

# COMPONENT SIZES FOR LARGE QUANTUM ERDŐS–RÉNYI GRAPH NEAR CRITICALITY

BY AMIR DEMBO<sup>1</sup>, ANNA LEVIT AND SREEKAR VADLAMANI

*Stanford University, University of British Columbia and TIFR-CAM*

The  $N$  vertices of a quantum random graph are each a circle independently punctured at Poisson points of arrivals, with parallel connections derived through for each pair of these punctured circles by yet another independent Poisson process. Considering these graphs at their critical parameters, we show that the joint law of the rescaled by  $N^{2/3}$  and ordered sizes of their connected components, converges to that of the ordered lengths of excursions above zero for a reflected Brownian motion with drift. Thereby, this work forms the first example of an inhomogeneous random graph, beyond the case of effectively rank-1 models, which is rigorously shown to be in the Erdős–Rényi graphs universality class in terms of Aldous’s results.

**1. Introduction.** The Erdős–Rényi random graph [8] is the simplest and most studied example of a random graph ensemble. Such a graph, denoted by  $G(N, p)$ , has  $N$  vertices, with each pair of vertices connected with probability  $p$ , independently of all other pairs. Its phase transition phenomena are well understood. In particular, for  $p = \frac{c}{N}$  with  $c > 1$ , the largest component in  $G(N, p)$  has  $\Theta(N)$  vertices and the second largest  $O(\ln N)$  vertices (as  $N \rightarrow \infty$ , with probability 1), for  $p = \frac{c}{N}$  with  $c < 1$  the largest component has  $O(\ln N)$  vertices and when  $p = \frac{1}{N}$ , the largest component of  $G(N, p)$  has  $\Theta(N^{2/3})$  vertices (cf. [5, 8, 17]).

Aldous [2] considered the asymptotic behavior of  $G(N, p)$  inside the “scaling window” of this phase transition, namely for  $N \rightarrow \infty$  and  $|p - 1/N|$  small enough, showing that the ordered set of component sizes rescaled by  $N^{2/3}$  then converges to an ordered set of excursion lengths of reflected inhomogeneous Brownian motion with a certain drift. Various other random graph models exhibit a phase transition phenomenon similar to the Erdős–Rényi random graph. While some further follow the same behaviour as  $G(N, p)$  in their near-critical regime, the near-critical regime of others falls into different universality classes.

For example, Nachmias and Peres [19] prove that the random graph ensemble obtained by performing percolation on a random  $d$ -regular ( $d \geq 3$ ) graph on  $N$  vertices with percolation probability  $p = 1/(d - 1) + aN^{-1/3}$  for  $a \in \mathbb{R}$  fixed, falls

---

Received June 2015; revised April 2017.

<sup>1</sup>Research funded by NSF Grants DMS-11-06627 and DMS-1613091.

*MSC2010 subject classifications.* 05C80, 82B10, 60F17.

*Key words and phrases.* Quantum random graphs, critical point, scaling limits, Brownian excursions, weak convergence.

into the same universality class as the Erdős–Rényi random graph. The random multi-graph whose  $N$  vertices are constructed using the configuration model, with its vertex degrees being i.i.d. variables, each having the distribution  $\nu$ , has a richer behavior. Indeed, Joseph [15] shows that when  $\nu$  has a finite third moment, the near-critical regime of this model falls into the Erdős–Rényi random graph’s universality class, whereas if  $\nu_k \sim ck^{-\tau}$  as  $k \rightarrow \infty$ ,  $c > 0$ ,  $\tau \in (3, 4)$  then the relevant scaling changes to  $N^{-(\tau-2)/(\tau-1)}$  and the limit is an ordered set of the excursion lengths of some other drifted process with independent increments above past minima (near-critical regime of the Erdős–Rényi universality class, is also obtained in [20] for more general class of degree distributions of finite third moment). A similar behavior has been found in the near critical regime of the Rank-1 model (a special case of the general in-homogeneous random graph studied in [6], which has received much attention recently). Such graph has random i.i.d. weights  $\{x_i\}$  associated to its vertices, and edges chosen independently, with the edge  $(i, j)$  chosen with probability  $p_{i,j} = \min\{c \frac{x_i x_j}{N}, 1\}$ , for some positive constant  $c = c(N)$  (cf. [21]). For  $x_i$  having finite third moment, the near-critical regime corresponds to  $c(N) = 1 + aN^{-1/3}$ , in which case [2] shows that this model (formulated slightly differently), falls into the Erdős–Rényi graph’s universality class (similar results have been later proved in [3, 22]). In contrast, for the Rank-1 model with power-law degrees of exponent  $\tau \in (3, 4)$ , [4] show that the sizes of the components, rescaled by  $N^{-(\tau-2)/(\tau-1)}$ , converge to hitting times of certain thinned Lévy process.

Our aim here is to study the near-critical behavior of the so-called quantum version of Erdős–Rényi random graph (QRG). We note in passing, that both the motivation and terminology come from the stochastic geometric (Fortuin–Kasteleyn type) representation of the quantum Curie–Weiss model at inverse temperature  $\beta > 0$  (we exclude here the ground state case of  $\beta = \infty$ ), in transverse magnetic field of strength  $\lambda > 0$  (at  $\lambda = 0$  it reduces to the Erdős–Rényi ensemble, see Remark 1.1). We refer the reader to [12] for more information on such stochastic geometric representations (that were originally developed in [1, 7] for the general ferromagnetic context), moving on instead, to the precise description of the QRG (as in [13]).

