirable events is the most likely outcome.

Proposition 5.1. *If constants $\epsilon, \delta, 1/B > 0$ are small enough and $\ldots$ then each of the following occurs with high probability as $n \to \infty$.*

(a) $u(T_2) > 2\epsilon$ *or* $|z(T_2) - z^*| \le \delta$.
(b) $\sqrt{n}u(T_3) > 2B$.
(c) $u(T_4) > 2\epsilon$.
(d) $u(T_5) \ge 1-\epsilon$.
(e) $u(T_6) = 1$.

We then need to estimate how much time passes between each event. For $i \in \{1,\ldots,6\}$ let $W_i = T_i - T_{i-1}$. The next result gives estimates on the times W_i.

Proposition 5.2. *Let W_i be as above. Fix any $\alpha > 0$ and any initial distribution. If constants $\epsilon, \delta, 1/B > 0$ are small enough and $\ldots$ then*

(a) $P(W_i > (\alpha/6)\log n) \to 0$ *as* $n \to \infty$ *for every* $i \in \{1,2,3,5\}$.
(b) $P(W_4 > (1/2z^* + \alpha/6)\log n \to 0$ *as* $n \to \infty$.
(c) $P(W_4 < (1/2z^* - \alpha/2)\log n \mid \sqrt{n}u(0) \le 2B$ *and* $|z(0) - z^*| \le \delta) \to 0$ *as* $n \to \infty$.
(d) $P(W_6 > (1+\alpha/6)\log n)) \to 0$ *as* $n \to \infty$.
(e) $P(W_6 < (1-\alpha/2)\log n \mid u(0) < 1-\epsilon) \to 0$ *as* $n \to \infty$.

Theorem 2.2 then follows easily from the above two propositions.

Proof of Theorem 2.2. If $2B/\sqrt{n} < 2\epsilon < 1-\epsilon$ it is easy to check that $T_6 \le T_c$. From Proposition 5.1, whp $u(T_6) = 1$ which implies $T_c \le T_6$, so whp $T_c = T_6 = W_1 + \cdots + W_6$. Using the events in Proposition 5.2 and taking a union bound,

$$\begin{aligned}P(T_c > (1+1/2z^*+\alpha)\log n) \le &\sum_{i\in\{1,2,3,5\}} P(W_i > (\alpha/6)\log n)\\ &+ P(W_4 > (1/2z^*+\alpha/6)\log n)\\ &+ P(W_6 > (1+\alpha/6)\log n)\\ &+ P(T_c \neq W_1+\cdots+W_6)\\ &\to 0 \ \text{ as } \ n\to\infty,\end{aligned}$$

which proves statement (i). If $|x(0)-y(0)| = O(1/\sqrt{n})$ then $\sqrt{n}u(0) \le B$ for large enough constant $B>0$ and large n, and if $|z(0)-z^*| = o(1)$ then $|z(0)-z^*| \le \delta$ for any constant $\delta > 0$ and large n. Using again Proposition 5.2 and taking a union bound as before gives

$$\begin{aligned}&P(T_c < (1+1/2z^*-\alpha)\log n \mid |x(0)-y(0)| = O(1/\sqrt{n}) \ \text{ and } \ |z(0)-z^*| = o(1))\\ &\le P(W_4 < (1/2z^*-\alpha/2)\log n \mid \sqrt{n}u(0) \le 2B \text{ and } |z(0)-z^*| \le \delta)\\ &+ P(W_6 < (1-\alpha/2)\log n \mid u(0) < 1-2\epsilon)\\ &+ P(T_c \neq W_1+\cdots+W_6)\\ &\to 0 \ \text{ as } \ n\to\infty.\end{aligned}$$

Combining with the previous estimate and noting that $\alpha > 0$ is arbitrary proves statement (ii). □

Proof of Propositions 5.1 and 5.2

For technical reasons it's easier to let z^* be the positive solution to $z^*(4(1-1/n)+z^*) = 1$. Each of the next six lemmas corresponds respectively to the times $T_1,\dots,T_6$. A few notes on the proofs:

- The key estimates are given in statements (i), (ii) and occasionally (iii) of each lemma, in terms of conditions on initial values. Estimates on the relevant T_i and W_i are then deduced.
- When deducing estimates on T_i, W_i for $i > 1$, the strong Markov property is used without mention.
- Since the process is a finite state Markov chain, if τ is a hitting time such that $P(\tau = 1)$, then τ has an exponential tail, so optional stopping can be invoked at time τ.
- When we say "fix a small $\epsilon > 0$" or "fix a large $B > 0$", we mean "if the constant $\epsilon > 0$ is chosen small enough" or "if the constant $B > 0$ is chosen large enough".
- If conditions are given on initial values, it means the statement holds for any initial distribution satisfying those conditions. If none are given, it holds for any initial distribution.

Lemma 5.3. *Fix small $\epsilon, \delta > 0$ such that $\epsilon \ge 18\delta$. Then*

(i) $E[T_1] \le 12\delta/\epsilon$ *and*
(ii) $P(\tau_\delta^-(z) \le \tau_\epsilon^-(1-u) \wedge \exp(\Omega(n)) \mid z(0) \ge 2\delta) = \exp(-\Omega(n))$.

In particular, for any $\alpha > 0$, $P(W_1 > (\alpha/6)\log n) \to 0$ as $n \to \infty$.

Proof. Recall $T_1 = \tau_{2\delta}^+(z) \wedge \tau_\epsilon^-(1-u)$.

Statement (i). If $z < 2\delta$, $u < 1-\epsilon$ and $\delta \le 1$ then $4(1-1/n)z + z^2 \le 8\delta + \delta^2 \le 9\delta$ and $1-u^2 = (1-u)(1+u) > \epsilon$, so if in addition $\epsilon \ge 18/\delta$ then $\mu(z) \ge (\epsilon - 9\delta)/2 \ge \epsilon/4$, so the process with $X(t) = z(t \wedge T_1) - (\epsilon/4)(t \wedge T_1)$ is a submartingale. Using optional stopping,

$$E[z(0)] = E[X(0)] \le E[X(T_1)] = E[z(T_1)] - (\epsilon/4)E[T_1],$$

so $E[T_1] \le (4/\epsilon)E[z(T_1) - z(0)]$. Since $z(T_1) \le 2\delta + 2/n \le 3\delta$ for large n and $z(0) \ge 0$, $E[T_1] \le 12\delta/\epsilon$, which is statement (i).

Statement (ii). From (5.4), $|\mu(z)| \le 2$ and $\sigma^2(z) \le 4/n$, use Lemma 3.6 with $X = 2\delta - z$, $x = \delta$, $\tau = \tau_\epsilon^-(1-u, 0)$, $\mu_\star = \epsilon/4, \sigma_\star = 4/n$, $C_\mu = 2$ and $C_\Delta = 1/2$. Since $\Delta_\star(z) = 2/n$, $\Delta_\star(z) \le x/2$ for large n and $\Delta_\star(z)\mu_\star/\sigma_\star^2 = \epsilon/8 \le 1/2$. By the above, (3.5) is satisfied. We then compute $\Gamma = \exp(\epsilon\delta n/2^8)$ and (3.6) gives

$$P(\tau_\delta^-(z) \le \tau_\epsilon^-(1-u) \wedge \epsilon\lfloor\Gamma\rfloor/32 \mid z(0) \ge 3\delta/2) \le 4/\Gamma,$$

which implies statement (ii).

Estimate on W_1. Since $W_1 = T_1$, this follows from (i) and Markov's inequality. □

Lemma 5.4. *Fix small $\delta, \epsilon > 0$ such that $\epsilon \ge 18\delta$ and $\delta \ge 4\epsilon^2$. Then*

(i) $P(\tau_\delta^-(|z - z^*|) \wedge \tau_{2\epsilon}^+(u) > \log(1/\delta) + t) \le e^{-t}$ *for any $t > 0$, and*
(ii) $P(\tau_{2\delta}^+(|z - z^*|) \le \tau_{2\epsilon}^+(u) \wedge \exp(\Omega n) \mid |z(0) - z^*| \le \delta) = \exp(-\Omega(n))$.

