Scaling limits and Generic Bounds for Exploration Processes

Paola Bermolen   Matthieu Jonckheere   Jaron Sanders

March 7, 2018

Abstract

We consider exploration algorithms of the random sequential adsorption type both for homogeneous random graphs and random geometric graphs based on spatial Poisson processes. At each step, a vertex of the graph becomes active and its neighboring nodes become explored. Given an initial number of vertices N growing to infinity, we study statistical properties of the proportion of explored nodes in time using scaling limits. We obtain exact limits for homogeneous graphs and prove an explicit central limit theorem for the final proportion of active nodes, known as the *jamming constant*, through a diffusion approximation for the exploration process. We then focus on bounding the trajectories of such exploration processes on random geometric graphs, i.e. random sequential adsorption. As opposed to homogeneous random graphs, these do not allow for a reduction in dimensionality. Instead we build on a fundamental relationship between the number of explored nodes and the discovered volume in the spatial process, and obtain generic bounds: bounds that are independent of the dimension of space and the detailed shape of the volume associated to the discovered node. Lastly, we give two trajectorial interpretations of our bounds by constructing two coupled processes that have the same fluid limits.

1   Introduction

Exploration algorithms occur in a wide a range of processes, for example in the evolution of parking processes [2], as well as random sequential adsorption processes [7]. A typical exploration algorithm works as follows: Assume there exists a binary relation between items $V = \{1, \dots, N\}$, to which we associate a relation graph where nodes are items such that two items are neighbors if they are related. Let A_t be the set of active items at step t and B_t the set of explored items. We assume that, initially, $A_0 = B_0 = \{\emptyset\}$. Then we consider the following exploration process:

(i) select $I_t \subseteq V \setminus \{A_t \cup B_t\}$ and determine its neighbors in the set of non-explored items, $\mathcal{N}_t \subseteq V \setminus \{A_t \cup B_t \cup I_t\}$, and

(ii) update A_t and B_t by setting $A_{t+1} = A_t \cup I_t$ and $B_{t+1} = B_t \cup \mathcal{N}_t$.

In case I_t is a single item at each step, the algorithm discovers in a greedy manner independent sets of the relation graph. If I_t is a subgraph, $\mathcal{N}_t$ is the set of non-active, non-explored neighbors of this subgraph at step t.

Exploration algorithms can also be identified in diverse applications. For example, they can be linked to communication procedures in communication networks [7]. In this context, the relation graph might be the outcome of random spatial effects, i.e. nodes could

be interacting through a geometry which itself can be random. This is certainly the case in wireless networks in which acceptable radio conditions have been defined based on the level of admissible interference between two competing nodes, so conditions that determine whether simultaneous communication is possible. These interference constraints can be modeled using a hardcore interference graph, in which an edge is present between two nodes if their radio conditions would impede a simultaneous communication [3].

Relation graphs and associated exploration processes can also be used to describe blockade effects in complex systems of interacting particles. Consider for instance gases of ultracold atoms that can change between a so-called Rydberg state and a ground state [8]. The essential feature of these gases is that once an atom is in its Rydberg state, it prevents neighboring atoms from reaching their Rydberg state. By modeling the dynamics of the excitation process as an exploration process on a relation graph (specifically, the canonical Erdös–Rényi (ER) graph), the statistical properties of the proportion of atoms ultimately in the Rydberg state were studied in [18]. This article formalizes and generalizes the heuristical arguments provided there.

The dimension needed to represent exploration algorithms as a Markov process is in general equal to the size of the underlying graph, N, and this impedes both the analysis and the ability to perform explicit calculations. To overcome these difficulties, we first assume that a homogeneity assumption holds for the relation between items. This allows us to describe the dynamics using a one-dimensional Markov process, instead of a prohibitively more complicated N-dimensional description. While this is a coarse simplification for many problems, it has the advantage that the analysis remains tractable. We then use classical tools of probability theory, specifically fluid limits and diffusion approximations, to derive computable characterizations of the performance of these exploration algorithms. In particular we provide not only convergence results but also error bounds. These techniques furthermore allow us to prove a Law of Large Numbers (LLN) as well as a Central Limit Theorem (CLT) for the proportion of active nodes in the jamming limit, i.e. at the end of the exploration.

While the techniques to obtain a LLN for the exploration process of the ER graph are well-known in the literature, see for instance the survey [5], our first contribution is to use such approaches to also obtain a LLN and CLT for the jamming constant (the final proportion of active nodes). Our results also cover exploration processes on other, sufficiently symmetric graphs. By next building on the recent results in [1], we can also give error bounds for the diffusion approximation, which in turn allows to construct confidence intervals for the jamming constant. These results have earlier been made available in the PhD thesis [17], as well as in the technical report [4].

Our new, second contribution consists of using the fluid and diffusion limits of exploration processes to obtain bounds for the more involved case of random geometric graphs. When the neighborhood relation is constructed using a spatial Poisson process and a geometric distance (say the Euclidean one), the correlations between edges of the graph have a much more complex structure. As underlined before, the exploration process can then not be reduced to the analysis of a one-dimensional Markov process. However, by building on a fundamental relationship between the number of explored nodes and the discovered volume in the spatial process (translating the properties of the Poisson process), we obtain bounds for the fluid limit of the exploration process. For both our lower and upper bound on the fluid limit, we propose a coupling of the original process with processes which have a lower and higher exploration rate, respectively.

The rest of the paper is structured as follows. In Section 2, we set the model for homogeneous and symmetric random graphs and obtain classical scaling limits in this context. Then, building on diffusion approximations errors bounds, we provide diffusion approximations for both the exploration process and the jamming constant. In Section 3, we characterize an important asymptotic relation between the mean discovered volume and the mean number of explored points for random relation graphs associated to Poisson point processes. Building on this relation, we obtain asymptotic bounds for the exploration process and provide trajectorial couplings for these bounds. Finally, in Section 4, we provide simulations of the exploration process and the jamming constant of the geometric random graph, which we compare to our bounds.

2 Fluid limit and diffusion approximation for homogeneous graphs

Assume from here on that precisely one item is selected in each step. Assume also the following homogeneity property on the graph $\mathcal{G}$:

Assumption 2.1. *If $(\mathcal{G}_1, \mathcal{G}_2)$ is a partition of $\mathcal{G}$, then the mean number of edges from $\mathcal{G}_1$ to $\mathcal{G}_2$ depends only on the size of the partition, i.e. $|\mathcal{G}_1|$ and $|\mathcal{G}_2|$.*

While Assumption (2.1) is not valid in cases such as random geometric graphs and random graphs with generic degree distribution, it is crucial to be able to describe the exploration process in one dimension. Assumption (2.1) can however be considered a reasonable approximation for many systems, and it is for example satisfied by ER random graphs. We refer the reader to [2] for a study of scaling limits in infinite dimension that applies to a larger class of problems.

If we let Z_n denote the number of explored items at step n, so $Z_n = |A_n \cup B_n|$, we have that

$$Z_0 = 0, \quad \text{and} \quad Z_n = Z_{n-1} + 1 + \xi_n. \tag{1}$$

Here, ξ_n denotes the number of neighbors that the selected item has at step n in the remaining non-explored portion of the graph.

The distribution of ξ_n depends under Assumption (2.1) only on Z_{n-1}, which we will denote by $\xi_{Z_{n-1}}$ with a slight abuse of notation. Assumption (2.1) also implies that the process $\{Z_n\}_{n\in\mathbb{R}}$ is a discrete Markov process that takes values in $\{0, \ldots, N\}$, is strictly increasing, and has in N an absorbing state. The transition probabilities at step n of the process $\{Z_n\}_{n\in\mathbb{R}}$ are therefore given by

$$p_{xy}(n) = \mathbb{P}[Z_n = y | Z_{n-1} = x] = \mathbb{P}[\xi_x = y - x - 1] \quad \text{for} \quad y > x. \tag{2}$$

If we now denote by $p_N(\cdot, x)$ the distribution of the number of neighbors in $\mathcal{G}_2$ of any vertex $i \in \mathcal{G}_2$ given that $(\mathcal{G}_1, \mathcal{G}_2)$ is a partition of $\mathcal{G}$ with $|\mathcal{G}_1| = x$, the transition probabilities can be written as

$$p_{x,x+k+1}(n) = \mathbb{P}[\xi_x = k] = p_N(k, x), \quad \text{with} \quad k \geq 0. \tag{3}$$

In case of the ER random graph, in which an edge exists between a pair of nodes with probability p, the transition probabilities are given by the Binomial distribution, i.e.

$$p_N(k, x) = \binom{N - x - 1}{k} p^k (1 - p)^{N - x - k - 1}. \tag{4}$$

Given a partition $(\mathcal{G}_1, \mathcal{G}_2)$ of $\mathcal{G}$ such that $|\mathcal{G}_1| = x$, we denote by $\gamma_N(x)$ the mean and by $\psi_N(x)$ the variance of the number of neighbors in $\mathcal{G}_2$ of a given vertex $i \in \mathcal{G}_2$, i.e.

$$\gamma_N(x) = \sum_{k=0}^{N-1} k p_N(k, x), \quad \psi_N(x) = \sum_{k=0}^{N-1} (k - \gamma_N(x))^2 p_N(k, x). \tag{5}$$

We also define $\bar{\gamma}_N = \sup_x \gamma_N(x)$, $\bar{\psi}_N = \sup_x \psi_N(x)$.

Now consider the scaled process defined as the piece-wise constant trajectory process of

$$Z_t^N = \frac{Z_{[tN]}}{N} \tag{6}$$

for all $t \geq 0$. Here, $[x]$ is the integer part of x. We derive a fluid limit for Z_t^N in Proposition 2.2. While the proof of convergence relies on classical techniques [5], we leverage these tools to also obtain error bounds along the way. The proof of Proposition 2.2 is deferred to Appendix A.

Proposition 2.2. *If there exists a (C_L)-Lipschitz continuous function γ on $\mathbb{R}^+$ such that*

$$\sup_x \left| \gamma_N(x) - \gamma\left(\frac{x}{N}\right) \right| \leq \delta_N, \tag{7}$$

then for $p > 1$ and $T > 0$,

$$\| \sup_{s \in [0,T]} |Z_s^N - z(s)| \|_p \leq \left(\delta_N T + \frac{1 + \bar{\gamma}_N}{N} + \kappa_p \left\| \frac{M_{[TN]}}{N} \right\|_p \right) e^{C_L T}, \tag{8}$$

where $\kappa_p = p/(p-1)$, and $z(t)$ denotes the solution to the deterministic differential equation

$$\dot{z}(t) = 1 + \gamma(z(t)), \quad \text{with} \quad z(0) = 0, \quad \text{for} \quad t \leq T^* = \min\{s : z(s) = 1\}. \tag{9}$$

Here, $M_n = Z_n - \sum_{i=0}^n (1 + \gamma_N(Z_i))$ denotes a global martingale.