*The model:* With  $G_N = \{1, \dots, N\}$  and  $\mathbb{S}_\beta$  denoting the circle of length  $\beta$ , let  $\mathcal{G}_N^\beta = G_N \times \mathbb{S}_\beta$ , associating to each site  $i \in G_N$  the copy  $\mathbb{S}_\beta^i = i \times \mathbb{S}_\beta$  of  $\mathbb{S}_\beta$ , so a point in  $\mathcal{G}_N^\beta$  has two coordinates, its site (in  $G_N$ ) and time (in  $\mathbb{S}_\beta$ ) coordinates. The QRG is then the following random subset  $\mathcal{G}_N^\beta \setminus \mathcal{H}$  of  $\mathcal{G}_N^\beta$ , equipped with random links  $\bigcup_{i,j} \mathcal{L}_{i,j}$  between pairs of points of the type  $\{(i, t) \text{ and } (j, t)\}$ , for  $i \neq j$ . To construct the QRG, we first punch within each  $\mathbb{S}_\beta^i$  finitely many holes, according to independent Poisson point processes  $\mathcal{H}_i$ ,  $i \in G_N$ , of intensity  $\lambda > 0$ , so each resulting punctured circle  $\mathbb{S}_\beta^i \setminus \mathcal{H}_i$  consists of  $m_i$  disjoint connected intervals

$$(1.1) \quad \mathbb{S}_\beta^i \setminus \mathcal{H}_i = \bigcup_{l=1}^{m_i} I_i^l$$

(the number of holes  $\#\mathcal{H}_i = m_i$ , except when  $\#\mathcal{H}_i = 0$ , in which case  $m_i = 1$ ). We next add links between pairs of points in  $\mathcal{G}_N^\beta$  of the same time coordinates (i.e., between points  $(i, t)$  and  $(j, t)$  where  $i \neq j$  and  $t \in \mathbb{S}_\beta$ ), as follows. With each (unordered) pair of sites  $i, j \in G_N$ , we associate a copy  $\mathbb{S}_\beta^{i,j}$  of  $\mathbb{S}_\beta$  and a Poisson point process of links  $\mathcal{L}_{i,j}$  on  $\mathbb{S}_\beta^{i,j}$  with intensity  $\frac{1}{N}$ . The processes  $\mathcal{L}_{i,j} = \mathcal{L}_{j,i}$  are assumed to be independent for different  $(i, j)$  and also independent of the processes of holes  $\mathcal{H}_i$ . Two intervals  $I_i^l$  and  $I_j^k$  of the decomposition (1.1) are then considered to be directly connected if there exists some  $t \in \mathcal{L}_{i,j}$  such that both  $(i, t) \in I_i^l$  and  $(j, t) \in I_j^k$ . Setting  $\mathcal{H} := \bigcup_i \mathcal{H}_i$  (a finite collection of points), the decomposition

$$(1.2) \quad \mathcal{G}_N^\beta \setminus \mathcal{H} = \mathcal{C}_1 \vee \dots \vee \mathcal{C}_\ell$$

of  $\mathcal{G}_N^\beta \setminus \mathcal{H}$  into maximal connected components is, thereby, well defined (see Figure 1 for an example with  $N = 4$ ). Further, each fixed  $x \in \mathcal{G}_N^\beta$  is a.s. not in  $\mathcal{H}$ , hence the notion of the connected component  $\mathcal{C}(x)$  containing  $x$  in the decomposition (1.2), is also well defined, and hereafter the size of a connected component  $\mathcal{C}_j$  (or  $\mathcal{C}(x)$ ), means the number of intervals it contains, and  $\mathcal{P}(\mathcal{C}(x)) = \sum_I |I| 1_{I \in \mathcal{C}(x)}$  denotes the cumulative length of intervals constituting the component  $\mathcal{C}(x)$ .