In particular,

- *for any $\alpha > 0$, $P(W_2 > (\alpha/6)\log n) \to 0$ as $n \to \infty$.*
- $P(u(T_2) > 2\epsilon$ *or* $|z(T_2) - z^*| \le \delta) = 1 - \exp(\Omega(n))$.

Proof. We first prove (i) and (ii).

Statement (i). Let $f(z) = 1 - 4(1-1/n)z - z^2$, then $f(z^*) = 0$. Moreover $f'(0) = -4(1-1/n)$ and $f''(z) = -2 < 0$, so $f'(z) \le -4(1-1/n)$ for $z \ge 0$. Letting $b = z - z_n^*$, we have $f(z_n^* + b)/b \le -4(1-1/n)$ for $b \ne 0$ and

$$\mu(b) = \frac{1}{2}(f(z_n^* + b) - u^2)$$

so if $b > \delta$ and $u < 2\epsilon$ then $\mu(b) \le -2(1-1/n)b + 2\epsilon^2 \le -b$ for large n, if $\delta \ge 4\epsilon^2$. Letting $\tau = \tau_\delta^-(b) \wedge \tau_{2\epsilon}^+(u)$, the process with $X_t = e^{t\wedge\tau}b_{t\wedge\tau}$ has $\mu(X) \le 0$, so is a supermartingale, and since $b_0 \le 1$, $X_0 \le 1$. Using Markov's inequality and $E[X_t] \le E[X_0] \le 1$,

$$P(\tau > t) \le P(X_t \ge e^t\delta) \le e^{-t}/\delta.$$

Thus $P(\tau > \log(1/\delta) + t) \le e^{-t}$ for $t \ge 0$. The same argument with $-b$ in place of b gives the same result, and combining gives statement (i).

Statement (ii). Using again the estimate $\mu(b) \le -b$ for $b > \delta$ and $u < 2\epsilon$, let $\tau = \tau_{2\delta}^+(b) \wedge \tau_{2\epsilon}^+(u)$ and let $X_t = b_t - \delta$. We will use Lemma 3.6 with $x = \delta$, $\mu_\star = \delta$, $\sigma_\star^2 = 4/n$, and $C_\mu = C_\Delta = 1/2$. We have $\Delta_\star(X) = 2/n \le \delta = x/2$ for large n. Again, (3.5) is easily verified. Then, $\Gamma = \exp(\delta^2 n/64)$, and (3.6) gives

$$P(\tau_{2\delta}^+(|z - z^*|) \le \tau_{2\epsilon}^+(u) \wedge \delta\lfloor\Gamma\rfloor/4 \mid |z(0) - z^*| \le 3\delta/2) \le 4/\Gamma,$$

which implies statement (ii).

Estimate on W_2. Since $T_2 \le \tau_\delta^-(|z-z^*|, T_1) \wedge \tau_{2\epsilon}^+(u, T_1)$, statement (i) implies

$$P(W_2 \ge \log(1/\delta) + t) \le e^{-t} \text{ for any } t > 0. \tag{5.6}$$

The estimate on W_2 follows by letting $t = (\alpha/8)\log n$.

Statement on T_2. If $u(T_1) \ge 1-\epsilon > 2\epsilon$ then $\tau_{2\epsilon}^+(u, T_1) = T_1$, so $T_2 = T_1$ and $u(T_2) > 2\epsilon$. Otherwise, $z(T_1) \ge 2\delta$. Since $T_2 \le \tau_{2\epsilon}^+(u, T_1) \le \tau_\epsilon^-(1-u, T_1)$, using Lemma 5.3, statement (ii) it follows that

$$P(\tau_\delta^-(z, T_1) \le T_2 \wedge (T_1 + \exp(\Omega(n))) \mid u(T_1) < 1-\epsilon) = \exp(-\Omega(n)).$$

Letting $t = n$ in (5.6), $P(T_2 \ge T_1 + \log(1/\delta) + n) \le e^{-n} = \exp(-\Omega(n))$. Since $\log(1/\delta) + n = o(\exp(\Omega(n)))$ and $\tau_\delta^-(z, T_1)$, combining the last two estimates and noting that $T_2 \le \tau_\delta^-(z, T_1)$ it follows that

$$P(\tau_\delta^-(z, T_1) = T_2 \mid u(T_1) < 1-\epsilon) = \exp(-\Omega(n)). \tag{5.7}$$

If E_1, E_2, F are events with $E \subseteq E_2, F$ then $P(E_1) = P(E_1 \cap F) \le P(E_2 \cap F) \le P(E_2 \cap F)/P(F) = P(E_2 \mid F)$. Let $E_1 = \{u(T_2) \le 2\epsilon \text{ and } |z(T_2) - z^*| > \delta\}$, $E_2 = \{z(T_2) \le \delta\}$ and $F = \{u(T_1) < 1-\epsilon\}$. Then, $E_1 \subseteq \{u(T_2) \le 2\epsilon\} \subseteq F$, $E_1 \subseteq E_2$ by definition of T_2, and (5.7) says that $P(E_2 \mid F) = \exp(-\Omega(n))$, so $P(E_1) = \exp(-\Omega(n))$, which implies the result. □

Lemma 5.5. *Fix small $\delta, \epsilon > 0$ and large $B > 0$ such that $\epsilon \ge 18\delta$ and $\delta \ge 4\epsilon^2$. Then for large n,*

(i) $E[\tau_{2B}^+(\sqrt{n}u)] \le 16B^2$, *and*
(ii) $P(\tau_B^-(\sqrt{n}u) \le \tau_{2\epsilon}^+(u) \wedge \tau_{2\delta}^+(|z-z^*|) \mid \sqrt{n}u(0) > 2B) = \exp(-\Omega(B^2))$.

In particular,

- *for any $\alpha > 0$, $P(W_3 > (\alpha/6)\log n) \to 0$ as $n \to \infty$ and*
- *$P(\sqrt{n}u(T_3) > 2B) \ge 1 - O(1/n)$.*

Proof. We begin with (i) and (ii).

Statement (i). Let $C = (2B)^2$ so that $\tau_{2B}^+(\sqrt{n}u) = \tau_C^+(nu^2)$. From (5.3) we note $\mu(u) \ge 0$, and if $u \le 2\epsilon$ then from (5.5), $\sigma^2(u) \ge (1-2\epsilon)/2n$. Using linearity of drift and the product rule of Lemma 3.4, if $u \le 2\epsilon$, ϵ is small and n is large then

$$\mu(nu^2) = n\mu(u^2) = 2nu\,\mu(u) + n\sigma^2(u) \ge (1-2\epsilon)/2 \ge 1/3.$$

It follows that $X(t) = nu^2(t \wedge \tau_C^+(nu^2)) - (1/3)t \wedge \tau_C^+(nu^2)$ is a submartingale. Since $(u+\Delta u)^2 - u^2 = (2u + \Delta u)\Delta u$ and $\Delta_\star(u) = 2/n$, if $nu^2 \le C$ and n is large, then $1/n \le \sqrt{C/n}$ and $\Delta_\star(nu^2) \le 8\sqrt{C/n}$, so $nu^2(\tau_C^+(nu^2)) \le C + 8\sqrt{C/n}$. Using optional stopping,

$$C + 8\sqrt{C/n} - (1/3)E[\tau_C^+(nu^2)] \ge E[X(\tau_C^+(nu^2)] \ge E[X_0] \ge 0,$$

so $E[\tau_C^+(nu^2)] \le 3C + 24\sqrt{C/n} \le 4C$ for large n, which is statement (i).