For $p = 2$, the bound reduces to

$$\| \sup_{t \in [0,T]} |Z^N(t) - z(t)| \|_2 \leq \left(\delta_N T + \frac{1 + \bar{\gamma}_N}{N} + 2\sqrt{\frac{\bar{\psi}_N T}{N}} \right) \exp(C_L T) = \omega_N. \tag{10}$$

Corollary 2.3. *If the distribution of the number of neighbors is such that $\delta_N \to 0$ as $N \to \infty$, $\bar{\gamma}_N = o(N)$, and $\bar{\psi}_N = o(N)$, then the scaled process Z_t^N converges to $z(t)$ in L^1 uniformly on compact time intervals.*

We now proceed and derive a diffusion approximation theorem for the scaled number of explored items Z_t^N. The convergence proof relies similarly on classical techniques, which we again leverage to determine error bounds. To that end, we apply results of [11] which are based on results by Komlós–Major–Tusnády [9, 10]. The results in [11] allow one to construct a Brownian motion and either a Poisson process or random walk on the same probability space. Since we are concerned with discrete time, we need to consider the random walk case, see also [1]. In order to obtain explicit error bounds, we impose stronger assumptions on the transitions probabilities than would be needed when only proving convergence. Our proof of Proposition 2.4 can be found in Appendix B.

Proposition 2.4. *If there exists a function p on $(\mathbb{N}, \mathbb{R}_+)$, and a sequence $(\epsilon_k)_{k=0,1,\ldots}$ such that*

$$|p_N(k, [Nx]) - p(k, x)| \leq \frac{\epsilon_k}{N},$$

$$|p(x, x + k) - p(y, y + k)| \leq M\epsilon_k |x - y|,$$

$$\sum_k k^2 |p(x, x+k)^{1/2} - p(y, y+k)^{1/2}|^2 \leq M|x-y|^2 \tag{11}$$

and

$$\sum_k k\epsilon_k^{1/2} < \infty, \tag{12}$$

and if γ is twice differentiable with bounded first and second derivatives, then the process

$$W_t^N = \sqrt{N}(Z_t^N - z(t)) \tag{13}$$

converges in distribution towards W_t, the unique solution of the stochastic differential equation

$$dW(t) = \gamma'(z(t))W(t)\,dt + \sqrt{\beta'(t)}\,dB_1(t). \tag{14}$$

Here, $B_1(t)$ denotes a standard Brownian motion, $\beta(t) = \int_0^t \psi(z(s))\,ds$, and $z(t)$ is the solution of (9). Furthermore,

$$\mathbb{E}\left[\sup_{t \leq T} |W_t^N - W_t|\right] \leq C\frac{\log(N)}{\sqrt{N}}. \tag{15}$$

2.1 LLN and CLT for the hitting time

The exploration algorithm finishes at time

$$T_N^* = \inf\{\tau \in \mathbb{N}_+ | Z_\tau = N\} \leq N < \infty, \tag{16}$$

which is a hitting time for the Markov process. Since the algorithm adds precisely one node at each step, we have that the final number of active items is exactly T_N^*, i.e. $A_{T_N^*} = T_N^*$. Because we wish to determine the statistical properties of $A_{T_N^*}$, we will seek not only a first-order approximation for T_N^*, but also prove a central limit theorem result as the initial number of items N goes to infinity.

Since the exploration process converges to the fluid limit $z(t)$, we can anticipate that an appropriately scaled hitting time T_N^* converges to T^*, the solution to $z(T^*) = 1$. This intuition is formalized in the LLN result for T_N^*/N in Proposition 2.5. Its proof is in Appendix C.

Proposition 2.5. *For all $\delta > 0$,*

$$\mathbb{P}\left[\left|\frac{T_N^*}{N} - T^*\right| \geq \delta\right] \leq \frac{2\omega_N}{\delta}. \tag{17}$$

Moreover if γ is continuous, non-increasing with $\gamma(1) = 0$, then there exists a constant C for sufficiently small δ so that

$$\left\|\frac{T_N^*}{N} - T^*\right\|_2 \leq C\omega_N := \Omega_N \tag{18}$$

Corollary 2.6. *If the distribution of the number of neighbors is such that $\delta_N \to 0$ as $N \to \infty$, $\bar{\gamma}_N = o(N)$, and $\bar{\psi}_N = o(N)$, then the proportion of active items T_N^*/N converges in L^1 to T^*.*

The LLN in Proposition 2.5 provides us formally with a candidate, T^*, around which to center T_N^*/N and subsequently prove the CLT result in Proposition 2.7. We defer to Appendix D for its proof.

Proposition 2.7. *There exist constants C_1, C_2 such that the expectation of the random variable $\sqrt{N}(T_N^*/N - T^*)$ centered around $-W_T^*$ is bounded by*

$$\mathbb{E}\left[\left|\sqrt{N}\left(\frac{T_N^*}{N} - T^*\right) + W_{T^*}\right|\right] \tag{19}$$

$$\leq C_1 \omega_N^2 \sqrt{N} + \left(\bar{\psi}_N \Omega_N + \frac{\bar{\psi}_N}{N}\right)^{\frac{1}{2}} + C_2 \frac{\log(N)}{\sqrt{N}} + \frac{1 + \bar{\gamma}_N}{\sqrt{N}}.$$

Moreover, if the distribution of the number of neighbors is such that $\delta_N = o(1/\sqrt{N})$, $\bar{\gamma}_N = o(\sqrt{N})$, and $\bar{\psi}_N = o(N^{1/4})$, then $\sqrt{N}(T_N^*/N - T^*)$ converges in L^1 to W_{T^*} that is a centered Gaussian random variable with variance

$$\sigma^2 = m(T^*) = \mathbb{E}[W_{T^*}^2], \tag{20}$$

and where $m(t) = \mathbb{E}[W_t^2]$ solves the differential equation

$$\dot{m}(t) = -2\dot{\gamma}(z(t))m(t) + \dot{\beta}(t), \quad with \quad m(0) = 0. \tag{21}$$

2.2 Case: ER random graph

We now illustrate our results through an application of Propositions 2.2, 2.7 to the case of the ER random graph. We point interested readers to [6] for the analysis of another, more involved example of an exploration on a graph with sufficient symmetry.

Suppose that given N the graph $\mathcal{G} = \mathcal{G}(N, c/N)$ is a sparse ER graph, i.e. $p_N(\cdot, x)$ is the probability mass function of the binomial distribution $\mathrm{Bin}(N - x - 1, c/N)$ with $c > 0$. Additionally, suppose that $N - 1$ is Poisson distributed with parameter h. The mean and variance of $p_N(., x)$ are then given by

$$\gamma_N(x) = (N - x - 1)\frac{c}{N}, \quad \psi_N(x) = (N - x - 1)\frac{c}{N}\left(1 - \frac{c}{N}\right). \tag{22}$$

Let $\gamma(x) = c(1 - x)$. Condition (7) is then satisfied with $\delta_N = c/N$, and as Lipschitz constant $C_L = c$ suffices. Moreover, $\bar{\gamma}_N, \bar{\psi}_N \leq c$. The deterministic differential equation in (9) reads

$$\dot{z}(t) = 1 + c(1 - z(t)) = (1 + c) - cz(t), \quad with \quad z(0) = 0. \tag{23}$$

This differential system can be explicitly solved, giving

$$z(t) = \frac{1 + c}{c}\left(1 - e^{-ct}\right). \tag{24}$$

Observe that $\lim_{t \to \infty} z(t) = (1 + c)/c > 1$, implying the existence of a finite, constant solution T^* to $z(T^*) = 1$. Solving this equation, we find that

$$T^* = \frac{\ln(1 + c)}{c}, \tag{25}$$

which agrees with literature [12].

We next calculate T^*'s variance using Proposition 2.7. First, we verify its assumptions. Relations between the binomial coefficients and the Poisson distribution are well studied. Defining

$$p_N(k, [xN]) = \binom{N - [Nx] - 1}{k} \left(\frac{c}{N}\right)^k \left(1 - \frac{c}{N}\right)^{N - [Nx] - 1}, \tag{26}$$

and using (for instance) the Stein–Chen method [20], we have that

$$|p_N(k, [xN]) - p(k, x)| \leq \frac{c}{N} p(k, x), \tag{27}$$

which shows that the assumptions of Proposition 2.4 are satisfied. Moreover, the differential equation for $\beta(t)$ is given by

$$\dot{\beta}(t) = \psi(z(s)) = (1 + c)\mathrm{e}^{-ct} - 1, \tag{28}$$

and the solution to (21) is then

$$m(t) = \mathrm{e}^{-2ct}(1 - \mathrm{e}^{ct})(\mathrm{e}^{ct} - 2c - 1)\frac{1}{2c}, \tag{29}$$

ultimately leading to

$$\sigma^2 = \frac{m(T^*)}{(1 - \gamma(1))^2} = \frac{c}{2(c+1)^2}. \tag{30}$$

3 Random Geometric Graph

In this section we consider the problem of random sequential adsorption (RSA) [7], which can be described as an exploration process on a Random Geometric Graph (RGG). Due to the strong spatial correlation between points in RSA, a quantitative analysis is notoriously difficult. We propose instead to use the fluid limits discussed in Section 2 as a stepping stone to obtain trajectorial bounds on the actual exploration process. As a consequence, we obtain a lower and upper bound for the jamming constant.

3.1 RSA as an exploration process $\{Z_n\}_{n \geq 0}$ on a RGG

Let Φ be a homogeneous Poisson point process with intensity $\lambda > 0$ in a finite box $\mathcal{C} \subset \mathbb{R}^d$. We consider a RGG $\mathcal{G}$ in which we identify each vertex with a point of Φ, and there is an edge between every two points $X_i, X_j \in \Phi$ within distance $r > 0$ of each other, i.e. $|X_i - X_j| \leq r$. The points X_i and X_j are then considered neighbors. Recall that given the number of points $N = \Phi(\mathcal{C})$ in $\mathcal{C}$, the points $X_1, X_2, \ldots, X_N$ are independently and uniformly at random distributed in $\mathcal{C}$.

Let Z_n denote the number of points (vertices of $\mathcal{G}$) explored at step n, and set $Z_0 = 0$. Consider any sequence of sets $\{S_1, S_2, \ldots\}$ defined on $\mathcal{C}$. Let $E_n = \cup_{k \leq n} S_k$ denote the union of these sets until step n. Each k-th set S_k corresponds to the new area of $\mathcal{C}$ being explored at step k, while E_n corresponds to the total explored area up to and including step n.

In particular, for RSA at each n-th step of the process, we choose a point X_{i_n} uniformly at random among all non-explored points, and next consider X_{i_n} and all of its neighbors in $S_n = B^*(X_{i_n}, r) \cap E_{n-1}^c$ explored. Here, $B^*(x, r) = B(x, r) \setminus \{x\}$ denotes a ball of radius r

centered around x but excluding its center x. The number of explored vertices up to time n, Z_n, satisfies the stochastic recursion:

$$Z_n = 1 + Z_{n-1} + \Phi(S_n). \tag{31}$$

Note that Z_n thus always increases at least by one. The exploration process ends at the first time T_N^* such that $Z_{T_N^*} = N$, i.e. when all points of $\mathcal{G}$ have been explored.