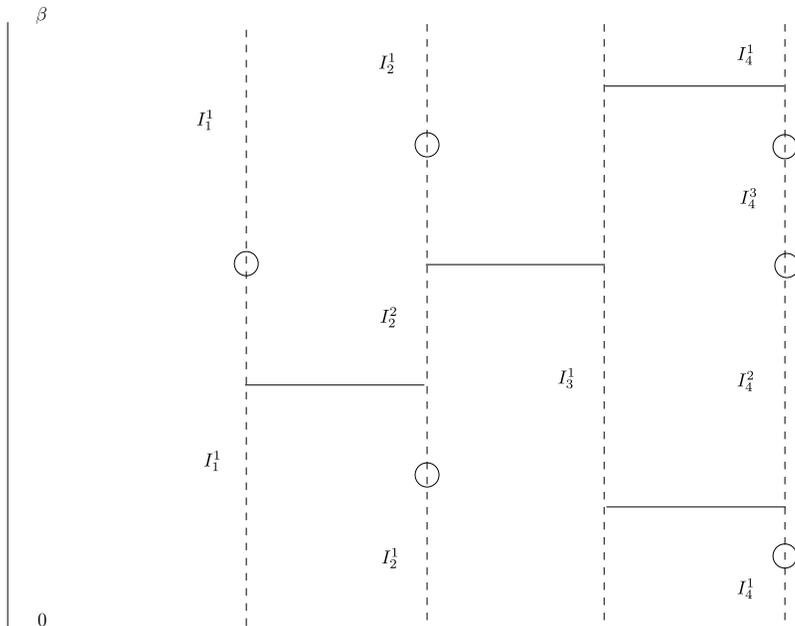


FIG. 1. An example of the decomposition of  $\mathcal{G}_N^\beta$  after all the holes are punched and the links are drawn:  $\mathcal{G}_N^\beta \setminus \mathcal{H} = \mathcal{C}_1 \vee \mathcal{C}_2 \vee \mathcal{C}_3$ , where  $\mathcal{C}_1 = I_1^1 \cup I_2^1 \cup I_4^1$ ,  $\mathcal{C}_2 = I_2^2$  and  $\mathcal{C}_3 = I_1^2 \cup I_3^1 \cup I_4^2 \cup I_4^3$ .

REMARK 1.1. For  $\lambda = 0$ , there are no holes, so each  $\mathbb{S}_\beta^i \setminus \mathcal{H}_i$  consists of one connected component, which equals to  $\mathbb{S}_\beta^i$  itself. We are then back to the Erdős–Rényi random graph  $G(N, p)$  with  $p = 1 - e^{-\frac{\beta}{N}}$  (the probability that  $\mathbb{S}_\beta^i$  and  $\mathbb{S}_\beta^j$  are directly connected).

Treating each interval  $I_i^k$  as a vertex, Janson in [14] notices that the QRG is an instance of the general in-homogeneous model of [6]. However, the probability of direct connection between two intervals depends on the size of their overlap and not only on the individual lengths of these intervals. Beyond separating our model from the class of rank-1 models, this property makes it inherently different from the other models we have mentioned thus far (all of whom mimic the idea of rank-1 random graphs, in the sense that certain *vertex related weights* determine the probabilities in which edges are present in the graph).

An equivalent description of the QRG in case  $\lambda > 0$ , which we adopt hereafter, has  $N$  circles of length  $\theta \triangleq \lambda\beta$  with a unit intensity Poisson process of holes on them, using now i.i.d. Poisson processes of intensity  $1/(\lambda N)$  for creating links between each pair of (punched) circles. The critical curve for the QRG model in the  $(\beta, \lambda)$ -parameter space, is obtained in [13] by comparisons with a critical branching process whose offspring distribution is the *cut-gamma* distribution  $\Gamma_\theta(2, 1)$  (namely, the law of  $J := (J_- + J_+) \wedge \theta$  for  $J_-, J_+$  i.i.d.  $\text{Exp}(1)$  variables). Using the preceding parametrization, the resulting curve  $\beta = \beta_c(\lambda)$  corresponds to

$$(1.3) \quad \beta_c = \frac{\theta}{F(\theta)}, \quad \lambda = F(\theta), \quad \text{for } F(\theta) = 2(1 - e^{-\theta}) - \theta e^{-\theta}$$

(where  $F(\theta)$  is precisely the expected length  $J$  of the interval  $I$  in the QRG upon our rescaling by  $\lambda$ ). It is easy to check that  $\lambda(\theta) : [0, \infty) \mapsto [0, 2)$  is concave, increasing and  $\beta_c(\theta) : [0, \infty) \mapsto [1, \infty)$  is strictly increasing, such that the curve  $\beta_c(\lambda) : [0, 2) \mapsto [1, \infty)$  is strictly increasing. The critical curve is alternatively given by

$$(1.4) \quad F(\beta, \lambda) := \lambda^{-1} F(\lambda\beta) = 1,$$

and it is further shown in [13] that taking  $F(\beta, \lambda) > 1$  (equivalently,  $\beta > \beta_c(\lambda)$ ), yields the emergence of an  $\Theta(N)$ -giant connected component in the disjoint decomposition (1.2), whereas when  $F(\beta, \lambda) < 1$  (equivalently,  $\beta < \beta_c(\lambda)$ ), all connected components are typically of order  $O(\ln N)$ . Our first result complements [13] by proving that at criticality the largest component is of size  $\Theta(N^{2/3})$  (so the QRG admits a version of the Erdős–Rényi phase transition).

THEOREM 1.2. *Suppose  $(\beta, \lambda)$  is a critical point, namely  $F(\beta, \lambda) = 1$ . Then, for the largest component  $\mathcal{C}_{\max}$  of the QRG, we have that:*

(a) *There exist  $c_*$ ,  $N_0$  and  $A_0$  finite, such that for all  $N > N_0$  and  $A > A_0$ ,*

$$(1.5) \quad \mathbb{P}(\mathcal{P}(\mathcal{C}_{\max}) > AN^{2/3}) \leq c_* A^{-3/2}.$$

(b) *There exists  $N_1$  finite such that for all  $N > N_1$  and  $\delta > 0$ ,*  
 (1.6) 
$$\mathbb{P}(\mathcal{P}(\mathcal{C}_{\max}) < \lfloor \delta N^{2/3} \rfloor) \leq (6 + 4\beta^2)\delta^{3/5}.$$