Statement (ii). The relevant stopping time is

$$\tau = \tau_{2\delta}^+(|z-z^*|) \wedge \tau_B^-(\sqrt{n}u) \wedge \tau_{2\epsilon}^+(u).$$

If $|z-z^*| \le 2\delta$ then $\mu(u) \ge (z^* - 2\delta)u$, and $\sigma^2(u) \le 4/n$. If we can find $\theta > 0$ such that $\mu_t(e^{-\theta\sqrt{n}u}) \le 0$ for $t < \tau$ then $\exp(-\theta\sqrt{n}u(t \wedge \tau))$ is a supermartingale and

$$P(\sqrt{n}u(\tau) \le B) = P(\exp(-\theta\sqrt{n}u(\tau)) \ge e^{-\theta B}) \le e^{\theta B}E[\exp(-\theta\sqrt{n}u(0))],$$

and so $P(\tau = \tau_B^-(\sqrt{n}u) \mid \sqrt{n}u(0) > 2B) \le e^{-\theta B}$. From (5.4) we have $\sigma^2(\sqrt{n}u) \le 4$, and if $\sqrt{n}u > B$ and $|z - z^*| \le 2\delta$ then $\mu(\sqrt{n}u) \ge (z^* - 2\delta)B \ge z^*B/2$ if $\delta > 0$ is small, so using Lemma 3.5,

$$\begin{aligned}\mu(\exp(-\theta\sqrt{n}u)) &\le -\theta e^{-\theta\sqrt{n}u}\mu(\sqrt{n}u) + \frac{\theta^2}{2}\,\sigma^2(\sqrt{n}u)\, e^{-\theta\sqrt{n}u + 2\theta/\sqrt{n}} \\ &= \theta e^{-\theta\sqrt{n}u}(-z^*B/2 + 2\theta e^{2\theta/\sqrt{n}}).\end{aligned}$$

Letting $\theta = z^*B/6$ the above is ≤ 0 for large n, and so $P(\tau = \tau_B^-(\sqrt{n}u) \mid \sqrt{n}u(0) > 2B) \le e^{-z^*B^2/6}$, which implies statement (ii).

Estimate on W_3. Since $T_3 \le \tau_{2B}^+(\sqrt{n}u, T_2)$, statement (i) implies $E[W_3] \le 16B^2$, so the desired estimate follows from Markov's inequality.

Estimate on T_3. Let $E = \{\sqrt{n}u(T_3) \le 2B\}$ and $F = \{|z(T_2) - z^*| \le \delta\}$; we wish to show $P(E) = O(1/n)$. If $u(T_2) > 2\epsilon$ then for large n, $2B/\sqrt{n} \le 2\epsilon$ so $T_3 = T_2$ and $\sqrt{n}u(T_3) > 2\epsilon\sqrt{n} \ge 2B$, which means $E \subseteq \{u(T_2) \le 2\epsilon\}$. Using Lemma 5.4, $P(\{u(T_2) \le 2\epsilon\} \cap F^c) = \exp(-\Omega(n))$ which implies $P(E \cap F^c) = \exp(-\Omega(n))$. On the other hand, if F holds then since $T_3 \le \tau_{2B}^+(\sqrt{n}u, T_2) \le \tau_{2\epsilon}^+(u, T_2)$, from Lemma 5.4 statement (ii),

$$P(\tau_{2\delta}^+(|z - z^*|, T_2) \le T_3 \wedge (T_2 + \exp(\Omega(n)) \mid F) = \exp(-\Omega(n)).$$

Using $T_3 = T_2 + W_3$ and $E[W_3] \le 16B^2$, $P(T_3 \ge T_2 + n) \le 16B^2/n$.
Combining with the above and noting that $T_3 \le \tau_{2\delta}^+(|z - z^*|, T_2)$,

$$P(\tau_{2\delta}^+(|z - z^*|, T_2) = T_3 \mid F) \le \exp(-\Omega(n)) + 16B^2/n = O(1/n).$$

If $\sqrt{n}u(T_3) \le 2B$ then by definition of T_3, $|z(T_3) - z^*| > 2\delta$, so we deduce $P(E \mid F) = O(1/n)$.
Noting that $P(E \cap F) = P(E \mid F)P(F) \le P(E \mid F)$ and combining the two main estimates,

$$P(E) = P(E \cap F) + P(E \cap F^c) \le P(E \mid F) + P(E \cap F^c) = O(1/n),$$

which is the desired result. □

Lemma 5.6. *Fix small $\delta, \epsilon > 0$ and large $B > 0$ such that $\epsilon \ge 18\delta$ and $\delta \ge 4\epsilon^2$. Let $\tau = \tau_{2\delta}^+(|z - z^*|) \wedge \tau_B^-(\sqrt{n}u) \wedge \tau_{2\epsilon}^+(u)$. Then*

(i) $P(\tau \ge ((1/2z^*) + \alpha/6)\log n \mid \sqrt{n}u(0) > 2B) \to 0$ *as* $n \to \infty$*, and*
(ii) $P(\tau = \tau_{2\epsilon}^+(u) \le ((1/2z^*) - \alpha/6)\log n \mid \sqrt{n}u(0) \le 2B + 2/\sqrt{n}) \to 0$ *as* $n \to \infty$.

In particular,

- $P(W_4 > ((1/2z^*) + \alpha/6)\log n) \to 0$ *as* $n \to \infty$,
- $P(u(T_4) > 2\epsilon) \ge 1 - \exp(-\Omega(B^2)) - o(1)$*, and*
- $P(W_4 < ((1/2z^*) - \alpha/6)\log n \mid u(0) \le 2B$ *and* $|z(0) - z^*| \le \delta) \le o(1) + \exp(-\Omega(B^2)) + o(1)$.

Proof. We first prove (i) and (ii).

Statement (i). Let $z_-^* = z^* - 2\delta$ and let $v = 1/u$. Recall from (5.4) that $\sigma^2(u) \le 4/n$. If $t < \tau$ then $\mu(u) \ge z_-^* u$ and since $u \ge B/\sqrt{n}$, $u - 2/n \ge u/2^{1/3}$ for large n and $v^3 \le nv/B^2$, so using Lemma 3.5,

$$\begin{aligned}\mu(v) &\le -\frac{1}{u^2}\mu(u) + \frac{1}{2}\sigma^2(u)\frac{1}{(u - 2/n)^3} \\ &\le -z_-^* v + 4v^3/n \\ &\le -(z_-^* - 1/B^2)v.\end{aligned}$$

Let $z_B = z_-^* - 1/B^2$, then the process with $X(t) = v(t \wedge \tau)\exp(z_B(t \wedge \tau))$ is a supermartingale. If $u(0) > 2\epsilon$ then $\tau = 0$. If $u(0) \le 2\epsilon$ and $\tau \ge t$ then $u(\tau) \le 2\epsilon + 2/n$ and $X(\tau) \ge \exp(z_B t)/(2\epsilon + 2/n) \ge \exp(z_B t)/\epsilon$ for large n. Thus for $t > 0$

$$P(\tau \ge t) \le P(X(\tau) \ge e^{z_B t}/\epsilon) \le \epsilon\, e^{-z_B t}\, E[X_0].$$

If $\sqrt{n}u(0) > 2B$ then $X(0) = v(0) < \sqrt{n}/2B$, so the above is at most $\epsilon\, e^{-z_B t}\sqrt{n}/2B$. Let

$$t = (1/z_B)(\log(n)/2 + \log(\epsilon) + \log(1/2B)) + (\alpha/4)\log n,$$

then the above probability is at most $n^{-z_B\alpha/4}$ which is $o(1)$. Moreover, since $|z_B - z^*| \to 0$ uniformly in n as $\min(1/B, \delta) \to 0$, if $1/B, \delta$ are small enough then $t \ge ((1/2z^*) + \alpha/6))\log n$ for large n. Together this gives (i).

Statement (ii). Let $z_+^* = z^* + 2\delta$. If $t < \tau$ then $u_t \ge 2/n$ and $z \le z_+^*$ so from (5.3), $\mu(u) \le z_+^* u$ and the process with $X(t) = \exp(-z_+^*(t \wedge \tau))u(t \wedge \tau)$ is a supermartingale. Thus

$$P(\tau = \tau_{2\epsilon}^+(u) \le t) \le P(X(t) \ge 2\epsilon\exp(-z_+^* t)) \le E[X(0)]\exp(z_+^* t)/(2\epsilon).$$

If $\sqrt{n}u(0) \le 2B + 2/\sqrt{n}$ then $X(0) \le 3B/\sqrt{n}$ for large n and the above is at most $(3B/\sqrt{n})e^{z_+^* t}/(2\epsilon)$. Let

$$t = (1/z_+^*)(\log(1/3B) + \log(n)/2 + \log(2\epsilon)) - (\alpha/4)\log n,$$

then the above probability is at most $n^{-z_+^*\alpha/4}$ which is $o(1)$. Moreover, since $|z_+^* - z^*| \to 0$ uniformly in n as $\delta \to 0$, if δ is small enough then for large n, $t \le ((1/2z^*) - (\alpha/6))\log n$. Together this gives (ii).