The process Z_n is a discrete Markov process with respect to the filtration

$$\mathcal{F}_n = \sigma(\cup_{i \leq n}\{Z_i, S_i\}), \tag{32}$$

and lives on a finite state space $\{0, 1, \ldots, N\}$ with absorbing state N. However the computation of its scaling limit and of the stopping time T_N^* are prohibitively complicated, because the need to track the history of the explored volume impedes a direct analysis of its drift. We therefore instead resort to deriving upper and lower bounds for Z_n and T_N^*.

3.2 Fluid limit properties of $\{Z_n\}_{n \geq 0}$

In the following proposition, we state the fluid limit convergence result for the exploration process as proven by Penrose et al. [15]. From this fundamental result, we derive a similar convergence result for the fraction of explored volume. Let $Z_t^N := Z_{[tN]}/N$ denote the scaled version of the exploration process Z_n in (31).

Proposition 3.1. *There exists a deterministic function $z : \mathbb{R}^+ \to [0, 1]$ such that*

$$Z_t^N \underset{N \to \infty}{\to} z(t) \text{ almost surely, and in } \mathcal{L}^\infty. \tag{33}$$

As a consequence there also exists a function $\eta : \mathbb{R}^+ \to [0, 1]$ such that

$$\frac{1}{|\mathcal{C}|} \sum_{i=1}^{[tN]} |S_i| \underset{N \to \infty}{\to} \eta(t) \text{ almost surely, and in } \mathcal{L}^\infty. \tag{34}$$

Proof. The first result, (33), was proved in [14, 13]. The second claim, (34), follows using a similar proof methodology. Specifically, we aim to apply [13, Thm. 3.2], which requires us to verify a few properties of the functional $t \to (1/|\mathcal{C}|)\sum_{i=1}^{[tN]}|S_i|$.

First, observe that $t \to (1/|\mathcal{C}|)\sum_{i=1}^{[tN]}|S_i|$ is translation invariant. Second, using percolation estimates, it is proved in [13] that that for all possible realizations of the marked point process there exists a (random) radius R such that the exploration process at the origin (i.e. the state explored or not of a point placed at the origin) stays unmodified by any change in the realization of the marked point process outside a ball of radius R. This implies that the same property holds for the exploration of the volumes associated to the points of the point process. Lastly, note that $t \to (1/|\mathcal{C}|)\sum_{i=1}^{[tN]}|S_i|$ is polynomially bounded since $(1/|\mathcal{C}|)\sum_{i=1}^{[tN]}|S_i| \leq t\lambda|B(\cdot, r)|$.

Having established these three properties, we are now in position to apply [13, Thm. 3.2], which completes the proof. $\square$

3.3 Bounding RSA's fluid limit $z(t)$

Having obtained a fluid limit for the fraction of explored volume, we now prove a differential equation fundamentally relating the limiting fraction of explored volume $\eta(t)$ and the limiting fraction of explored points $z(t)$.

Proposition 3.2. *For all $t \leq T^*$,*

$$\dot{z}(t) = 1 + (1 - z(t))\frac{\dot{\eta}(t)}{1 - \eta(t)} = 1 + \dot{\eta}(t)\mathrm{e}^{-\int_0^t \frac{1}{1-z(s)}ds}. \tag{35}$$

Proof. First observe that $z(t)$ is a differentiable function. Indeed for all fixed N, t, and h,

$$\left|\mathbb{E}\Big[\frac{Z_{[(t+h)N]}}{N}\Big] - \mathbb{E}\Big[\frac{Z_{[tN]}}{N}\Big]\right| \leq vh, \tag{36}$$

which implies that $z(t)$ is globally Lipschitz continuous. Hence $z(t)$ is almost everywhere differentiable with respect to the Lebesgue measure. Using the same argument, we can show that η is differentiable.

Now recall that since Φ is a Poisson point process, given the number of points in $\mathcal{C}$, $N = \Phi(\mathcal{C})$, the position of each point is independently and uniformly distributed in $\mathcal{C}$. Thus by definition of the exploration dynamics, we obtain that

$$\mathbb{E}\Big[Z_{[tN]}^N|\mathcal{F}_{[tN]-1}\Big] \overset{(33)}{=} Z_{[tN]-1}^N + \frac{1}{N} + \frac{1}{N}\mathbb{E}[\Phi(S_{[tN]})|\mathcal{F}_{[tN]-1}]$$

$$= Z_{[tN]-1}^N + \frac{1}{N} + \frac{1}{N}\mathbb{E}[\mathrm{Bin}(N - Z_{[tN]-1} - 1, p_i)|\mathcal{F}_{[tN]-1}], \tag{37}$$

where $p_i = |S_i|/(|\mathcal{C}| - \sum_{j \leq [tN]-1}|S_j|)$. Since $|\mathcal{C}| = N/\lambda$, it follows that

$$\mathbb{E}\Big[N(Z_{[tN]}^N - Z_{[tN]-1}^N)|\mathcal{F}_{[tN]-1}\Big] = 1 + \Big(1 - Z_{[tN]-1}^N - \frac{1}{N}\Big)\frac{\lambda\mathbb{E}[|S_{[tN]}||\mathcal{F}_{[tN]-1}]}{1 - \frac{1}{N}\sum_{j\leq[tN]-1}|S_j|}. \tag{38}$$

Hence writing $\Gamma_N(i) = (1/N)\sum_{j \leq i}|S_j|$, we obtain after summing these differences that

$$\mathbb{E}\Big[\frac{Z_{[tN]}}{N}|\mathcal{F}_{[tN]-1}\Big] = \frac{[tN]}{N} + \frac{1}{N}\sum_{i=1}^{[tN]}\Big(1 - Z_{i-1}^N - \frac{1}{N}\Big)\frac{\lambda N\mathbb{E}[\Gamma_N(i) - \Gamma_N(i-1)|\mathcal{F}_{[tN]-1}]}{1 - \Gamma_N(i-1)}, \tag{39}$$

$$= \frac{[tN]}{N} + \sum_{i=1}^{[tN]}\Big(1 - Z_{i-1}^N - \frac{1}{N}\Big)\frac{\lambda\mathbb{E}[\Gamma_N(i) - \Gamma_N(i-1)|\mathcal{F}_{[tN]-1}]}{1 - \Gamma_N(i-1)}, \tag{40}$$

Using the almost sure convergence of both Z_t^N and Γ_N to continuous functions, and since both functional are increasing in time, we have uniform almost sure convergence on compact (macroscopic) time intervals. As a consequence, we obtain the almost sure convergence of the right-hand side of the previous equality to the Riemann-Stieljes integral $\int_0^t (1-z_s)/(1-\eta_s)\,\mathrm{d}\eta_s$ which is equal to $\int_0^t (1-z_s)/(1-\eta_s)\dot{\eta}_s\,\mathrm{d}s$, using the differentiability of η. Therefore,

$$\dot{z}(t) = 1 + (1 - z(t))\frac{\dot{\eta}(t)}{1 - \eta(t)} \quad \text{or equivalently} \quad \dot{z}(t) = 1 + \dot{\eta}(t)\mathrm{e}^{-\int_0^t \frac{1}{1-z(s)}ds}. \tag{41}$$

This concludes the proof. $\qquad\square$

It is important to note that we do not have an explicit expression available for either $z(t)$ or $\eta(t)$, hence the differential system is not solvable. But now that we have with Proposition 3.2 a relation between the limiting fraction of explored volume $\eta(t)$ and the limiting fraction of explored vertices $z(t)$ at our disposal, we can find upper and lower bounds $u(t)$, $l(t)$ for $z(t)$ by deriving upper and lower bounds for $\eta(t)$. These upper and lower bounds immediately also provide us with lower and upper bound for the jamming constant, since

$$T^{\text{lower}} := \inf\{t > 0 | u(t) = 1\} \leq T^* \leq \inf\{t > 0 | l(t) = 1\} =: T^{\text{upper}}. \tag{42}$$

Proposition 3.3. *For $t \in [0, T^*]$,*

$$l(t) \leq z(t) \leq u(t), \tag{43}$$

where $l(t)$ and $u(t)$ are the solutions to

$$\dot{l}(t) = 1 + c\left(1 - \frac{3ct}{1 - l(t)}e^{-\int_0^t \frac{1}{1-l(s)}\, \mathrm{d}s}\right)e^{-\int_0^t \frac{1}{1-l(s)}\, \mathrm{d}s}, \quad l(0) = 0, \tag{44}$$

and

$$\dot{u}(t) = 1 + ce^{-\int_0^t \frac{1}{1-u(s)}\, \mathrm{d}s}, \quad u(0) = 0, \tag{45}$$

i.e. $u(t) = ct + (t - (1/c))\ln(1 - ct)$, respectively. Here, $c = \lambda|B(\cdot, r)|$.

Corollary 3.4. *As $c \downarrow 0$, the spatial process has fluid limit $z(t) = (1 + c)t - \frac{1}{2}ct^2 + O(c^2)$.*

The proof of Corollary 3.4 can be found in Appendix E. We now prove Proposition 3.3, which is actually a direct consequence of the volume bounds in Lemma 3.5.

Lemma 3.5. *For every step $k = 1, \ldots, N$, with $v = |B(\cdot, r)|$,*

$$v\left(1 - \frac{(k-1)3v}{|\mathcal{C}| - \sum_{j<k} S_j}\right) \leq \mathbb{E}[|S_k|] \leq v. \tag{46}$$

Proof. The right inequality in (46) is immediate by noting that $|S_i^N| = |B^*(X_i, r) \cap E_{i-1}^c| \leq |B^*(\cdot, r)|$. The fluid limit upper bound in Proposition 3.3 then follows from η's definition, since

$$\dot{\eta}(t) = \lim_{N \to \infty} \lambda\mathbb{E}[|S_{[tN]}|] \leq \lambda v = c. \tag{47}$$

We next prove the left inequality in (46). Let X_{i_k} be the selected point at step k and define the event

$$\mathcal{E}_k = \{B^*(X_{i_k}, r) \cap E_{k-1} = \emptyset\}.$$

Consequently, after decomposing and by strict positivity of $|S_k|$,

$$\mathbb{E}[|S_k|] = \mathbb{E}[|S_k||\mathcal{E}_k]\mathbb{P}[\mathcal{E}_k] + \mathbb{E}[|S_k||\mathcal{E}_k^c]\mathbb{P}[\mathcal{E}_k^c] \geq \mathbb{E}[|S_k||\mathcal{E}_k]\mathbb{P}[\mathcal{E}_k] = v\mathbb{P}[\mathcal{E}_k]. \tag{48}$$

We next use that the unexplored points are uniformly distributed in E_{k-1}^c to obtain that

$$\mathbb{P}[\mathcal{E}_k] \geq 1 - (k-1)\mathbb{P}[\exists_{j<k} : X_{i_k} \in B(X_{i_j}, 2r) \setminus B(X_{i_j}, r)] = 1 - \frac{(k-1)3v}{|\mathcal{C}| - \sum_{j<k}|S_j|}. \tag{49}$$

Together with (48), this gives the left inequality in (46). To prove the fluid limit lower bound in Proposition 3.3, gather that

$$\dot{\eta}(t) = \lim_{N \to \infty} \lambda\mathbb{E}[|S_{[tN]}|]$$

$$\overset{(49)}{\geq} \lambda v \lim_{N \to \infty} \mathbb{P}[\mathcal{E}_{[tN]}] = c\left(1 - \frac{3ct}{1 - \eta(t)}\right) \overset{(41)}{=} c\left(1 - \frac{3ct}{1 - z(t)}\right)e^{-\int_0^t \frac{1}{1-z(s)}ds} \tag{50}$$

This completes the proof. □

10

3.4 Trajectorial bounds for $\{Z_n\}_{n\geq 0}$

We now construct trajectorial upper and lower bounds for the process $\{Z_n\}_{n\geq 0}$ by constructing two couplings that have the same fluid limits as the upper and lower bound in Section 3.3. These lead in turn to lower and upper bounds for the jamming constant $T_Z^* = \inf\{t \geq 0 | Z_t = N\}$, respectively.