Our primary objective is to further analyze the QRG model, and in particular its component sizes near criticality, thereby confirming that the QRG is in the same universality class as the Erdős–Rényi random graph. Whereas our proofs also rely on an exploration process for estimating the connected components sizes, in contrast to all cases dealt with before (i.e., [2, 3, 15, 18, 19]), here we may have many intervals sharing the same vertex (i.e., a site  $i \in G_N$ , or alternatively, the corresponding circle  $\mathbb{S}_\theta^i$ ). Thus, our exploration process (or breadth first walk), may re-visit an already visited vertex (circle), as many times as the number of intervals sharing such vertex. The latter is an *unbounded* random variable, thereby posing a serious challenge to our analysis. While Theorem 1.2 is rough enough that we can surpass this difficulty by showing that multiple returns to same vertex are rare enough to not matter, this is no longer true for our main result, Theorem 1.3, about the scaling limits of ordered component sizes. Indeed, our limiting process drift differs from that of [2] by additional quadratic factor representing the already explored portion of the relevant circle. Indeed, the question of convergence of such quantum random graphs, as metric spaces, is completely open due to this precise problem of multiple visits to the same vertex.

The following quantities are required for our main result. First, let

(1.7) 
$$\sigma^2(\theta) = \frac{\mathbb{E}[J^2]}{F(\theta)^2},$$

(1.8) 
$$\gamma(\theta) = \frac{\mathbb{E}[\widehat{F}(\theta - J)]}{\theta F(\theta)}, \quad \text{for } \widehat{F}(x) = 2(x - 1 + e^{-x}) + \frac{x^3}{3}e^{-\theta}.$$

Then, for standard Brownian motion  $\{W(s), s \geq 0\}$  and any  $a \in \mathbb{R}$ , consider the processes

(1.9) 
$$W^{a,\theta}(s) := \sigma(\theta)W(s) + \rho^{a,\theta}(s),$$

(1.10) 
$$\rho^{a,\theta}(s) := as - \frac{s^2}{2}(1 - \gamma(\theta)),$$

and the associated process of nonnegative excursions

(1.11) 
$$B^{a,\theta}(s) = W^{a,\theta}(s) - \min_{0 \leq u \leq s} W^{a,\theta}(u).$$

**THEOREM 1.3.** *Fix  $a \in \mathbb{R}$  and  $(\beta, \lambda)$  a point on the critical curve of (1.4). Consider parameters  $(\beta_N, \lambda_N) \rightarrow (\beta, \lambda)$  such that  $F(\beta_N, \lambda_N) = 1 + aN^{-1/3}$ . Then, denoting the ordered sizes of components of the graph by  $|\mathcal{C}_1^{a,N}|, |\mathcal{C}_2^{a,N}|, \dots$ , we have when  $N \rightarrow \infty$  that*

$$(N^{-2/3}|\mathcal{C}_1^{a,N}|, N^{-2/3}|\mathcal{C}_2^{a,N}|, \dots) \xrightarrow{d} (\gamma_1, \gamma_2, \dots),$$

where  $\{\gamma_j\}$  denote the ordered lengths of the excursions of the process  $B^{a,\theta}$  above zero, and the convergence of component sizes holds with respect to the  $l^2_{\searrow}$  topology (as defined in [2]).

In Section 2, we prove Theorem 1.2 by adapting to our context the ideas set forth in [18]. Specifically, the main task here is to construct a pair of manageable auxiliary counting processes, which are not too far apart, while stochastically dominating (from above and below, resp.), the counting process that determines the size of our components (thereby circumventing much of the difficulty associated with the precise counting).

Section 3 is devoted to the proof of Theorem 1.3 which requires finer estimates and thereby some new ideas. What sets our analysis apart of all those mentioned before, is Proposition 3.4 which provides rough estimates on the number of sites of QRG visited twice, or more, during the first  $k = O(N^{2/3})$  steps of the exploration process. It shows in particular that only the *first return to a site* plays a crucial role, with subsequent returns playing no role when the relevant limit is considered. Combined with a further rough estimate on the number of sites visited exactly once during the first  $k$  steps, it thus allows us to thereafter adapt the program of [2] to the QRG setting. Specifically, Section 3.1 deals with weak convergence of the law induced by the rescaled breadth first walk to the law of  $W^{a,\theta}$  defined on the space of RCLL functions  $D([0, \infty))$ , equipped with the topology of uniform convergence on finite intervals. Finally, in Section 3.2 we collate all the above results into a proof of Theorem 1.3.

A further insight gained from our proof is that the QRG model is in the Erdős–Rényi universality class by the confluence of two reasons: first, the small probability of many returns to the same vertex (circle); second, and more crucial is the relatively fast relaxation of its exploration process, which thereby behaves approximately as a Markov process. One may examine the latter feature in many other inhomogeneous random graph models, and where it is present, proceed to try proving that they too belong to the Erdős–Rényi universality class.

**2. Proof of Theorem 1.2.** We shall first prove that  $|C_{\max}| = \Theta_{\mathbb{P}}(N^{2/3})$ , then to conclude the result of Theorem 1.2, we shall use natural bounds arising from the arguments used to prove the former.

In particular, our first step toward proving Theorem 1.2 will be the following proposition.