Upper bound on W_4. Let $E = \{W_4 \ge ((1/2z^*) + \alpha/6)\log n\}$ and $F = \{\sqrt{n}u(T_3) > 2B\}$. From Lemma 5.5, $P(F^c) = O(1/n)$, and from statement (i), $P(E \mid F) \to 0$ as $n \to \infty$. Therefore

$$P(E) = P(E \cap F) + P(E \cap F^c) = P(E \mid F) + P(F^c) \to 0 \text{ as } n \to \infty.$$

Statement on T_4. If $u(T_3) > 2\epsilon$ or if $|z(T_3) - z^*| > 2\delta$ then $T_3 = T_4$, so we can also express T_4 as

$$T_4 = \tau_{2\delta}^+(|z - z^*|, T_2) \wedge \tau_B^-(\sqrt{n}u, T_3) \wedge \tau_{2\epsilon}^+(u, T_2).$$

From Lemma 5.5, $P(\sqrt{n}u(T_3) \le 2B) = O(1/n)$. Noting that $\tau_x^\pm(X, s) \le \tau_x^\pm(X, t)$ for any x, X and $s \le t$ and combining with (ii) from Lemma 5.5, it follows that

$$\begin{aligned}
&P(T_4 \ne \tau_{2\delta}^+(|z - z^*|, T_2) \wedge \tau_{2\epsilon}^+(u, T_2)) \\
&= P(\tau_B^-(\sqrt{n}u, T_3) < \tau_{2\delta}^+(|z - z^*|, T_2) \wedge \tau_{2\epsilon}^+(u, T_2)) \\
&\le P(\tau_B^-(\sqrt{n}u, T_3) < \tau_{2\delta}^+(|z - z^*|, T_3) \wedge \tau_{2\epsilon}^+(u, T_3)) \\
&\le P(\sqrt{n}u(T_3) \le 2B) + P(\tau_B^-(\sqrt{n}u, T_3) < \tau_{2\delta}^+(|z - z^*|, T_3) \wedge \tau_{2\epsilon}^+(u, T_3) \mid \sqrt{n}u(T_3) > 2B) \\
&\le O(1/n) + \exp(-\Omega(B^2)).
\end{aligned}$$

Using (ii) from Lemma 5.4,

$$P(\tau_{2\delta}^+(|z - z^*|, T_2) \le \tau_{2\epsilon}^+(u, T_2) \wedge (T_2 + \exp(\Omega(n))) \mid |z(T_2) - z^*| \le \delta) = \exp(-\Omega(n)).$$

By Lemmas 5.4 and 5.5, $P(W_2 + W_3 \ge (\alpha/3)\log n) \to 0$. Since $\tau_{2\delta}^+(|z - z^*|, T_2) \ge T_4 = T_2 + W_2 + W_3$, it follows that

$$P(\tau_{2\delta}^+(|z - z^*|, T_2) \le \tau_{2\epsilon}^+(u, T_2) \mid |z(T_2) - z^*| \le \delta) = \exp(-\Omega(n)) + o(1) = o(1).$$

If $u(T_2) > 2\epsilon$ then $T_2 = T_3 = T_4$ so $u(T_4) > 2\epsilon$, and from Lemma 5.4, $P(u(T_2) \leq 2\epsilon$ and $|z(T_2) - z^*| > \delta) = \exp(-\Omega(n))$. Combining with the previous estimates,

$$\begin{aligned} P(u(T_4) \leq 2\epsilon) &\leq P(T_4 \neq \tau^+_{2\delta}(|z-z^*|, T_2) \wedge \tau^+_{2\epsilon}(u, T_2)) \\ &\quad + P(\tau^+_{2\delta}(|z-z^*|, T_2) \leq \tau^+_{2\epsilon}(u, T_2) \mid |z(T_2) - z^*| \leq \delta) \\ &\quad + P(u(T_2) \leq 2\epsilon \text{ and } |z(T_2) - z^*| > \delta) \\ &= \exp(-\Omega(B^2)) + o(1). \end{aligned}$$

Lower bound on W_4. If $\sqrt{n}u(0) \leq 2B$ and $|z(0) - z^*| \leq \delta$ then $T_2 = 0$ and so $\sqrt{n}u(T_3) \leq 2B + 2/\sqrt{n}$. Using statement (ii) above,

$$\begin{aligned} &P(W_4 \leq ((1/2z^*) - \alpha/6)\log n \mid \sqrt{n}u(0) \leq 2B \text{ and } |z(0) - z^*| \leq \delta) \\ &\leq P(u(T_4) \leq 2\epsilon) + P(\tau = \tau^+_{2\epsilon}(u, T_3) \leq ((1/2z^*) - \alpha/6)\log n \mid \sqrt{n}u(T_3) \leq 2B + 2/\sqrt{n}) \\ &= \exp(-\Omega(B^2)) + o(1). \end{aligned}$$

□

Lemma 5.7. *Fix small $\delta, \epsilon > 0$ and large $B > 0$ such that $\epsilon \geq 18\delta$ and $\delta \geq 4\epsilon^2$. Let $\tau = \tau^-_\delta(z) \wedge \tau^-_\epsilon(u) \wedge \tau^-_\epsilon(1-u)$. Then*

(i) $E[\tau] \leq 1/(\delta\epsilon)$ *and*
(ii) $P(\tau = \tau^-_\epsilon(u) \leq \exp(\Omega(n)) \mid u(0) > 2\epsilon) = \exp(-\Omega(n))$.

In particular,

- *for any $\alpha > 0$, $P(W_5 > (1/6\alpha)\log n) \to 0$ as $n \to \infty$, and*
- $P(u(T_5) \geq 1 - \epsilon) \geq 1 - o(1) - \exp(-\Omega(B^2))$.

Proof. From (5.3), $\mu(u) \geq uz$. If $t < \tau$ then $z_t > \delta$ and $u_t > \epsilon$ so $\mu_t(u) \geq \delta\epsilon$. Thus, the process with $X(t) = u(t\wedge\tau) - \delta\epsilon(t \wedge \tau)$ is a submartingale. Moreover $X(0) = u(0) \geq 0$ and $X(\tau) = u(\tau) - \delta\epsilon\tau \leq 1 - \delta\epsilon\,\tau$. Using optional stopping, $0 = E[X(0)] \leq E[X(\tau)] \leq 1 - \delta\epsilon\,E[\tau]$ and (i) follows.

To show (ii), use Lemma 3.6 with $x = \epsilon$, $X = 2\epsilon - u$, $\mu_\star = \delta\epsilon$, $\sigma^2_\star = 4/n$, $C_\mu = 2$ and $C_\Delta = 1/2$. Then $\Delta_\star(X) = 2/n$ so $\Delta_\star(X)\mu_\star/\sigma^2_\star = \epsilon\delta/2 \leq 1/2$ for small $\epsilon, \delta > 0$ and (3.5) follows from $\mu_t(u) \geq \delta\epsilon$ for $t < \tau$ and (5.4). Then, $\Gamma = \exp(\delta\epsilon^2 n/64) = \exp(\Omega(n))$, $\lfloor\Gamma\rfloor x/16C_\mu = \lfloor\Gamma\rfloor\epsilon/32 = \exp(\Omega(n))$ and (ii) follows from (3.6).