3.4.1 Upper bound process $\{U_n\}_{n\geq 0}$

We define a new process $\{U_n\}_{n\geq 0}$ that corresponds to an exploration process of $\mathcal{G}$ with a higher discovery rate. As before, at each step (say step n) we choose a point X_{i_n} uniformly at random among the non-explored points of the process $\{Z_n\}_{n\geq 0}$. There are now two possible situations:

Case 1: X_{i_n} is also unexplored for $\{U_n\}$. We define a new set $\tilde{S}_n = B^*(X_{i_n}, \tilde{r}_n) \cap \tilde{E}_{n-1}^c$, with $\tilde{r}_n$ chosen such that the area of $\tilde{S}_n$ and that of a free ball $v = |B^*(\cdot, r)|$ coincide, that is $|\tilde{S}_n| = v$. Here $\tilde{E}_i^c = \cup_{j=1}^{i-1}\tilde{S}_j$.

Case 2: X_{i_n} is already explored for $\{U_n\}$. We now choose a different point X'_{i_n} uniformly at random from the non-explored points of the process $\{U_n\}$, and consider it instead as the point that will be explored for the process $\{U_n\}$. We then proceed as in the previous case, by again letting $\tilde{S}_n = B^*(X_{i_n}, \tilde{r}_n) \cap \tilde{E}_{n-1}^c$ with $\tilde{r}_n$ such that $|\tilde{S}_n| = v$.

In both cases, the exploration process is updated – as for the original process – according to the following recursion equation:

$$U_n = 1 + U_{n-1} + \Phi(\tilde{S}_n). \tag{51}$$

Note that crucially both S_n and $\tilde{S}_n$ are constructed using the point X_{i_n} initially chosen by the process $\{Z_n\}_{n\geq 0}$.

Now let $T_U^* = \inf\{t > 0 | U_t = N\}$ be the time at which the coupled process $\{U_n\}_{n\geq 0}$ completes the exploration of the graph. The coupling has been constructed such that the following holds.

Proposition 3.6. *The process $\{U_n\}_{n\geq 0}$ is such that $U_n \geq Z_n$ almost surely, and $T_U^* \leq T_Z^*$ with probability one.*

Proof. We couple the two explorations processes by using the same spatial point process Φ and the same exploration order for both processes. Next, we prove that the two processes stay ordered using induction: for the first step $n = 1$, both process coincide and inequality holds. Consider next any $n > 1$ for which $U_{n-1} \geq Z_{n-1}$. By construction of S_n and $\tilde{S}_n$ we obtain that $U_n = 1 + U_{n-1} + \Phi(\tilde{S}_n) \geq 1 + Z_{n-1} + \Phi(S_n) = Z_n$ almost surely.

Since we have now shown that at each step $U_n \geq Z_n$ almost surely, it follows immediately that $T_U^* \leq T_Z^*$ with probability one. This completes the proof. $\square$

While the transition probabilities of the process $\{Z_n\}_{n\geq 1}$ cannot readily be explicitly calculated due to the spatial correlations between the points $\{X_1, X_2, \ldots\}$, they can be calculated for the process $\{U_n\}_{n\geq 1}$, as we prove in the Proposition 3.7.

Proposition 3.7. *Consider the scaled process $U_t^N = U_{[tN]}/N$, then for $T > 0$,*

$$\lim_{N \to \infty} \mathbb{E}[\sup_{s \in [0,T]} |U_s^N - u(s)|] = 0, \tag{52}$$

where $u(t)$ is the solution to (45).

Proof. By Proposition 3.6 we have that $Z_n \leq U_n$ almost surely. This implies that the corresponding fluid limits verify the same inequality, i.e. $z(t) \leq u(t)$. The convergence of the bounding process towards its fluid limit follows from the arguments given in Section 2, see in particular the arguments surrounding (8). $\qquad\square$

3.4.2 Lower bound process $\{L_n\}_{n \geq 0}$

Simultaneously with the construction of the process $\{Z_n\}_{n \geq 0}$, we are now going to construct a new process $\{L_n\}_{n \geq 0}$ that also corresponds to an exploration on $\mathcal{G}$ but then with a lower discovery rate. This then provides us with a trajectorial lower bound on Z_n, leading to an upper bound for the jamming constant.

Before choosing a new point X_{i_n} to explore for step n, consider the event

$$\mathcal{P}(d_n) = \Big\{ X_{i_n} \in \mathcal{C} \backslash \bigcup_{j < n} B(X_{i,j}, r + d_n) \Big\}, \tag{53}$$

where $d_n \geq r$ denotes a perimeter radius. This event has two properties, namely that:

(i) $\mathbb{P}[\mathcal{P}(d_n)]$ is decreasing and continuous in d_n and,

(ii) $\mathbb{P}[\mathcal{P}(r)] = 1 - P(X_{i_n} \in \cup_{j<n} B(X_{i_j}, 2r)) \geq \max\Big\{ 0, 1 - \frac{3(n-1)v}{|\mathcal{C}| - \sum_{j<n} S_j} \Big\} =: \alpha_n$.

Hence there exists a $d_n \geq r$ such that $\mathbb{P}[\mathcal{P}(d_n)] = \alpha_n$. We choose d_n as such. Now choose a point X_{i_n} uniformly at random among the non-explored points *according to* the process $\{Z_n\}_{n \geq 0}$, i.e.

$$X_{i_n} \overset{d}{=} \mathrm{Unif}\Big(\mathcal{C} \backslash \bigcup_{j < n} B(X_{i_j}, r) \Big). \tag{54}$$

While for the process $\{Z_n\}_{n \geq 0}$ the point X_{i_n} and its neighbors are always added to the set of explored points, for the process $\{L_n\}_{n \geq 0}$ we add X_{i_n}'s neighbors to the set of explored points only if $\mathcal{P}(d_n)$ turned out to be true. I.e.

Case 1: $\mathcal{P}(d_n)$ **is true.** Both X_{i_n} as well as its neighbors will be considered explored. Since $d_n \geq r$, it must be that $\tilde{S}_n = B^*(X_{i_n}, r)$.

Case 2: $\mathcal{P}(d_n)$ **is false.** Only the point X_{i_n} will be considered explored, i.e. $\tilde{S}_n = \phi$.

Consequentially, the exploration process $\{L_n\}_{n \geq 0}$ is updated similarly to the process $\{Z_n\}_{n \geq 0}$, i.e. $L_n = 1 + L_{n-1} + \Phi(\tilde{S}_n)$, but note that now

$$|\tilde{S}_n| = \begin{cases} 0 & \text{with probability } 1 - \alpha_n, \\ v & \text{with probability } \alpha_n. \end{cases} \tag{55}$$

A proof similar to that of Proposition 3.6 implies that $L_n \leq Z_n$ almost surely. We now again consider its scaled version $L_t^N = L_{[tN]}/N$ and analyze its limit behavior as $N \to \infty$.

Proposition 3.8. *For $T > 0$,*

$$\lim_{N \to \infty} \mathbb{E}[\sup_{s \in [0,T]} |L_s^N - l(s)|] = 0, \tag{56}$$

where $l(t)$ is the solution to (44).

Proof. Using similar arguments as in the proof of Proposition 35, we can prove that

$$\dot{l}(t) = 1 + \dot{\eta}_l(t) e^{-\int_0^t \frac{1}{1-l(s)} \, ds} \tag{57}$$

where

$$\dot{\eta}_l(t) = \lim_{N \to \infty} \mathbb{E}[\lambda \tilde{v}_{[tN]}^N] \quad \text{with} \quad \tilde{v}_i^N = \mathbb{E}[|\tilde{S}_i||\mathcal{F}_{i-1}] = v\alpha_i. \tag{58}$$

Then by definition of the exploration process $\{L_n\}_{n \geq 0}$ and equation (50),

$$\dot{\eta}_l(t) = \lambda v \lim_{N \to \infty} \mathbb{P}[\alpha_{[tN]}] = c\Big(1 - \frac{3ct}{1 - l(t)} e^{-\int_0^t \frac{1}{1-l(s)} \, ds}\Big). \tag{59}$$

This concludes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

4 Simulation Results

We now verify our results by simulating the exploration process $\{Z_n\}_{n \geq 0}$ of Section 3.3. We simulate the two-dimensional variant in a box of size $l \times w$ with fixed density $\lambda > 0$, and then choosing for each of the $N = \lambda l w$ particles a position distributed uniformly at random. As part of the initialization, we calculate the distance matrix $D_{i,j} := |\boldsymbol{r}_i - \boldsymbol{r}_j|$ for all particles $i, j \in \{1, \ldots, n\}$, i.e. there are no periodic boundary conditions (the effects of which are negligible for large N), and set $\mathcal{A}_0 = \phi$. We then run the exploration process: At time $n + 1$, we select a non-explored particle $v \in \mathcal{A}_n^c$, identify all of its unexplored neighbors $\mathcal{N}_v = \{w \in \mathcal{A}_n^c | D_{v,w} < r\}$, and update $\mathcal{A}_{n+1} = \mathcal{A}_n \cup \{v\} \cup \mathcal{N}_v$. This is repeated until all particles have been explored at time τ and consequently $\mathcal{A}_\tau^c = \phi$.

On the left in Figure 1 we depict one resulting jammed state at time $t = T_Z^*$. On the right, we show the corresponding scaled process Z_t^N together with our lower and upper bounds. Note that our bounds are tighter for small $t \lesssim 0.2$, and become looser for $t \gtrsim 0.2$. This happens because initially, newly arriving disks do not overlap with already deposited disks. Notice also that our lower bound becomes linear from $t \approx 0.2$ onward. This corresponds to us deleting only one particle at a time in the exploration process, and occurs because this is where our lower bound of Lemma 3.5 becomes invalid. We expect that if we can find a tighter lower bound than Lemma 3.5, for example by using higher-order statistical information on the overlapping area between disks, this point should move to the right so that our lower bound for the scaled process will improve.