**PROPOSITION 2.1.** *Suppose  $(\beta, \lambda)$  is a critical point, namely  $F(\beta, \lambda) = 1$ . Then, for the largest component  $C_{\max}$  of the QRG, we have that:*

- (a) *There exist  $c_*$ ,  $N_0$  and  $A_0$  finite, such that for all  $N > N_0$  and  $A > A_0$ ,*
- $$(2.1) \quad \mathbb{P}(|C_{\max}| > AN^{2/3}) \leq c_* A^{-3/2}.$$

(b) *There exists  $N_1$  finite such that for all  $N > N_1$  and  $\delta > 0$ ,*  
 (2.2) 
$$\mathbb{P}(|C_{\max}| < \lfloor \delta N^{2/3} \rfloor) \leq (6 + 4\beta^2)\delta^{3/5}.$$

In proving  $|C_{\max}| = \Theta_{\mathbb{P}}(N^{2/3})$ , to bypass the problem of multiple visits of the same vertex by our exploration process (as described in Section 2.1), we stochastically sandwich it between the overcounting process of Section 2.2, and the undercounting process of Section 2.3. By combining the upper and lower bounds provided by these two auxiliary processes, we complete the proof of Theorem 1.2.

2.1. *Exploration process.* Taking advantage of conditional independence properties of Poisson processes, we start with an algorithmic definition of the exploration process for our model (following [13], who used it for examining a single component). This algorithm allows us to sequentially construct (or sample), the rescaled QRG, interval by interval. In this description, our vertices (circles) have first been labeled  $\{1, 2, \dots, N\}$ , and after  $k$  steps of the algorithm, we fully explore  $k$  intervals, having  $A_k \geq 0$  active points (unexplored ends of connections with the already explored intervals, or the point around which a new component starts), while the rest of the space is declared to be neutral (note that we start exploring a new component of the graph upon arriving at  $A_{k-1} = 0$ , but not before).

*Initial stage:* We fix the vertex  $w_0 = 1$  and choose a point  $t$  uniformly at random on this vertex. At end of step  $k = 0$ , we have  $A_0 = 1$ , with  $(w_0, t)$  as our sole active point and the whole space considered neutral.

*At step  $k \geq 1$ :*

(a) If  $A_{k-1} > 0$ , we choose an active point  $(w_k, t)$  whose vertex has the smallest index among all active points. In case of a tie, choose the active point which chronologically appeared earlier than the others on the same vertex.

(b) If  $A_{k-1} = 0$  and there exists at least one neutral circle, we choose  $w_k$  to be the neutral vertex with the smallest index and uniformly at random mark a new active point  $(w_k, t)$  on this vertex.

(c) If  $A_{k-1} = 0$  and there is no neutral circle, we choose  $w_k$  to be the vertex of smallest index among the vertices having some neutral part, marking new active point  $(w_k, t)$  uniformly at random on the neutral part of  $\mathbb{S}_\theta^{w_k}$ .

(d) If  $A_{k-1} = 0$  and there is no neutral part available on any circle, then this ends the exploration process.

Using i.i.d.  $\text{Exp}(1)$  variables  $J_-, J_+$ , we carve out of the maximal neutral interval  $\{w_k\} \times (t_1, t_2)$  around  $(w_k, t)$ , the subinterval  $I_k := \{w_k\} \times \tilde{I}$  for  $\tilde{I} = (t_1 \vee (t - J_-), t_2 \wedge (t + J_+))$ . For  $\mathbb{S}_\theta^{w_k}$  completely neutral (apart from active points), we take  $t_2 = -t_1 = \infty$  and  $\tilde{I} = \mathbb{S}_\theta$  whenever  $J_- + J_+ \geq \theta$  (resulting with the length of  $\tilde{I}$  having the  $\Gamma_\theta(2, 1)$  law). We then remove from the list of active points all those points which got encompassed by the interval  $I_k$ , including the base point  $(w_k, t)$ . The links in the graph connected to all such points other than  $(w_k, t)$ , are considered to be surplus edges.

*Connections of  $I_k$ :* With  $I_k = \{w_k\} \times \tilde{I}_k$ , for each  $i \neq w_k$ , we view  $\tilde{I}_k$  as a subset of  $\mathbb{S}^{w_k, i}$  and sequentially for  $i = 1, 2, \dots, N$ , sample the process of links  $\mathcal{L}_{w_k, i}$  for  $t' \geq 0$  restricted to  $\tilde{I}_k$ . We erase all links between  $I_k$  and points on already explored intervals, and register each link end  $(i, t')$  on the neutral space as an active point, labeled with the time (order) of its registration.

When done examining all the connections from  $I_k$ , we change its status from neutral to that of an explored interval and increase  $k$  by one, continuing with this procedure until no neutral space remains (which happens after finitely many steps, since the number of intervals in the QRG is finite). To recover the resulting QRG, we need only to keep track of the explored intervals (end-points), and the  $\zeta_k$  new links that have been formed in each step.

Now, let  $\eta_k = \zeta_k - (\text{sur}(k) - \text{sur}(k - 1))$ , where  $\text{sur}(k)$  counts all the surplus edges found by the end of each of the first  $k$  steps of exploration. Then, by definition,

$$(2.3) \quad A_k = \begin{cases} A_{k-1} + \eta_k - 1, & \text{if } A_{k-1} > 0, \\ \eta_k, & \text{if } A_{k-1} = 0. \end{cases}$$

As mentioned before, the exploration of the first component containing the point  $(1, t)$  sampled at the initial stage ends at  $\tau_1 = \min\{k \geq 1 : A_k = 0\}$ , with its size  $|\mathcal{C}(1, t)|$  being  $\tau_1$  (the number of explored intervals thus far). A new component whose size is  $\tau_2 - \tau_1$  is then explored from step  $\tau_1$  until the end of step  $\tau_2 = \min\{k > \tau_1 : A_k = 0\}$ , and so on.