Estimate on W_5. This follows from (i) and Markov's inequality.
Statement on T_5. For any $i \in \{1,2,3,4\}$, if $u(T_i) \geq 1 - \epsilon$ then $T_5 = T_i = \tau^-_\epsilon(1-u, T_i)$, and similarly if $z(T_i) \leq \delta$ then $T_5 = T_i = \tau^-_\delta(z, T_i)$. In particular,

$$\{T_5 = \tau^-_\delta(z, T_4)\} = \{T_5 = \tau^-_\delta(z, T_1) \leq \tau^-_\epsilon(1-u, T_1)\}$$

If $u(T_1) > 1 - \epsilon$ then $T_5 = T_1$ so $u(T_5) > 1 - \epsilon$, and if $u(T_1) < 1 - \epsilon$ then $z(T_1) \geq 2\delta$. Combining,

$$\begin{aligned} \{u(T_5) < 1 - \epsilon\} &\subseteq \{T_5 = \tau^-_\delta(z, T_4)\} \cup \{T_5 = \tau^-_\epsilon(u, T_4)\} \\ &\subseteq \{T_5 \geq T_4 + (\alpha/6)\log n\} \\ &\cup \{u(T_4) \leq 2\epsilon\} \\ &\cup (\{T_5 = \tau^-_\delta(z, T_1) \leq \tau^-_\epsilon(1-u, T_1) \wedge (T_4 + \alpha/6)\log n\} \cap \{z(T_1) \geq 2\delta\}) \\ &\cup (\{T_5 = \tau^-_\epsilon(u, T_4) \leq (T_4 + \alpha/6)\log n\} \cap \{u(T_4) > 2\epsilon\}). \end{aligned}$$

Using the above estimate on W_5, Lemma 5.3 statement (ii), and statement (ii) above,

$$\begin{aligned}
P(u(T_5) < 1-\epsilon) \le\ & P(T_5 \ge T_4 + (\alpha/6)\log n) \\
&+ P(u(T_4) \le 2\epsilon) \\
&+ P(T_5 = \tau_\delta^-(z, T_1) \le \tau_\epsilon^-(1-u, T_1) \wedge (T_4 + (\alpha/6)\log n) \mid z(T_1) \ge 2\delta) \\
&+ P(T_5 = \tau_\epsilon^-(u, T_4) \le (T_4 + (\alpha/6)\log n) \mid u(T_4) > 2\epsilon) \\
&= o(1) + \exp(-\Omega(B^2)). \qquad \square
\end{aligned}$$

Lemma 5.8. *Fix small $\delta, \epsilon > 0$ and large $B > 0$ such that $\epsilon \ge 18\delta$ and $\delta \ge 4\epsilon^2$. Let $\tau = \tau_{4\epsilon}^+(1-u) \wedge \tau_0^-(1-u)$. Then,*

(i) $P(u(\tau) < 1 \mid u(0) \ge 1-\epsilon) = \exp(-\Omega(n))$,
(ii) $P(\tau > (1+\alpha/6)\log n) \to 0$ *as* $n \to \infty$, *and*
(iii) $P(\tau < (1-\alpha/6)\log n \mid u(0) \le 1-\epsilon+2/n) \to 0$ *as* $n \to \infty$.

In particular,

- $P(u(T_6) = 1) \ge 1 - o(1) - \exp(-\Omega(B^2))$,
- *for any* $\alpha > 0$, $P(W_6 > (1+\alpha/6)\log n) \to 0$ *as* $n \to \infty$, *and*
- *for any* $\alpha > 0$, $P(W_6 < (1-\alpha/6)\log n) \mid u(0) < 1-\epsilon) \to 0$ *as* $n \to \infty$.

Proof. Let $b = 1-u$. From (5.3), if $u \ge 1-4\epsilon$ then $u \ge 2/n$ for large n and

$$\begin{aligned}
&\mu(b) = -\mu(u) = -uz = bz - z \\
&\mu(z) = \frac{1}{2}((1-u)(1+u) - 4(1-1/n)z - z^2) = \frac{1}{2}(b(2-b) - 4z - z^2) + O(z/n),
\end{aligned}$$

so linearizing around $(b, z) = (0, 0)$,

$$\begin{pmatrix}\mu(b)\\ \mu(z)\end{pmatrix} = \left(\begin{pmatrix}0 & -1\\ 1 & -2\end{pmatrix} + O(\epsilon + 1/n)\right)\begin{pmatrix}u\\ z\end{pmatrix}.$$

The above matrix has trace -2 and determinant 1, so repeated eigenvalues $-1, -1$, and $(1,1)^\top$ is an eigenvector. Let $\xi(t) = (1,1)^\top(b(t), z(t)) = b(t) + z(t)$, so that for $t < \tau$,

$$\mu_t(\xi) = -(1 + O(\epsilon + 1/n))\xi_t.$$

Using the inequality $(x+y)^2 \le 2x^2 + 2y^2$,

$$\sigma^2(\xi) = \sum_i q_i(\Delta_i(u) + \Delta_i(z))^2 \le 2\sum_i q_i(\Delta_i(u)^2 + \Delta_i(z)^2) = 2(\sigma^2(u) + \sigma^2(z)).$$

Using (5.5), $\sigma^2(b), \sigma^2(z) \le 8b/n \le 8\,\xi/n$, so $\sigma^2(\xi) \le 32\,\xi/n$. If $\theta > 0$ is such that $\mu_t(e^{\theta\xi}) \le 0$ for $t < \tau$ then $\exp(\theta\xi(t \wedge \tau))$ is a supermartingale. By definition of τ, if $b(\tau) \ne 0$ then $b(\tau) > 4\epsilon$, so

$$P(b(\tau) \ne 0) = P(\exp(\theta\xi(\tau)) > e^{4\theta\epsilon}) \le e^{-4\theta\epsilon}E[e^{\theta\xi(0)}]. \tag{5.8}$$

Using Lemma 3.5, the above estimates on $\mu(\xi), \sigma^2(\xi)$, and the fact that $\Delta_\star(\xi) \le 4/n$, if $t < \tau$ then

$$\begin{aligned}
\mu_t(e^{\theta\xi}) &\le e^{\theta\xi_t}(-(1 - O(\epsilon + 1/n))\,\theta\,\xi_t + \frac{\theta^2}{2}e^{4\theta/n}\frac{32\,\xi_t}{n}) \\
&= \theta\,\xi_t\,e^{\theta\xi_t}(-1 + O(\epsilon + 1/n)) + \frac{16\,\theta}{n}\,e^{4\theta/n}).
\end{aligned}$$

Letting $\theta = n/32$, the above is ≤ 0 for small $\epsilon > 0$ and large n. Recall that $z \leq 1 - u = b$, so $\xi = b + z \leq 2b$. Using (5.8),

$$P(b(\tau) \neq 0 \mid b(0) \leq \epsilon) \leq e^{-4\theta\epsilon + 2\theta\epsilon} = e^{-2\theta\epsilon} = e^{-n\epsilon/16},$$

which gives statement (i).

Statement (ii). For small $\epsilon > 0$ and large n, from the estimate on $\mu_t(\xi)$ the process with $X(t) = e^{(1-\alpha/8)(t\wedge\tau)}\xi(t \wedge \tau)$ is a supermartingale. If $\tau > t$ then $\xi(t) \geq b(t) \geq 1/n$, and if $b(0) > 4\epsilon$ then $\tau = 0$. Since $\xi(0) \leq 2b(0)$, it follows that

$$P(\tau > t) \leq P(X(t) \geq \exp((1 - \alpha/8)t)/n \mid b(0) \leq 2\epsilon) \leq ne^{-(1-\alpha/8)t}(8\epsilon).$$

Letting $t = (1 + (\alpha/6) \log n$, if $\alpha > 0$ is small enough the above probability is $n^{-\Omega(1)}$, so $\to 0$ as $n \to \infty$.