In Figure 2, we have plotted our upper bound for the jamming constant T^{upper}, our lower bound T^{lower}, the Erdös–Rényi solution $(\ln{(1 + c)})/c$, as well as an average of the jamming constant T_Z^*/N obtained from 20 samples per value of c, for different values of c. The upper and lower bounds for the jamming constant are obtained by numerically solving the differential equations in (44) and (45), respectively[1]. The dashed line indicates a 99% confidence interval. Here, $N = 1000$.

---

[1] We solved (44) numerically by reformulating it as a system of differential equations. Specifically, we solved $\dot{w}_1(t) = 1 + \max\{0, c(1 - (3ctw_2(t))/(1 - w_1(t)))w_2(t)\}$, and $\dot{w}_2(t) = -w_2(t)/(1 - w_1(t))$ for $w_1(t)$ with initial conditions $w_1(0) = 0$, $w_2(0) = 1$.
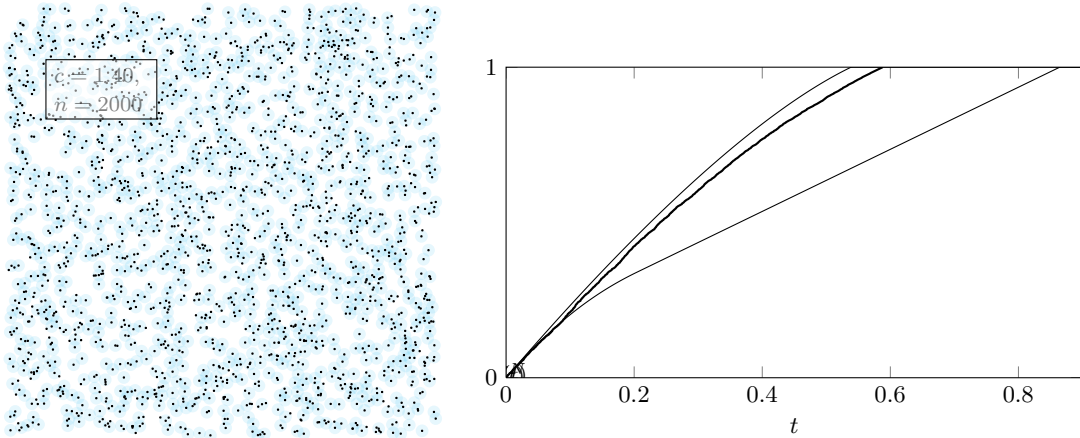
Figure 1: The resulting jammed-state for a two-dimensional simulation of the spatial process (left). The corresponding scaled process Z_t^N as a function of time, together with our lower bound $l(t)$ and upper bound $u(t)$ (right). Here, $c \approx 1.40$ and $N = 2000$.
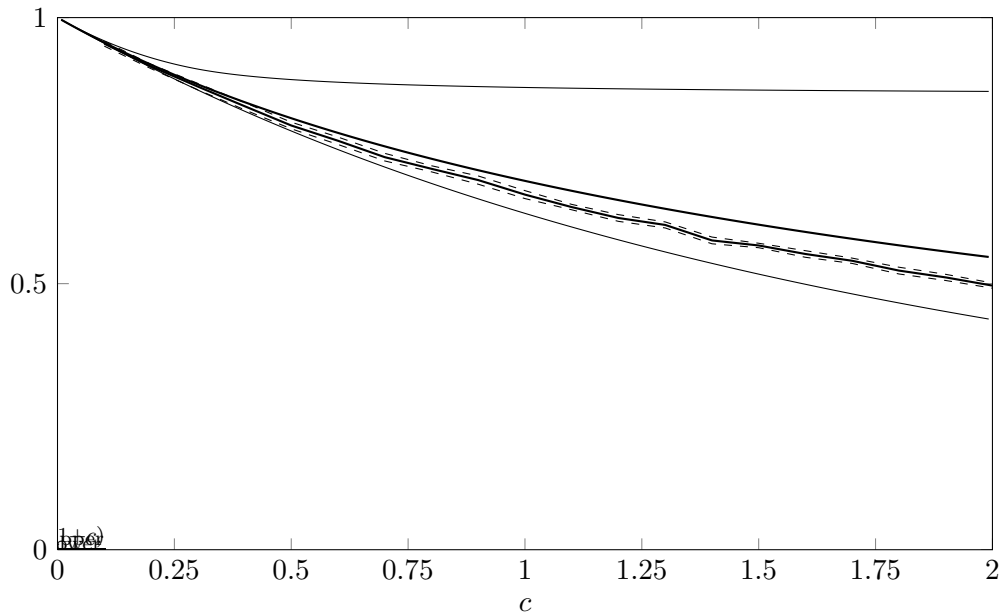

Figure 2: Sample averages of the jamming constant as a function of c, together with our lower and upper bounds, and the Erdös–Rényi solution $(\ln(1 + c))/c$.

Figure 2 illustrates, as Corollary 3.4 has proven, that our lower and upper bounds become tight as $c \downarrow 0$. As c increases, our bounds loosen, because of the increased overlap between deposited disks. The Erdös–Rényi solution is surprisingly close to the actual jamming constant, particularly when we realize there is no geometry associated with the exploration process. The graph suggests that the Erdös–Rényi solution is an upper bound for the jamming constant, tighter than our upper bound. To prove this, it would be sufficient (but not neccesary) to show that $\dot{\eta}/(1 - \eta)$ is upper bounded by c. Our preliminary investigations have proven inconclusive thus far, but we believe the observation to be of interest for future research.

5   Conclusions

We have shown that for sufficiently symmetric random graphs, exploration processes and their jamming limits can be asymptotically described with dynamical systems of lower complexity and with diffusion approximations. For geometric random graphs corresponding to the exploration of a marked Poisson process where this symmetry breaks, we build on fundamental properties of the Poisson process to provide generic bounds for the scaling limits, together with associated trajectorial coupling. These bounds are independent of the shape associated with the exploration and to the dimension of the point process. As future work, we want to investigate tighter bounds by taking into account both dimension and shapes, and we want to link these bounds to the scaling limits of symmetric random graphs.

References

[1] I. Berkes, W. Liu, and W. B. Wu. Komlós–Major–Tusnády approximation under dependence. *Ann. Probab.*, 42(2):794–817, 03 2014.

[2] P. Bermolen, M. Jonckheere, and P.l Moyal. The jamming constant of uniform random graphs. *Stochastic Processes and their Applications*, 2016.

[3] Paola Bermolen, Matthieu Jonckheere, Federico Larroca, and Pascal Moyal. Estimating the transmission probability in wireless networks with configuration models. *ACM Trans. Model. Perform. Eval. Comput. Syst.*, 1(2):9:1–9:23, April 2016.

[4] Paola Bermolen, Matthieu Jonckheere, and Jaron Sanders. Scaling limits for exploration algorithms. Technical report, 2015.

[5] R.W.R. Darling and J.R. Norris. Differential equation approximations for Markov chains. *Probability Surveys*, 5:37–79, 2008.

[6] S. Dhara, J.S.H. van Leeuwaarden, and D. Mukherjee. Solvable random network model for disordered sphere packing. *arXiv preprint arXiv:1611.05019*, 2016.

[7] J. W. Evans. Random and cooperative sequential adsorption. *Rev. Mod. Phys.*, 65:1281–1329, 10 1993.

[8] Thomas F. Gallagher. *Rydberg Atoms (Cambridge Monographs on Atomic, Molecular and Chemical Physics)*. Cambridge University Press, 1994.

[9] J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent rv's-s, and the sample df. i. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 32(1-2):111–131, 1975.

[10] J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent rv's, and the sample df. ii. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 34(1):33–58, 1976.

[11] T. G. Kurtz. Strong approximation theorems for density dependent Markov chains. *Stochastic Processes and their Applications*, 6(3):223 – 240, 1978.

[12] C. McDiarmid. Colouring random graphs. *Annals of Operations Research*, 1(3):183–200, 1984.

[13] Mathew Penrose. *Random geometric graphs.* Number 5. Oxford University Press, 2003.

[14] Mathew D. Penrose. Random parking, sequential adsorption, and the jamming limit. *Commun. Math. Phys.*, 218(1):153–176, 2001.

[15] Mathew D. Penrose and J.E. Yukich. Limit theory for random sequential packing and deposition. *Ann. Appl. Probab.*, 12(1):272–301, 02 2002.

[16] Walter Rudin. *Real and Complex Analysis.* Tata McGraw-Hill, 1987.

[17] Jaron Sanders. *Stochastic optimization of large-scale complex systems.* PhD thesis, Technische Universiteit Eindhoven, 2016.

[18] Jaron Sanders, Matthieu Jonckheere, and Servaas Kokkelmans. Sub-Poissonian statistics of jamming limits in ultracold Rydberg gases. *Phys. Rev. Lett.*, 115:043002, 7 2015.

[19] J. Michael Steele. *Stochastic Calculus and Financial Applications.* Springer, 2001.

[20] K. Teerapabolaan. A bound on the binomial–Poisson relative error. *International Journal of Pure and Applied Mathematics*, 87(4):535 – 540, 2013.

Appendix

A   Proof of Proposition 2.2

Proof. Doob's martingale decomposition [19] for the Markov process $\{Z_n\}_{n \geq 0}$ gives that for $n \geq 0$,

$$Z_n = \sum_{i=0}^{n} (1 + \gamma_N(Z_i)) + M_n. \tag{60}$$

Here, we have used that $Z_0 = 0$, and M_n denotes a local martingale that is actually a global martingale since the state space is finite.

We will now examine the scaled random variable Z_t^N, for which

$$Z_t^N = \frac{Z_{[tN]}}{N} = \frac{1}{N} \sum_{i=0}^{[tN]} \big(1 + \gamma_N(Z_i)\big) + \frac{M_{[tN]}}{N} \tag{61}$$

$$\overset{(i)}{=} \frac{1}{N} \int_0^{[tN]} \big(1 + \gamma_N(Z_s)\big)\, ds + \frac{M_{[tN]}}{N} \overset{(ii)}{=} \int_0^{\frac{[tN]}{N}} \big(1 + \gamma_N(Z_{uN})\big)\, du + M_t^N,$$

since we (i) view each trajectory as being path-wise continuous, and (ii) use the change of variables $u = s/N$, and introduce the notation $M_t^N = M_{[tN]}/N$ for a scaled martingale.