*2.2. Overcounting.* Let  $m_i^{(T)} \leq m_i$  count the intervals in vertex  $i$  which belong to components of  $\mathcal{G}_N^\beta$  whose sizes exceed  $T$  and  $\mathcal{C}(i, *)$  denotes the connected component of  $\mathcal{G}_N^\beta$  containing  $\mathbb{S}_\theta^i$ , after erasing all the holes punched in  $\mathbb{S}_\theta^i$  by  $\mathcal{H}_i$ . Since the size of the component containing interval  $I_i^l$  of (1.1) is at most  $|\mathcal{C}(i, *)| + m_i - 1$ , it follows that  $m_i^{(2T)} = 0$  whenever both  $m_i \leq T$  and  $|\mathcal{C}(i, *)| \leq T$ . Hence, by Markov's inequality

$$(2.4) \quad \begin{aligned} \mathbb{P}(|\mathcal{C}_{\max}| \geq 2T) &\leq \mathbb{P}\left(\sum_{i=1}^N m_i^{(2T)} \geq 2T\right) \\ &\leq \frac{1}{2T} \sum_{i=1}^N \mathbb{E}[m_i^{(2T)}] \\ &\leq \frac{1}{2T} \sum_{i=1}^N \mathbb{E}[m_i(\mathbf{1}_{|\mathcal{C}(i, *)| > T} + \mathbf{1}_{m_i > T})]. \end{aligned}$$

Further,  $|\mathcal{C}(i, *)|$ ,  $i \in G_N$ , are identically distributed random variables, each of which is independent of the corresponding variable  $m_i$  which in turn has the

Poisson( $\theta$ )  $\vee$  1 distribution (so  $\mathbb{E}[m_i] = \theta + e^{-\theta}$ ). Consequently,

$$(2.5) \quad \mathbb{P}(|\mathcal{C}_{\max}| \geq 2T) \leq \frac{N(\theta + e^{-\theta})}{2T} [\mathbb{P}(|\mathcal{C}(1, *)| > T) + \mathbb{P}(m_1 \geq T)].$$

Taking  $T = (A/2)N^{2/3} = H^2$ , we thus establish part (a) of Theorem 1.2, upon showing that for some  $c$  finite and all  $H$  and  $N$  large enough

$$(2.6) \quad \mathbb{P}(|\mathcal{C}(1, *)| > H^2) \leq \frac{c}{H}.$$

Since  $\mathcal{C}(1, *)$  corresponds to the exploration process starting at  $I_1 = \{1\} \times \mathbb{S}_\theta$ , it suffices to consider the value of  $\tau_1$  when the corresponding  $\{\eta_k\}_{k \geq 1}$  are replaced in (2.3) by another, simpler to analyze, collection  $\{\xi_k\}$  that stochastically dominate them.

For us to be able to estimate tail probabilities of  $\tau_1$  using the i.i.d. sequence  $\{\xi_k\}$ , we must define appropriate coupling between  $\{\eta_k\}$  and  $\{\xi_k\}$ . To this end, let us define

$$(2.7) \quad S_n = S_1 + \sum_{k=2}^n (\xi_k - 1),$$

and consider the following monotone coupling between  $A_k$  associated to  $\mathcal{C}(1, *)$  and  $S_k$  up to time  $\tau_1$ .

Like  $A_0$ , we begin with setting  $S_0 = 1$ . Since  $\mathcal{C}(1, *)$  corresponds to the exploration process starting at  $I_1 = \{1\} \times \mathbb{S}_\theta$ ,  $\eta_1$  follows Poisson( $\frac{(N-1)\theta}{N\lambda}$ ) distribution. Let  $\xi_1$  be  $\eta_1$  together with *self-links* of the interval to itself. Therefore,  $\xi_1$  follows Poisson( $\frac{\theta}{\lambda}$ ). Since the coupling is up to time  $\tau_1$ , we only have to consider the  $A_{k-1} > 0$  case at step  $k \geq 2$ . As explained above, we choose an active point  $(w_k, t)$  and sample links included in the counting toward  $\zeta_k$ . In order to define  $\xi_k$ , recall the procedure of sampling connections and consider the following addition to it. In addition to carving out of the maximal neutral interval  $I_k$  around  $(w_k, t)$ , consider also the full interval  $I_k^S$  around  $(w_k, t)$  having the length law  $\Gamma_\theta(2, 1)$ . In addition to the links sampled by  $\mathcal{L}_{w_k, i}$ -s  $\forall i \neq w_k$  restricted to  $\tilde{I}_k$ , run a unit intensity Poisson process on  $I_k^S \cap I_k^c$ , and another independent Poisson process of ( $\frac{1}{N}$ ) intensity on the interval  $\tilde{I}_k$  counting self-links to the same interval. Let  $\nu_k$  count the arrival points of these additional Poisson processes. Define  $\xi_k = \zeta_k^S + \nu_k$  with  $\zeta_k^S$  counting all the links created by  $\mathcal{L}_{w_k, i}$ -s restricted to  $\tilde{I}_k$  without erasing those whose end points fall on already explored intervals. Obviously,  $\xi_k \geq \eta_k$  for all  $k \geq 1$ . Note also that for  $k \geq 2$ , the random variables  $\xi_k$  are i.i.d. each following Poisson( $\frac{J_k}{\lambda}$ ) conditioned on  $J_k$ , which are i.i.d.  $\Gamma_\theta(2, 1)$ , and independent of the  $\xi_1$ .