Statement (iii). Let $\rho = \tau^-_{n^{-1+\epsilon}}(b) \wedge \tau^+_{4\epsilon}(b)$ so that $\rho \leq \tau$ and let $v = 1/\xi$ so that $v(t)$ is defined at least for $t < \rho$. If $t < \rho$ then $\xi(t) \geq b(t) \geq n^{-1+\epsilon}$. Recall that $\sigma^2(\xi) \leq 32\,\xi$. Using Lemma 3.5, if $t < \rho$ and n is large enough that $n^{-1+\epsilon} - 4/n \geq 2^{-1/3}n^{-(1+\epsilon)}$ then

$$\begin{aligned}
\mu_t(v) &\leq \mu_t(\xi)\frac{-1}{\xi_t^2} + \frac{\sigma_t^2(\xi)}{2}\,\frac{2}{(\xi_t - 4/n)^3} \\
&\leq (1 + O(\epsilon + 1/n))\frac{\xi_t}{\xi_t^2} + \frac{16\,\xi_t}{n\xi_t^3} \\
&\leq ((1 + O(\epsilon + 1/n) + 16n^{-\epsilon})v(t).
\end{aligned}$$

If $\epsilon > 0$ is small and n large then the process with $X(t) = e^{-(1+\alpha/8)(t\wedge\rho)}v(t \wedge \rho)$ is a supermartingale. If $b \leq n^{-1+\epsilon}$ then $v \geq n^{1-\epsilon}$, so

$$P(\rho = \tau^-_{n^{-1+\epsilon}}(b) \leq t) \leq P(X(t) \geq e^{-(1+\alpha/8)t}n^{1-\epsilon}) \leq e^{(1+\alpha/8)t}n^{-1+\epsilon}E[X(0)].$$

If $\tau = \tau^+_{4\epsilon}(b)$ then $\rho = \tau^+_{4\epsilon}(b)$, so by statement (i), $P(\rho \neq \tau^-_{n^{-1+\epsilon}}(b)) = o(1)$. If $b(0) \geq \epsilon - 2/n$ then $X(0) \leq 1/(\epsilon - 2/n) \leq 2/\epsilon$ for large n. Letting $t = (1 - \alpha/6)\log n$, if $\alpha, \epsilon > 0$ are small and $b(0) \geq \epsilon - 2/n$ the above probability is $n^{-\Omega(n)} = o(1)$, and gives statement (iii).

Statement on T_6. From Lemma 5.7, $P(u(T_5) < 1 - \epsilon) = o(1) + \exp(-\Omega(B^2))$. Using this and statement (i) above,

$$\begin{aligned}
P(u(T_6) < 1) &\leq P(u(T_5) < 1 - \epsilon) + P(u(T_6) = 1 \mid u(T_5) \geq 1 - \epsilon) \\
&\leq o(1) + \exp(-\Omega(B^2)).
\end{aligned}$$

Upper bound on W_6. This follows from statement (ii) and Markov's inequality.
Lower bound on W_6. If $u(0) < 1 - \epsilon$ then $u(T_5) \leq 1 - \epsilon + 2/n$. Using statement (ii) above,

$$\begin{aligned}
&P(W_6 < (1 - \alpha/6)\log n \mid u(0) < 1 - \epsilon) \\
&\leq P(W_6 < (1 - \alpha/6)\log n \mid u(T_5) \leq 1 - \epsilon + 2/n) = o(1). \qquad \square
\end{aligned}$$

Proof of Propositions 5.1 and 5.2. The results follow directly from Lemmas 5.3–5.8, letting $B \to \infty$ slowly enough as $n \to \infty$. □

Acknowledgements

The author wishes to thank Nicolas Lanchier for suggesting the model, and for many helpful conversations while working on the paper.

Appendix

Results of Section 3

Proof of Lemma 3.1. For a semimartingale (s-m) X and locally bounded, predictable H, the notation $H \cdot X$ refers to the stochastic integral, defined in [8, I.4d]. Since M has bounded jumps it is locally square-integrable. Let $V = (e^c/2)\langle M\rangle$, noting that V is locally integrable. Since M is a local martingale, $M - V$ is a s-m. Let $E = \exp(M - V)$. Then by [8, I.4.57] (Itô's formula), E is a s-m and

$$E = 1 + E_- \cdot (M - V + \frac{1}{2}\langle M^c\rangle + Y), \quad \text{where} \tag{5.9}$$
$$Y_t = \sum_{s\le t} e^{\Delta M_s} - 1 - \Delta M_s.$$

Since $E \ge 0$, if E is the sum of a local martingale and a non-increasing process, it is a supermartingale, so our goal is to obtain such a decomposition for the right-hand side of the first line in (5.9).

If $|x| \le c$ then $e^x - 1 - x \le \frac{1}{2}e^c x^2$, so letting $W_t = \sum_{s\le t}(\Delta M_s)^2$ and $Z = Y - (e^c/2)W$, Z is non-increasing. By [8, I.4.52], $[M] = \langle M^c\rangle + W$, so

$$\begin{aligned}
-V + \frac{1}{2}\langle M^c\rangle + Y &= -\frac{e^c}{2}\langle M\rangle + \frac{1}{2}\langle M^c\rangle + Z + \frac{e^c}{2}W \\
&= -\frac{e^c}{2}\langle M\rangle + \frac{1-e^c}{2}\langle M^c\rangle + Z + \frac{e^c}{2}(\langle M^c\rangle + W) \\
&= \frac{e^c}{2}([M] - \langle M\rangle) + \frac{1-e^c}{2}\langle M^c\rangle + Z.
\end{aligned}$$

By [8, I.4.50(b)], the first term is a local martingale, and since $\langle M^c\rangle$ is non-decreasing and $c \ge 0$, both the second and third terms are non-increasing. Combining with (5.9), we find there is a local martingale N and a non-decreasing process Q such that

$$E = 1 + E_- \cdot (N + Q).$$

Using [8, I.4.34(b)], since N is a local martingale, so is $E_- \cdot N$, and since $E \ge 0$, $1 + E_- \cdot Q$ is non-increasing. Since E has the desired form, it is a supermartingale. □

Proof of Lemma 3.2. $\mathbf{1} \Rightarrow \mathbf{2}$. By [8, I.3.21], $\Delta(X^p) = {}^p(\Delta X)$. If X is qlc, then by [8, I.2.35] its predictable projection ${}^pX$ has ${}^pX = X_-$. From uniqueness and property (ii) in [8, I.2.28] it follows that the ${}^p(\,\cdot\,)$ operation is linear. Therefore

$$\Delta(X^p) = {}^p(\Delta X) = {}^pX - {}^pX_- = 0,$$

i.e., X^p is continuous. Using (3.1), $\Delta X^m = \Delta X$ which implies that X^m is qlc, by definition of qlc.
$\mathbf{2} \Rightarrow \mathbf{1}$. If X^p is continuous then $\Delta X = \Delta X^m$. Thus if, in addition, X^m is qlc then X is qlc.
$\mathbf{2} \Leftrightarrow \mathbf{3}$. By [8, I.4.2], $\langle X\rangle$ is continuous iff X^m is qlc. □

Proof of Lemma 3.3. Since X has bounded jumps, it is special, so X^m, X^p are defined. Note that for $\lambda > 0$ and $\bullet \in \pm$, $\langle \bullet\lambda X^m\rangle = \lambda^2\langle X^m\rangle$. Since X is qlc, by Lemma 3.2, X^m is qlc, and as noted below (3.1), $\Delta_\star(X^m) \le \Delta_\star(X) \le c$. Let $M = \bullet\lambda X^m$ in Lemma 3.1, which has $|\Delta M| \le \lambda c$, and use Doob's inequality to find

$$P(\sup_{t\ge 0} \bullet\lambda X^m_t - (\lambda^2 e^{\lambda c}/2\,)\langle X^m\rangle_t \ge \lambda a) \le e^{-\lambda a}.$$

Since $1/2 \le \log 2$, if $\lambda c \le 1/2$ then $e^{\lambda c} \le 2$, and the result follows. □

Proof of Lemma 3.4. By definition of quadratic variation,

$$XY = X_0Y_0 + [X,Y] + X_- \cdot Y + Y_- \cdot X.$$

Since $X = X_0 + X^p + X^m$, $Y = Y_0 + Y^p + Y^m$ and $X_0 + X^p$, $Y_0 + Y^p$ have locally finite variation, $[X,Y] = [X^m, Y^m]$. Since X^m, Y^m are locally square-int, $[X^m, Y^m]$ has compensator $\langle X^m, Y^m\rangle$, so XY has compensator

$$(XY)^p = \langle X^m, Y^m\rangle + X_- \cdot Y^p + Y_- \cdot X^p.$$

The result will follow if we can show $\langle X^m, Y^m\rangle$ is absolutely continuous. For any $s < t$, applying the Cauchy-Schwarz inequality to the symmetric, bilinear and semidefinite map $(X,y) \mapsto \langle X,Y\rangle_t - \langle X,Y\rangle_s$ gives

$$|\langle X^m, Y^m\rangle_t - \langle X^m, Y^m\rangle_s| \le \sqrt{(\langle X^m\rangle_t - \langle X^m\rangle_s)(\langle Y^m\rangle_t - \langle Y^m\rangle_s)}.$$