We can replace the integral $\int_0^{[tN]/N} \cdots du$ by the integral $\int_0^t \cdots du$, which introduces an error $\Delta_{N,t}$. Specifically, we can write

$$\int_0^{\frac{[tN]}{N}} \big(1 + \gamma_N(Z_{uN})\big)\, du = \int_0^t \big(1 + \gamma_N(Z_{uN})\big)\, du + \Delta_{N,t} \tag{62}$$

where

$$\Delta_{N,t} = \int_0^{\frac{[tN]}{N}} \big(1 + \gamma_N(Z_{uN})\big)\, du - \int_0^t \big(1 + \gamma_N(Z_{uN})\big)\, du. \tag{63}$$

For large N such replacement has negligible impact, since independently of t,

$$|\Delta_{N,t}| \leq \sup_{u \in [0,1]} \{1 + \gamma_N(Z_{uN})\} \Big| \frac{[tN]}{N} - t \Big| \leq \frac{1 + \bar{\gamma}_N}{N}, \tag{64}$$

where in the last inequality we have used that $\bar{\gamma}_N = \sup_x \gamma_N(x)$ and $|[tN] - tN| \leq 1$.

Using (i) the integral version of (9), the triangle inequality [16], and (ii) Lipschitz continuity of γ, condition (7) and bound (64), we find that

$$\sup_{s \in [0,t]} |Z_s^N - z(s)| \overset{(i)}{\leq} \sup_{s \in [0,t]} \Big(\int_0^s \big| \gamma_N(Z_{uN}) - \gamma(z(u)) \big|\, du + |\Delta_{N,s}| + |M_s^N| \Big) \tag{65}$$

$$\overset{(ii)}{\leq} C_L \int_0^t \sup_{u \in [0,s]} |Z_u^N - z(u)|\, ds + \delta_N t + \frac{1 + \bar{\gamma}_N}{N} + \sup_{s \in [0,t]} |M_s^N|.$$

Next, we define $\epsilon_N(T) = \sup_{s \in [0,T]} |Z_s^N - z(s)|$ for notational convenience and to prepare for an application of Grönwall's lemma [19]. Eq. (65) then shortens for $T > 0$ to

$$\epsilon_N(T) \leq \delta_N T + \frac{1 + \bar{\gamma}_N}{N} + \sup_{s \in [0,T]} |M_s^N| + C_L \int_0^T \epsilon_N(s)\, ds. \tag{66}$$

Because $\delta_N T + (1 + \bar{\gamma}_N)/N + \sup_{s \in [0,T]} |M_s^N|$ is nondecreasing in T, it follows from Grönwall's lemma that

$$\epsilon_N(T) \leq \Big(\delta_N T + \frac{1 + \bar{\gamma}_N}{N} + \sup_{s \in [0,T]} |M_s^N| \Big) e^{C_L T}. \tag{67}$$

Using Minkowsky's inequality for $p \in [1, \infty)$, strict monotonicity of $\exp(C_L T)$ and $\delta_N T$, and the triangle inequality, we find that

$$\|\epsilon_N(T)\|_p \leq \Big(\delta_N T + \frac{1 + \bar{\gamma}_N}{N} + \| \sup_{s \in [0,T]} |M_s^N| \|_p \Big) e^{C_L T}. \tag{68}$$

Finally, using Doob's martingale inequality [19] for $p > 1$, we obtain

$$\|\epsilon_N(T)\|_p \leq \Big(\delta_N T + \frac{1 + \bar{\gamma}_N}{N} + \kappa_p \|M_T^N\|_p \Big) \mathrm{e}^{C_L T}, \tag{69}$$

completing the first part of the proof.

For $p = 2$, this inequality can be further simplified by computing the increasing process associated to the martingale. Note specifically that for $l \geq 0$ we have

$$\mathbb{E}[(M_l)^2] = \mathbb{E}[\langle M_l \rangle] = \mathbb{E}\Big[\sum_{i=0}^{l} \mathrm{Var}[\gamma_N(Z_i)] \Big] \tag{70}$$

where

$$\mathrm{Var}[\gamma_N(x)] = \sum_{k=0}^{N-x-1} (k+1)^2 p_{x,x+k+1} - \Big(\sum_{k=0}^{N-x-1} (k+1) p_{x,x+k+1} \Big)^2 = \psi_N(x). \tag{71}$$

Therefore for the scaled martingale M_t^N, we find by combining (70) and (71) that for $t > 0$

$$\|M_t^N\|_2^2 = \mathbb{E}[(M_t^N)^2] = \frac{\mathbb{E}[M_{[tN]}^2]}{N^2} = \frac{1}{N^2} \sum_{i=0}^{[tN]} \psi_N(Z_i) \leq \frac{\bar{\psi}_N t}{N}. \tag{72}$$

This completes the second part of the proof. □

B   Proof of Proposition 2.4

Proof. We adapt the results of Kurtz which were derived for continuous time Markov jump processes. For doing so, we can replace the Poisson processes involved in the construction of the jump processes by some random walks that can be used to construct discrete time Markov chains. We can then use exactly the same steps as in [11], by first comparing the original process Z^N to a diffusion of the form

$$\tilde{Z}_t^N = \frac{1}{N} \sum_{l \leq N} l B_l(N \sum_{0}^{t} p_N(l, \tilde{Z}_s^N) \, \mathrm{d}s), \tag{73}$$

that is a sum of a finite number of scaled independent Brownian motions B_l.

Rewriting the inequalities in [11, (3.6)], and using a random walk version of the approximation lemma of Komlós–Major–Tusnády [1], we obtain

$$\mathbb{E}\Big[\sup_{t \leq T} |\tilde{Z}_t^N - Z_t^N| \Big] \leq C_2 \frac{\log(N)}{N}. \tag{74}$$

This leads using the results of [11, Section 3] to

$$\mathbb{E}\Big[\sup_{t \leq T} |W_t^N - W_t| \Big] \leq C_3 \frac{\log(N)}{\sqrt{N}}, \tag{75}$$

which concludes the proof. □

C Proof of Proposition 2.5

Proof. Remark that if $|z(s) - Z_s^N| \leq \delta/2$ for all $s > 0$, that then

$$\left| \frac{T_N^*}{N} - T^* \right| \leq \left| (z - \tfrac{1}{2}\delta)^{-1}(1) - (z + \tfrac{1}{2}\delta)^{-1}(1) \right| \leq |(T^* + \tfrac{1}{2}\delta) - (T^* - \tfrac{1}{2}\delta)| = \delta. \tag{76}$$

Here, the last inequality follows from the fact that $\dot{z}(s) = 1 + \gamma(z(s)) \geq 1$, since

$$(z - \tfrac{1}{2}\delta)(T^* + \tfrac{1}{2}\delta) = (z - \tfrac{1}{2}\delta)(z^{-1}(1) + \tfrac{1}{2}\delta)$$
$$\geq z(z^{-1}(1)) + \tfrac{1}{2}\delta - \tfrac{1}{2}\delta = 1 = (z - \tfrac{1}{2}\delta)((z - \tfrac{1}{2}\delta)^{-1}(1)). \tag{77}$$

Thus the first claim follows directly from the observation that the event

$$\left\{ \left| \frac{T_N^*}{N} - T^* \right| \geq \delta \right\} \subseteq \left\{ |z(s) - Z_s^N| \geq \tfrac{1}{2}\delta \right\}, \tag{78}$$

and then using (i) Markov's inequality [19], and (ii) invoking Proposition 2.2, so that

$$\mathbb{P}\left[\left| \frac{T_N^*}{N} - T^* \right| \geq \delta \right] \overset{(78)}{\leq} \mathbb{P}\left[|z(s) - Z_s^N| \geq \tfrac{1}{2}\delta \right] \overset{(i)}{\leq} \frac{2}{\delta}\mathbb{E}[|z(s) - Z_s^N|] \overset{(ii)}{\leq} \frac{2\omega_N}{\delta}. \tag{79}$$

Now (i) using that $Z_{T_N^*}/N = z(T^*) = 1$ together with (9) and (61), and (ii) after expanding the integrals, we find that

$$\frac{T_N^*}{N} - T^* \overset{(i)}{=} \int_0^{T^*} \gamma(z(s))\,\mathrm{d}s - \int_0^{\frac{T_N^*}{N}} \gamma_N(Z_{sN})\,\mathrm{d}s - \frac{M_{T_N^*}}{N}$$
$$\overset{(ii)}{=} \int_0^{\frac{T_N^*}{N} \wedge T^*} (\gamma(z(s)) - \gamma_N(Z_{sN}))\,\mathrm{d}s - \frac{M_{T_N^*}}{N}$$
$$+ \int_{\frac{T_N^*}{N} \wedge T^*}^{T^*} \gamma(z(s))\,\mathrm{d}s - \int_{\frac{T_N^*}{N} \wedge T^*}^{\frac{T_N^*}{N}} \gamma_N(Z_{sN})\,\mathrm{d}s. \tag{80}$$

Then taking the absolute value and using the triangle inequality, it follows that

$$\left| \frac{T_N^*}{N} - T^* \right| \leq \int_0^{\frac{T_N^*}{N} \wedge T^*} |\gamma(z(s)) - \gamma_N(Z_{sN})|\,\mathrm{d}s + |M_{T_N^*/N}^N|$$
$$+ \int_{\frac{T_N^*}{N} \wedge T^*}^{T^*} |\gamma(z(s))|\,\mathrm{d}s + \int_{\frac{T_N^*}{N} \wedge T^*}^{\frac{T_N^*}{N}} |\gamma_N(Z_{sN})|\,\mathrm{d}s. \tag{81}$$

Approximating γ_N by γ via (7), using Lipschitz continuity of γ, and recalling that $\max\{T_N^*/N, T^*\} \leq 1$, we find that

$$\left| \frac{T_N^*}{N} - T^* \right| \leq 2C_L \sup_{s \leq 1} |z(s) - Z_s^N| + 2\delta_N + |M_{T_N^*/N}^N| + \int_{\frac{T_N^*}{N} \wedge T^*}^{\frac{T_N^*}{N} \vee T^*} |\gamma(z(s))|\,\mathrm{d}s. \tag{82}$$

The continuity of $\gamma(x)$ guarantees that there exist constants $C_1, \varepsilon > 0$ such that (i) $\gamma(z(s)) \leq 1 - \varepsilon$ for all $s \geq C_1$, and (ii) $C_1 < T^* - \delta$, provided that δ is sufficiently small. There

are now two possible cases: either (a) $C_1 \leq T_N^*/N \wedge T^*$, or (b) $T_N^*/N \wedge T^* < C_1 < T_N^*/N \vee T^*$. For convenience, we first split the integral according to

$$\int_{\frac{T_N^*}{N} \wedge T^*}^{\frac{T_N^*}{N} \vee T^*} |\gamma(z(s))| \, \mathrm{d}s = \int_{\frac{T_N^*}{N} \wedge T^*}^{\frac{T_N^*}{N} \vee T^*} |\gamma(z(s))| (\mathbb{1}[s < C_1] + \mathbb{1}[s \geq C_1]) \, \mathrm{d}s. \tag{83}$$