Defining  $\tau = \min\{n \geq 1 : S_n = 0\}$  as the first hitting time of zero by the process  $S_k$ , and using the monotone coupling argument, the inequality in (2.6) follows from the bound

$$(2.8) \quad \mathbb{P}(\tau > H^2) \leq \frac{c}{H}.$$

Having  $\{-1, 0, 1, 2, \dots\}$ -valued increments, recall Kemperman’s formula for such a random walk, stating that for any  $\ell \geq 0$  and  $n \geq 1$ ,

$$\mathbb{P}(\tau = n + 1 | S_1 = \ell) = \frac{\ell}{n} \mathbb{P}(S_{n+1} = 0 | S_1 = \ell)$$

(see [10], Theorem 7, page 165). Then we can write

$$\begin{aligned} \mathbb{P}(\tau = n + 1) &= \sum_{\ell=0}^n \frac{\ell}{n} \mathbb{P}(S_{n+1} - S_1 = -\ell) \mathbb{P}(S_1 = \ell) \\ (2.9) \qquad \qquad &\leq \frac{\mathbb{E}(S_1)}{n} \sup_{\ell} \{\mathbb{P}(S_{n+1} - S_1 = -\ell)\}. \end{aligned}$$

Our assumption that  $F(\beta, \lambda) = 1$ , implies that  $\mathbb{E}\xi_2 = \lambda^{-1} \mathbb{E}J_2 = 1$ , so  $\{S_n\}_{n \geq 2}$  has zero-mean i.i.d. increments of finite exponential tails. Thus, applying the local CLT for the lattice random walk  $S_{n+1} - S_1$  (see [16], Proposition 2.4.4), we deduce from (2.9) that

$$(2.10) \qquad \qquad \mathbb{P}(\tau = n + 1) \leq cn^{-3/2},$$

for some  $c$  finite and all  $n$ , which together with (2.8), proves that

$$(2.11) \qquad \mathbb{P}(|C_{\max}| > AN^{2/3}) \leq c_* A^{-3/2}.$$

**2.3. Undercounting.** To bound the lower tail of  $|C_{\max}|$ , we construct a stochastic lower bound for all component sizes by following a more restrictive exploration process, which after forming the first active point on each vertex  $w \in G_N$ , voids all space on that same vertex beyond the relevant interval around this active point (thus sequentially producing components with no more intervals than does the original exploration process). Specifically, after the initial stage, at each step  $k \geq 1$  the restrictive exploration considers for  $I_k$  only active intervals or completely neutral circles (as in parts (a) and (b) of the original exploration process defined in Section 2.1), until none such are left. It also keeps at most one connection from  $I_k$  to any, as of yet, never visited (in particular, completely neutral) circle  $S_{\theta}^i$ , ignoring (erasing) all the other links which are being formed in step  $k$  by the original exploration process. Note that this restrictive process has no surplus edges and its number of active points  $A_k^f$ , starts at  $A_0^f = 1$  and follows the recursion

$$(2.12) \qquad A_k^f = \begin{cases} A_{k-1}^f + \eta_k^f - 1, & \text{if } A_{k-1}^f > 0 \\ \eta_k^f, & \text{if } A_{k-1}^f = 0. \end{cases}$$

Here, conditioned on  $A_{k-1}^f$  and  $J_k$ , the variables  $\eta_k^f$  are independent variables distributed as  $\text{Bin}(N_{k-1}^f, 1 - e^{-J_k/(\lambda N)})$  for i.i.d.  $\Gamma_{\theta}(2, 1)$ -distributed collection  $\{J_k\}$  and  $N_k^f := N - k - (A_k^f \vee 1)$ . As before, the component sizes are given by

$\tau_r^f - \tau_{r-1}^f$ , for successive returns to zero  $\tau_r^f = \min\{k > \tau_{r-1}^f : A_k^f = 0\}$ , starting at  $\tau_0^f = 0$ .

Had we replaced  $J_k$  by  $\mathbb{E}J_k = \lambda$ , it would have resulted in the exploration process for the (effectively) critical Erdős–Rényi random graph  $G(N, 1 - e^{-1/N})$ , for which (2.2) is well known, for example, see [18], Theorem 2. As we are not aware of a study of component sizes for our inhomogeneous graph, we next adapt the proof of [18], Theorem 2, to our context.