Absolutely continuity of $t \mapsto \langle X^m\rangle_t$, $\langle Y^m\rangle_t$ means that for any $\epsilon > 0$ there is $\delta > 0$ so that if $\sum_i |t_i - s_i| < \delta$ then $\sum_i |\langle X^m\rangle_{t_i} - \langle X^m\rangle_{s_i}|$, $\sum_i |\langle Y^m\rangle_{t_i} - \langle Y^m\rangle_{s_i}| < \epsilon$. Using the Cauchy-Schwarz inequality to obtain the second line,

$$\begin{aligned}
\sum_i |\langle X^m, Y^m\rangle_{t_i} - \langle X^m, Y^m\rangle_{s_i}| &\le \sum_i \sqrt{(\langle X^m\rangle_{t_i} - \langle X^m\rangle_{s_i})(\langle Y^m\rangle_{t_i} - \langle Y^m\rangle_{s_i})} \\
&\le \left(\sum_i |\langle X^m\rangle_{t_i} - \langle X^m\rangle_{s_i}| \sum_i |\langle Y^m\rangle_{t_i} - \langle Y^m\rangle_{s_i}|\right)^{1/2} \\
&< (\epsilon \cdot \epsilon)^{1/2} = \epsilon
\end{aligned}$$

which shows that $\langle X^m, Y^m\rangle$ is absolutely continuous. ☐

Proof of Lemma 3.7. First we treat the case $X_0 = 1$, so that $Y_t = X_t/m(t) - \int_0^t b(s)/m(s)ds$. Given $y > 0$ define $\tau(y) = \inf\{t : Y_t \ge y\}$, and note that $\tau(y) < \zeta'$. Since $1/m(t) = e^{-\int_0^t \ell(s)ds}$, $(1/m(t))' = -\ell(t)/m(t)$, so using linearity of the drift and Lemma 3.4,

$$\mu(Y_t) \le (b(t) + \ell(t)X_t)/m(t) + X_t(-\ell(t)/m(t)) - b(t)/m(t) = 0,$$

which implies $Y^p \le 0$. Clearly $\sigma_t^2(Y) = (1/m(t))^2\sigma_t^2(X)$. Suppose that X satisfies hypothesis (i) of the lemma. Since X is non-decreasing, it has finite variation, so has zero continuous martingale part. Thus X^m is purely discontinuous and $\langle X\rangle_t = (\sum_{s\le t}(\Delta X_s)^2)^p$. In addition, $0 \le \Delta X_s \le c$, so $(\Delta X_s)^2 \le c\Delta X_s$. Using this and $\sum_{t\le s\le t+r} \Delta X_s \le X_{t+r} - X_t$, for any $t, r \ge 0$,

$$\langle X^m\rangle_{t+r} - \langle X^m\rangle_t \le c(\sum_{t\le s\le t+r} \Delta X_s)^p \le c(X_t^p - X_r^p)$$

which implies $\sigma_t^2(X) \le c\mu_t(X)$, i.e., hypothesis (ii) of the lemma. Using $\mu_t(X) \le b(t) + \ell(t)X_t = b(t) + \ell(t)m(t)Y_t$, hypothesis (ii), and continuity of $t \mapsto m(t)$ and $t \mapsto \int_0^t b(s)/m(s)ds$,

$$\sigma_t^2(Y) = \sigma_t^2(X)/m(t)^2 \le c\mu_t(X)/m(t)^2 = cb(t)/m(t)^2 + (c/m(t))\ell(t)Y_t.$$

Since $Y_t < y$ for $t < \tau(y)$,

$$\langle Y\rangle_{\tau(y)} \le c\int_0^{\tau(y)} b(s)/m(s)^2ds + yc\int_0^{\tau(y)} \ell(s)/m(s)ds = c\beta(\tau(y)) + yc\alpha(\tau(y)),$$

the last equality defining $\alpha(t)$ and $\beta(t)$. Taking the antiderivative,

$$\alpha(t) = \int_0^t e^{-\int_0^s \ell(r)dr}\ell(s)ds = 1 - e^{-\int_0^t \ell(s)ds} = 1 - 1/m(t) \le 1 \quad \text{for all} \quad t \ge 0.$$

Since $Y_0 = 1$, $Y_{\tau(y)} \ge y$ and $Y^p \le 0$, it follows that for $\lambda > 0$,

$$Y^m_{\tau(y)} - \lambda\langle Y\rangle_{\tau(y)} = Y_{\tau(y)} - Y_0 - Y^p_{\tau(y)} - \lambda\langle Y\rangle_{\tau(y)} \ge y - 1 - \lambda c(\beta + y).$$

Since $m(t) \ge 1$, $\Delta_\star(Y) \le \Delta_\star(X) \le c$, so using Lemma 3.3 with $a = y - 1 - \lambda c(\beta + y)$, assuming $\lambda c \le 1/2$ we find

$$P(\sup_{t<\zeta'} Y_t \ge y) \le 2e^{-\lambda a}.$$

Optimizing λa gives $\lambda = (y-1)/(2c(y+\beta))$ and

$$\lambda a \ge (y-1)^2/(4cy(1+\beta/y)) \ge (y-2)/(4c(1+\beta)),$$

and if $y \ge 1$ the assumption $c\lambda \le 1/2$ holds. For general X_0, first condition on X_0 and apply the above to X_t/X_0, whose jumps are bounded by c/X_0 instead of c. Then, integrate over X_0 to obtain the result.

To see that $\zeta \ge \zeta'$, note that $\{\zeta \ge \zeta'\} \supset \bigcup_y \{\sup_{t<\zeta'} Y_t < y\}$ and that the above estimate implies the latter event has probability 1. □

Miscellaneous estimates

Lemma 5.9. *Suppose $\beta > 0$ and $\beta c > 0 \vee (1 + a - \beta)x^{-\beta}$, then*

$$\int_x^\infty t^a e^{-ct^\beta} dt \le (\beta c - 0 \vee (1 + a - \beta)x^{-\beta})^{-1} x^{1+a-\beta} e^{-cx^\beta}.$$

Proof. Since

$$\frac{d}{dx}(x^{1+a-\beta}e^{-cx^\beta}) = ((1+a-\beta)x^{-\beta} - \beta c)x^a e^{-cx^\beta}$$

and $\beta c - (1+a-\beta)t^{-\beta} \ge \beta c - 0 \vee (1+a-\beta)x^{-\beta}$ for $t \ge x$ (to verify, consider separately the cases $1 + a - \beta \ge 0$ and $1 + a - \beta < 0$), if $\beta c - 0 \vee (1+a-\beta)x^{-\beta} > 0$ we have the upper bound

$$\begin{aligned} \int_x^\infty t^a e^{-ct^\beta} dt &\le (\beta c - 0 \vee (1+a-\beta)x^{-\beta})^{-1} \int_x^\infty (\beta c - (1+a-\beta)t^{-\beta}) t^a e^{-ct^\beta} dt \\ &= (\beta c - 0 \vee (1+a-\beta)x^{-\beta})^{-1} x^{1+a-\beta} e^{-cx^\beta}. \end{aligned} \tag{5.10}$$

□

Lemma 5.10. *Suppose $c, \lambda > 0$, $\alpha \in [0, 1/2]$ and $2c\lambda^{\alpha-1} \le 1$. Then*

$$(1 + \lambda - c\lambda^{1/2+\alpha})^{-1} \le (1+\lambda)^{-1} + (1+\lambda)^{-3/2+\alpha}.$$