Then splitting further into case (a), we have that

$$\int_{\frac{T_N^*}{N} \wedge T^*}^{\frac{T_N^*}{N} \vee T^*} |\gamma(z(s))| \mathbb{1}\left[s < C_1, C_1 \leq \frac{T_N^*}{N} \wedge T^*\right] \mathrm{d}s = 0, \tag{84}$$

and

$$\int_{\frac{T_N^*}{N} \wedge T^*}^{\frac{T_N^*}{N} \vee T^*} |\gamma(z(s))| \mathbb{1}\left[s \geq C_1, C_1 \leq \frac{T_N^*}{N} \wedge T^*\right] \mathrm{d}s \tag{85}$$

$$\leq (1 - \varepsilon) \left|\frac{T_N^*}{N} - T^*\right| \mathbb{1}\left[C_1 \leq \frac{T_N^*}{N} \wedge T^*\right].$$

Next let C_2 be a constant such that $C_2 \geq \int_{T_N^*/N \wedge T^*}^{C_1} |\gamma(z(s))| \, \mathrm{d}s$. We can then, after splitting further into case (b), bound

$$\int_{\frac{T_N^*}{N} \wedge T^*}^{\frac{T_N^*}{N} \vee T^*} |\gamma(z(s))| \mathbb{1}\left[s < C_1, \frac{T_N^*}{N} \wedge T^* < C_1 < \frac{T_N^*}{N} \vee T^*\right] \mathrm{d}s \tag{86}$$

$$= \int_{\frac{T_N^*}{N} \wedge T^*}^{C_1} |\gamma(z(s))| \, \mathrm{d}s \mathbb{1}\left[\frac{T_N^*}{N} \wedge T^* < C_1 < \frac{T_N^*}{N} \vee T^*\right]$$

$$\leq C_2 \mathbb{1}\left[\frac{T_N^*}{N} \wedge T^* < C_1 < \frac{T_N^*}{N} \vee T^*\right] \leq C_2 \mathbb{1}\left[\frac{T_N^*}{N} < C_1\right],$$

since if $\mathbb{1}[T_N^*/N \wedge T^* < C_1 < T_N^*/N \vee T^*] = 1$, clearly $T_N^*/N \wedge T^* < C_1$. But by construction $C_1 < T^*$, so it must hold that $T_N^*/N < C_1$ and thus $\mathbb{1}[T_N^*/N < C_1] = 1$. Next, we bound

$$\int_{\frac{T_N^*}{N} \wedge T^*}^{\frac{T_N^*}{N} \vee T^*} |\gamma(z(s))| \mathbb{1}\left[s \geq C_1, \frac{T_N^*}{N} \wedge T^* < C_1 < \frac{T_N^*}{N} \vee T^*\right] \mathrm{d}s \tag{87}$$

$$\leq (1 - \varepsilon) \left|\frac{T_N^*}{N} - T^*\right| \mathbb{1}\left[\frac{T_N^*}{N} \wedge T^* < C_1 < \frac{T_N^*}{N} \vee T^*\right].$$

Summarizing, there thus exists a constant C_2 such that

$$\left|\frac{T_N^*}{N} - T^*\right| \leq 2C_L \sup_{s \leq 1} |z(s) - Z_s^N| + 2\delta_N + |M_{T_N^*/N}^N| \tag{88}$$

$$+ (1 - \varepsilon) \left|\frac{T_N^*}{N} - T^*\right| + C_2 \mathbb{1}\left[\frac{T_N^*}{N} < C_1\right].$$

Now recall that if $|z(s) - Z_s^N| \leq \delta/2$, then $|T_N^*/N - T^*| \leq \delta$. Moreover then also $T_N^*/N \geq C_1$ since $C_1 < T^* - \delta$. Hence,

$$\left\{\frac{T_N^*}{N} < C_1\right\} \subset \left\{|z(s) - Z_s^N| > \tfrac{1}{2}\delta\right\}. \tag{89}$$

20

Then by (i) collecting terms in and subsequently using (88), and then (ii) applying Minkowski's inequality [16], we obtain

$$\varepsilon\Big\|\frac{T_N^*}{N} - T^*\Big\|_2 \overset{\text{(i)}}{\leq} \Big\|2C_L \sup_{s\leq 1}|z(s) - Z_s^N| + 2\delta_N + |M_{T_N^*/N}^N| + C_2\mathbb{1}\Big[\frac{T_N^*}{N} < C_1\Big]\Big\|_2$$

$$\overset{\text{(ii)}}{\leq} 2C_L\|\sup_{s\leq 1}|z(s) - Z_s^N|\|_2 + 2\delta_N + \|M_{T_N^*/N}^N\|_2 + C_2\Big\|\mathbb{1}\Big[\frac{T_N^*}{N} < C_1\Big]\Big\|_2. \tag{90}$$

We now note that (iii) since $f(y) = y^2$ is monotonically increasing for $y \geq 0$ and (iv) by Markov's inequality,

$$\Big\|\mathbb{1}\Big[\frac{T_N^*}{N} < C_1\Big]\Big\|_2 = \mathbb{P}\Big[\frac{T_N^*}{N} < C_1\Big]^{\frac{1}{2}} \overset{\text{(89)}}{\leq} \mathbb{P}[|z(s) - Z_s^N| > \tfrac{1}{2}\delta]^{\frac{1}{2}} \tag{91}$$

$$\overset{\text{(iii)}}{=} \mathbb{P}[|z(s) - Z_s^N|^2 > \tfrac{1}{4}\delta^2]^{\frac{1}{2}} \overset{\text{(iv)}}{\leq} \frac{2}{\delta}\mathbb{E}[|z(s) - Z_s^N|^2]^{\frac{1}{2}} = \frac{2}{\delta}\|z(s) - Z_s^N\|_2.$$

Therefore,

$$\varepsilon\Big\|\frac{T_N^*}{N} - T^*\Big\|_2 \leq \big(2C_L + \frac{2C_2}{\delta}\big)\|\sup_{s\leq 1}|z(s) - Z_s^N|\|_2 + 2\delta_N + \|M_{T_N^*/N}^N\|_2. \tag{92}$$

Thus by finally using Proposition 2.2 and (72), we have that there exist constants C_3, C_4 so that

$$\varepsilon\Big\|\frac{T_N^*}{N} - T^*\Big\|_2 \leq C_3\omega_N + 2\delta_N + \sqrt{\frac{\bar{\psi}_N}{N}} \leq C_4\omega_N, \tag{93}$$

which concludes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

D   Proof of Proposition 2.7

Proof. First, recall that by (9) and (61), see (80),

$$\frac{T_N^*}{N} - T^* = \int_0^{T^*} \gamma(z(s))\,\mathrm{d}s - \int_0^{\frac{T_N^*}{N}} \gamma_N(Z_{sN})\,\mathrm{d}s - M_{T_N^*/N}^N, \tag{94}$$

Note furthermore that

$$W_{T^*}^N \overset{\text{(13)}}{=} \sqrt{N}\big(Z_{T^*}^N - z(T^*)\big) \overset{\text{(61)}}{=} \sqrt{N}\Big(\int_0^{\frac{[T^*N]}{N}} (1 + \gamma_N(Z_{sN}))\,\mathrm{d}s + M_{T^*}^N - z(T^*)\Big)$$

$$= \sqrt{N}\Big(\int_0^{T^*} (1 + \gamma_N(Z_{sN}))\,\mathrm{d}s + M_{T^*}^N - z(T^*)\Big) + \sqrt{N}\Delta_{N,T^*} \tag{95}$$

where $M_{T^*}^N = M_{[T^*N]}/N$. Recall that the error Δ_{N,T^*} introduced by replacing the upper integration boundary, can readily be bounded by $|\Delta_{N,T^*}| \leq (1 + \bar{\gamma}_N)/N$, see (64).

Comparing (94) and (95), a subsequent natural series of steps would be to (i) add comparison terms $\pm W_{T^*}^N$ and use the triangle inequality, and then (ii) substitute (95), use the

triangle inequality, and upper bound $|\Delta_{N,T^*}| \le (1 + \bar{\gamma}_N)/N$, after which we arrive at

$$\left| \sqrt{N} \Big(\frac{T_N^*}{N} - T^* \Big) + W_{T^*} \right|$$

$$\overset{(i)}{\le} \left| \sqrt{N} \Big(\int_0^{T^*} \gamma(z(s)) \, \mathrm{d}s - \int_0^{\frac{T_N^*}{N}} \gamma_N(Z_{sN}) \, \mathrm{d}s - M_{T_N^*/N}^N \Big) + W_{T^*}^N \right| + |W_{T^*} - W_{T^*}^N|$$

$$\overset{(ii)}{\le} \left| \sqrt{N} \Big(\int_0^{T^*} \gamma(z(s)) \, \mathrm{d}s - \int_0^{\frac{T_N^*}{N}} \gamma_N(Z_{sN}) \, \mathrm{d}s - M_{T_N^*/N}^N \Big) \right.$$

$$\left. + \sqrt{N} \Big(\int_0^{T^*} (1 + \gamma_N(Z_{sN})) \, \mathrm{d}s + M_{T^*}^N - z(T^*) \Big) \right| + |W_{T^*} - W_{T^*}^N| + \frac{1 + \bar{\gamma}_N}{\sqrt{N}}$$

$$= \ \text{term I} \ + \ \text{term II} \ + \frac{1 + \bar{\gamma}_N}{\sqrt{N}}. \tag{96}$$

We will now proceed and bound term I and II.

Note that the expectation of term II can be directly bounded by Proposition 2.4, i.e. there exists a constant C_2 such that

$$\mathbb{E}[\text{term II}] = \mathbb{E}[|W_{T^*} - W_{T^*}^N|] \le \mathbb{E}[\sup_{t \le 1} |W_t - W_t^N|] \le C_2 \frac{\log(N)}{\sqrt{N}}. \tag{97}$$

Bounding term I requires more work. Using (95), the integral version of $\dot{z} = 1 + \gamma(z)$, and the triangle inequality, we find that

$$\text{term I} \le \sqrt{N} \left| \int_0^{T^*} \gamma_N(Z_{sN}) \, \mathrm{d}s - \int_0^{\frac{T_N^*}{N}} \gamma_N(Z_{sN}) \, \mathrm{d}s \right| + \sqrt{N} |M_{T^*}^N - M_{T_N^*/N}^N|$$

$$= \text{term Ia} + \text{term Ib}, \tag{98}$$

and we now proceed with bounding term Ia and Ib separately.

In order to bound term Ia, we add comparison terms $\pm\gamma(Z_{sN}/N)$ and $\pm\gamma(z(s))$ and use the triangle inequality, so that[2]

$$\text{term Ia} \le \sqrt{N} \left| \int_{\frac{T_N^*}{N}}^{T^*} \gamma_N(Z_{sN}) - \gamma\Big(\frac{Z_{sN}}{N} \Big) \, \mathrm{d}s \right| + \sqrt{N} \left| \int_{\frac{T_N^*}{N}}^{T^*} \gamma\Big(\frac{Z_{sN}}{N} \Big) - \gamma(z(s)) \, \mathrm{d}s \right|$$

$$+ \sqrt{N} \left| \int_{\frac{T_N^*}{N}}^{T^*} \gamma(z(s)) \, \mathrm{d}s \right| \tag{99}$$

Then by approximating γ_N by γ, using the Lipschitz continuity of γ, and upper bounding the first two integrands, we find that

$$\text{term Ia} \le \sqrt{N} \delta_N \left| \frac{T_N^*}{N} - T^* \right| + \sqrt{N} C_L \sup_{s \le 1} \left| \frac{Z_{sN}}{N} - z(s) \right| \left| \frac{T_N^*}{N} - T^* \right|$$

$$+ \sqrt{N} \left| \int_{\frac{T_N^*}{N}}^{T^*} \gamma(z(s)) \, \mathrm{d}s \right| \tag{100}$$

---

[2]Note that we use the notation that $\int_a^b = -\int_b^a$ when $a > b$.

Taking the expectation and using the triangle inequality, it follows that

$$\mathbb{E}[\text{term Ia}] \leq \sqrt{N}\delta_N \mathbb{E}\Big[\Big|\frac{T_N^*}{N} - T^*\Big|\Big] + \sqrt{N}C_L \mathbb{E}\Big[\sup_{s \leq 1}\Big|\frac{Z_{sN}}{N} - z(s)\Big|\Big|\frac{T_N^*}{N} - T^*\Big|\Big]$$
$$+ \sqrt{N}\mathbb{E}\Big[\int_{\frac{T_N^*}{N}}^{T^*} |\gamma(z(s))|\, \mathrm{d}s\Big]. \tag{101}$$

Applying Hölder's inequality [16],

$$\mathbb{E}[\text{term Ia}] \leq \sqrt{N}\delta_N \Big\|\frac{T_N^*}{N} - T^*\Big\|_2 + \sqrt{N}C_L \Big\|\sup_{s \leq 1}\Big|\frac{Z_{sN}}{N} - z(s)\Big|\Big\|_2 \Big\|\frac{T_N^*}{N} - T^*\Big\|_2$$
$$+ \sqrt{N}\mathbb{E}\Big[\int_{\frac{T_N^*}{N}}^{T^*} |\gamma(z(s))|\, \mathrm{d}s\Big], \tag{102}$$

and finally Propositions 2.2 and 2.5, we end up with

$$\mathbb{E}[\text{term Ia}] \leq \Omega_N\sqrt{N}(\delta_N + C_L\omega_N) + \sqrt{N}\mathbb{E}\Big[\int_{\frac{T_N^*}{N}}^{T^*} |\gamma(z(s))|\, \mathrm{d}s\Big]. \tag{103}$$

In order to deal with the last term in (103), we will use a Taylor expansion of order zero around T^*. Specifically, we write

$$\gamma(z(s)) = \gamma(z(T^*)) + c(s - T^*) + R_2 = c(s - T^*) + R_2, \tag{104}$$

where we have recalled that $\gamma(1) = 0$ by assumption and $z(T^*) = 1$. Then, by (i) the triangle inequality, (ii) upper bounding the integrand, and (iii) evaluating the integral, there exists a constant C_3 so that

$$\sqrt{N}\int_{\frac{T_N^*}{N}}^{T^*} |\gamma(z(s))|\, \mathrm{d}s \overset{(i)}{\leq} \sqrt{N}\int_{\frac{T_N^*}{N}}^{T^*} c|s - T^*| + |R_2|\, \mathrm{d}s$$
$$\overset{(ii)}{\leq} \sqrt{N}\int_{\frac{T_N^*}{N}}^{T^*} c\Big|\frac{T_N^*}{N} - T^*\Big| + |R_2|\, \mathrm{d}s \overset{(iii)}{\leq} C_3\sqrt{N}\Big|\frac{T_N^*}{N} - T^*\Big|^2, \tag{105}$$

where for the second term we have used that $|R_2| = O((s - T^*)^2)$, and that $|s - T^*| \leq 1$ for $s \in [T_N^*/N, T^*]$. Therefore, by Proposition 2.5,

$$\sqrt{N}\mathbb{E}\Big[\int_{\frac{T_N^*}{N}}^{T^*} |\gamma(z(s))|\, \mathrm{d}s\Big] \leq C_3\sqrt{N}\mathbb{E}\Big[\Big|\frac{T_N^*}{N} - T^*\Big|^2\Big]$$
$$\leq C_3\sqrt{N}\Big\|\frac{T_N^*}{N} - T^*\Big\|_2^2 \leq C_3\Omega_N^2\sqrt{N}. \tag{106}$$

Ultimately bounding (103) using (106), we conclude that there exists a constant C_1 such that

$$\mathbb{E}[\text{term Ia}] \leq \Omega_N\sqrt{N}(\delta_N + C_L\omega_N + C_3\Omega_N) \leq C_1\omega_N^2\sqrt{N}. \tag{107}$$

To finish the proof we still need to bound the expectation of term Ib, that is, $\sqrt{N}\mathbb{E}[|M_{T^*}^N - M_{T_N^*/N}^N|]$. By (i) Cauchy–Schwarz's inequality, (ii) definition of the scaled martingale, (iii)

23

calculating the increasing process similar to (70)–(72), and (iv) $\mathbb{E}[(M_t - M_s)^2] = \mathbb{E}[M_t^2] - \mathbb{E}[M_s^2]$ for $t > s$ as a consequence of M_t being a martingale, we find

$$\sqrt{N}\mathbb{E}[|M_{T_N^*/N}^N - M_{T^*}^N|] \overset{(i)}{\le} \sqrt{N}\mathbb{E}[|M_{T_N^*/N}^N - M_{T^*}^N|^2]^{\frac{1}{2}} \overset{(ii)}{=} \frac{1}{\sqrt{N}}\mathbb{E}[|M_{T_N^*} - M_{[T^*N]}|^2]^{\frac{1}{2}}$$

$$\overset{(iii)}{=} \frac{1}{\sqrt{N}}\mathbb{E}[< M_{T_N^*} - M_{[T^*N]} >]^{\frac{1}{2}} \overset{(iv)}{=} \mathbb{E}\Big[\frac{1}{N}\sum_{i=T_N^*\wedge[T^*N]}^{T_N^*\vee[T^*N]}\psi_N(Z_i)\Big]^{\frac{1}{2}}. \tag{108}$$

Then (v) upper bounding $\psi_N(Z_i) \le \bar{\psi}_N$, (vi) adding compensation terms $\pm T^*$, applying the triangle inequality and upper bounding $|T^*N - [T^*N]| \le 1$, it follows (vii) from Proposition 2.5 that

$$\mathbb{E}[\text{term Ib}] = \sqrt{N}\mathbb{E}[|M_{T_N^*/N}^N - M_{T^*}^N|] \overset{(v)}{\le} \mathbb{E}\Big[\bar{\psi}_N\Big|\frac{T_N^*}{N} - \frac{[T^*N]}{N}\Big|\Big]^{\frac{1}{2}}$$

$$\overset{(vi)}{\le} \Big(\bar{\psi}_N\Big\|\frac{T_N^*}{N} - T^*\Big\|_1 + \frac{\bar{\psi}_N}{N}\Big)^{\frac{1}{2}} \overset{(vii)}{\le} \Big(\bar{\psi}_N\Omega_N + \frac{\bar{\psi}_N}{N}\Big)^{\frac{1}{2}}. \tag{109}$$

Finally, we combine all bounds, resulting in

$$\mathbb{E}\Big[\Big|\sqrt{N}\Big(\frac{T_N^*}{N} - T^*\Big) + W_{T^*}\Big|\Big] \le \mathbb{E}[\text{term Ia}] + \mathbb{E}[\text{term Ib}] + \mathbb{E}[\text{term II}] + \frac{1 + \bar{\gamma}_N}{\sqrt{N}}$$

$$\le C_1\omega_N^2\sqrt{N} + \Big(\bar{\psi}_N\Omega_N + \frac{\bar{\psi}_N}{N}\Big)^{\frac{1}{2}} + C_2\frac{\log(N)}{\sqrt{N}} + \frac{1 + \bar{\gamma}_N}{\sqrt{N}}. \tag{110}$$

If the distribution of the number of neighbors is such that $\delta_N = o(1/\sqrt{N})$, $\bar{\gamma}_N = o(\sqrt{N})$ and $\bar{\psi}_N = o(N^{1/4})$, then $\omega_N = o(1/N^{3/8})$ and $\Omega_N = o(1/N^{3/8})$, and all the product-terms in (110) converge to 0 as $N \to \infty$. We have thus proven that under these conditions, the limit is a Gaussian random variable with variance

$$\sigma^2 = \mathbb{E}[W_{T^*}^2]. \tag{111}$$

Defining $m(t) = \mathbb{E}[W_t^2]$, and using Itô's formula [19], note that

$$\mathbb{E}[W_t^2] = \mathbb{E}\Big[\int_0^t 2W_s dW_s + \frac{1}{2}2\beta_t\Big] = 2\int_0^t \gamma'(z(s))\mathbb{E}[W_s^2]\,\mathrm{d}s + \beta(t), \tag{112}$$

and hence $m(t)$ satisfies the differential system

$$\dot{m} = -2\dot{\gamma}(z(t))m(t) + \dot{\beta}, \quad \text{with} \quad m_0 = 0. \tag{113}$$

This finishes the proof. □

E   Proof of Corollary 3.4

Proof. Consider the expansion $u(t) = t + \sum_{i=1}^{\infty} c^i u_i(t)$. Substitute into (45), and Taylor expand the right-hand side to obtain

$$1 + cu_1'(t) + O(c^2) = 1 + c \exp\left(-\int_0^t \frac{ds}{1 - u(s)}\right)$$

$$= 1 + c \exp\left(-\int_0^t \frac{1}{1-s} + \sum_{i=1}^{\infty} \frac{c^i u_i(s)}{(1-s)^{i+1}} \, ds\right)$$

$$= 1 + c(1-t)\exp\left(\sum_{i=1}^{\infty} \frac{c^i u_i(s)}{(1-s)^{i+1}} \, ds\right) = 1 + c(1-t)\big(1 + O(c)\big). \quad (114)$$

Comparing terms we find that $u_1'(t) = (1-t)$ with initial condition $u_1(0) = 0$, leading to the conclusion that $u_1(t) = t(1 - \frac{1}{2}t)$. Therefore,

$$u(t) = t + ct(1 - \tfrac{1}{2}t) + O(c^2) = (1+c)t - \tfrac{1}{2}ct^2 + O(c^2). \quad (115)$$

Exactly the same expansion is obtained for $l(t)$ when applying the approach up to and including order $O(c)$. Since $u(t)$ is an upper bound and $l(t)$ is a lower bound for the fluid limit $z(t)$ of the spatial process, and both bounds have the same asymptotic behavior as $c \downarrow 0$, this completes the proof. $\square$