Proof. Factoring and using the fact that $|(1+\lambda)^{-1}c\lambda^{1/2+\alpha}| \le c\lambda^{\alpha-1} \le 1/2$ and $(1-x)^{-1} \le 1 + 2x$ for $|x| \le 1/2$,

$$\begin{aligned} (1 + \lambda - c\lambda^{1/2+\alpha})^{-1} &= (1+\lambda)^{-1}(1 - (1+\lambda)^{-1}c\lambda^{1/2+\alpha})^{-1} \\ &\le (1+\lambda)^{-1}(1 + 2(1+\lambda)^{-1}c\lambda^{1/2+\alpha}) \\ &\le (1+\lambda)^{-1} + (1+\lambda)^{-3/2+\alpha}. \end{aligned} \tag{5.11}$$

□

Lemma 5.11. *Let X be a Poisson random variable with mean λ.*

$$\begin{aligned} \text{For} \quad 0 < x \le \lambda^{1/2}, \quad & P(X < \lambda - x\lambda^{1/2}) \le e^{-x^2/2} \quad \text{and} \\ & P(X > \lambda + x\lambda^{1/2}) \le e^{-x^2/3}. \end{aligned} \tag{5.12}$$

Proof. We have

$$\mathbb{E}[e^{\theta X}] = \sum_{k\geq 0} e^{\theta k} e^{-\lambda}\lambda^k/k! = e^{-\lambda}\sum_{k\geq 0}(\lambda e^{\theta})^k/k! = \exp(\lambda(e^{\theta}-1)).$$

Also,

$$P(e^{\theta X} \geq e^{\lambda\theta c}) = \begin{cases} P(X\geq c\lambda) & \text{if } \theta > 0 \\ P(X \leq c\lambda) & \text{if } \theta < 0. \end{cases}$$

Using Markov's inequality,

$$P(e^{\theta X}\geq e^{\lambda\theta c}) \leq e^{-\lambda\theta c}\mathbb{E}[e^{\theta X}] = \exp(\lambda(e^{\theta}-1-\theta c))$$

Optimizing in θ gives $\theta = \log c$ which is positive for $c > 1$ and negative for $c < 1$, and

$$\gamma(c) := e^{\theta} - 1 - \theta c = c - 1 - c\log c.$$

Expanding $\gamma(1+\delta)$ in an alternating Taylor series around $\delta = 0$,

$$\begin{aligned} \gamma(1+\delta) &\leq -\delta^2/2 + \delta^3/6 \quad \text{for} \quad |\delta| < 1, \text{ so} \\ &\leq \begin{cases} -\delta^2/2 & \text{for} \quad -1 < \delta \leq 0 \\ -\delta^2/3 & \text{for} \quad 0 \leq \delta < 1, \end{cases} \end{aligned}$$

using $\delta^3 \leq \delta^2$ for $\delta \in [0,1)$ and $\frac{1}{2} - \frac{1}{6} = \frac{1}{3}$. (5.12) follows for $0 < x < \lambda^{1/2}$ by letting $\delta = x\lambda^{-1/2}$. For $x = \lambda^{1/2}$ it follows by continuity of probability. □

Lemma 5.12. *Let (N_t) be a Poisson process with intensity r. Fix $\alpha \in (0, 1/2]$ and let*

$$\begin{aligned} \tau_1 &= \sup\{t : N_t - rt \geq 2(rt)^{1/2+\alpha}\} \quad \textit{and} \\ \tau_2 &= \sup\{t : N_t - rt \leq -2(rt)^{1/2+\alpha}\} \end{aligned}$$

denote the last passage time of N_t above/below the curve $rt \pm 2(rt)^{1/2+\alpha}$, respectively. If $r \in [1/2, 1]$, $t \geq 4$ and $t^{2\alpha} \geq 6/\alpha$ then

$$\begin{aligned} P(\tau_1 > t) &\leq 6t^{1-2\alpha}e^{-(rt)^{2\alpha}/3} \quad \textit{and} \\ P(\tau_2 > t) &\leq 6t^{1-2\alpha}e^{-(rt)^{2\alpha}/3}. \end{aligned}$$

Proof. Let f denote the function defined by $f(t) = rt + (rt)^{1/2+\alpha}$. By assumption, $rt \geq 1$ and $\alpha \leq 1/2$ so $(rt)^{\alpha} \leq (rt)^{1/2}$. Using Lemma 5.11,

$$P(N_t > f(t)) \leq e^{-(rt)^{2\alpha}/3}.$$

Since $|f'(t)| \leq 2r$ for any $t \geq 0$, f is Lipschitz with constant $2r$. Using this and the fact that $t \mapsto N_t$ is non-decreasing,

$$\{\sup_{s\in[t-1,t]} N_s - f(s) > 2r\} \subseteq \{N_t > f(t)\},$$

so taking a union bound over $t \in \{T+1, T+2, \dots\}$,

$$P(\sup_{t\geq T} N_t - f(t) > 2r) \leq \sum_{k\geq 1} e^{-(r(T+k))^{2\alpha}/3} \leq \int_T^{\infty} e^{-(rt)^{2\alpha}/3}dt.$$

Since $2(rt)^{1/2+\alpha}$ increases with t, we have $\{\tau_1 > t\} \subseteq \{\sup_{t\geq T} N_t - f(t) > 2r\}$ if $2r \leq (rt)^{1/2+\alpha}$, i.e., if $2r^{1/2-\alpha} \leq t^{1/2+\alpha}$, for which $t \geq 4$ suffices. It remains to estimate the above integral. Using Lemma 5.9 with $a = 0, \beta = 2\alpha, c = r^{2\alpha}/3$ and $x = T$ and noting $1 + a - \beta \geq 0$, if $2\alpha r^{2\alpha}/3 - (1-2\alpha)T^{-2\alpha} > 0$ then

$$\int_T^\infty e^{-(rt)^{2\alpha}/3} dt \leq (2\alpha r^{2\alpha}/3 - (1-2\alpha))T^{-2\alpha})^{-1} T^{1-2\alpha} e^{-(rT)^{2\alpha}/3}.$$

By assumption, $r^{2\alpha} \geq (1/2)^{2\alpha} \geq 1/2$, $1 - 2\alpha \in [0,1]$ and $T^{-2\alpha} \leq \alpha/6$, so $2\alpha r^{2\alpha}/3 - (1 - 2\alpha))T^{-2\alpha} \geq \alpha/6 > 0$ and the right-hand side above is at most $(6/\alpha)T^{1-2\alpha}e^{-(rT)^{2\alpha}/3}$. The lower bound is analogous – note Lemma 5.11 gives a somewhat better estimate in that case, but just be lazy and use $e^{-(rt)^{2\alpha/3}}$ as before. $\square$

References

[1] Andrea Baronchelli, Luca Dall'Asta, Alain Barrat, and Vittorio Loreto. Topology-induced coarsening in language games. *Physical Review E*, 73(1):015102, 2006.

[2] Andrea Baronchelli, Maddalena Felici, Vittorio Loreto, Emanuele Caglioti, and Luc Steels. Sharp transition towards shared vocabularies in multi-agent systems. *Journal of Statistical Mechanics: Theory and Experiment*, 2006(06):P06014, 2006.

[3] Andrea Baronchelli, Vittorio Loreto, and Luc Steels. In-depth analysis of the naming game dynamics: the homogeneous mixing case. *International Journal of Modern Physics C*, 19(05):785–812, 2008.

[4] A. Basak, R. Durrett, and E. Foxall. Diffusion limit for the partner model at the critical value. *Preprint on the ArXiv*, 2018. MR-3563195

[5] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. Statistical physics of social dynamics. *Reviews of modern physics*, 81(2):591, 2009.

[6] Luca Dall'Asta, Andrea Baronchelli, Alain Barrat, and Vittorio Loreto. Nonequilibrium dynamics of language games on complex networks. *Physical Review E*, 74(3):036105, 2006.

[7] K. Dzhaparidze and J.H. van Zenten. On bernstein-type inequalities for martingales. *Stochastic Processes and their Applications*, 93:109–117, 2001. MR-1819486

[8] Jean Jacod and Albert Shiryaev. *Limit Theorems for Stochastic Processes*. Springer, second edition, 2003. MR-1943877

[9] O. Kallenberg. *Foundations of Modern Probability Theory*. Springer, 2006. MR-1464694

[10] P. Protter. *Stochastic Integration and Differential Equations*. Springer-Verlag, second edition, 2005. MR-2273672