First, from the recursion (2.12) conditioned on the event  $\{A_{k-1}^f > 0\}$ , then

$$(A_k^f)^2 - (A_{k-1}^f)^2 = (\eta_k^f - 1)^2 + 2(\eta_k^f - 1)A_{k-1}^f.$$

Conditioned on the event  $\{0 < A_{k-1}^f \leq h\}$  for some arbitrary  $h > 0$ , which we shall specify later, we observe that

$$(2.13) \quad \mathbb{E}[(A_k^f)^2 - (A_{k-1}^f)^2 | A_{k-1}^f] \geq \frac{(N - h - k)^2}{(\lambda N)^2} \mathbb{E}[J_k^2] - 2\frac{h}{N}(h + k) + O\left(\frac{1}{N}\right).$$

Further, since  $\mathbb{E}[J_k^2] > \lambda^2$ , so for  $h = c_h N^{1/3}$ , for all  $k \leq T_h$  with  $T_h = c_1 N^{2/3}$ , and large enough  $N$ ,

$$(2.14) \quad \mathbb{E}[(A_k^f)^2 - (A_{k-1}^f)^2 | A_{k-1}^f] \geq 1 - 2c_h c_1,$$

where  $c_h$  and  $c_1$  are arbitrary positive constants to be chosen later.

The latter bound applies also when  $A_{k-1}^f = 0$ , as then  $A_k^f = \eta_k^f$ . Now, taking  $b := 1 - 2c_h c_1 > 0$ , we consider the stopping time

$$\tau_h = T_h \wedge \min\{k \geq 0 : A_k^f \geq h\}$$

noting that by the preceding calculation,  $L_k := (A_{k \wedge \tau_h}^f)^2 - b(k \wedge \tau_h)$  is a submartingale starting at  $L_0 = 1$ . Further, it is shown in [18], Proof of Lemma 5, that for  $\xi$  a  $\text{Bin}(n, p)$  variable, and any  $n \leq N$ , the distribution of  $\xi - r$ , conditioned on the event  $\{\xi \geq r\}$ , is stochastically dominated by  $\text{Bin}(N, p)$ . Thus, in our setting, given  $\tau_h = k \leq T_h$ ,  $N_{k-1}^f = n \leq N$ ,  $A_{k-1}^f = \ell$  and  $p_N = 1 - e^{-J_k/(\lambda N)}$ , with  $\{J_k\}$  i.i.d  $\Gamma_\theta(2, 1)$ , we have that conditioned on the event  $\{A_k^f \geq h\}$ , the distribution of  $(A_k^f - h)$  is stochastically dominated by  $\text{Bin}(N, p_N)$  conditioned on the same  $p_N$ . Averaging over all possible  $n, k, \ell, p_N$  values, we deduce that conditioned on  $\{A_{\tau_h}^f \geq h\}$  the overshoot  $(A_{\tau_h}^f - h)$  is stochastically dominated by  $\text{Bin}(N, p_N)$  conditioned on  $p_N = 1 - e^{-J/(\lambda N)}$  with  $J \sim \Gamma_\theta(2, 1)$ . Consequently,

$$\mathbb{E}[(A_{\tau_h}^f)^2] \leq h^2 + 2h\mathbb{E}[Np_N] + \mathbb{E}[(Np_N)^2] + \mathbb{E}[Np_N(1 - p_N)] \leq (h + 2)^2,$$

for  $h = c_h N^{1/3}$  and all  $N$  large enough. With  $\tau_h \leq T_h$ , upon applying the optional stopping theorem for the  $L^2$ -bounded submartingale  $\{L_k\}$ , we find that  $\mathbb{E}[(A_{\tau_h}^f)^2] \geq 1 + b\mathbb{E}(\tau_h)$ . Thus, by Markov's inequality,

$$(2.15) \quad \mathbb{P}(\tau_h = T_h) \leq \frac{\mathbb{E}[\tau_h]}{T_h} \leq \frac{(h + 2)^2}{bT_h} = \frac{c_h^2}{bc_1}(1 + o(1)).$$

Fixing  $T_0 = \delta N^{2/3}$  note that  $|\mathcal{C}_{\max}|$  exceeds the value of the stopping time

$$\tau_0 = T_0 \wedge \min\{s \geq 0 : A_{\tau_h+s}^f = 0\},$$

so the stated bound (2.2) follows once we show that for any  $b' > \lambda^{-2}\mathbb{E}[J^2] + 2c_h(c_1 + \delta)$ ,

$$(2.16) \quad \mathbb{P}(\tau_0 < T_0 | A_{\tau_h}^f \geq h) \leq \frac{b'\delta}{c_h^2}.$$

Indeed, we then choose  $c_h = \frac{1}{2}\delta^{1/5}$  and  $c_1 = 1/(4c_h)$ , so  $b = 1/2$  and  $b' = \beta^2 + 1$  works whenever  $\delta \leq 1/2$ .

To derive (2.16), consider the uniformly bounded, nonnegative process  $M_k = \max\{h - A_{\tau_h+k}^f, 0\}$ . If  $0 < M_{k-1} < h$  then by (2.12)

$$M_k^2 - M_{k-1}^2 \leq (\eta_{\tau_h+k}^f - 1)^2 + 2(1 - \eta_{\tau_h+k}^f)M_{k-1}.$$

The same inequality applies when  $M_{k-1} = 0$  (i.e.,  $A_{\tau_h+k-1}^f \geq h$ , so  $M_k \leq \max\{1 - \eta_{\tau_h+k}^f, 0\}$ ). By definition,  $A_{\tau_h+k-1}^f \leq h$  whenever  $M_{k-1} \neq 0$ , hence for  $N$  large enough and all  $k \leq T_0$ ,  $\tau_h \leq T_h$  we find, as in the derivation of (2.14) that

$$\mathbb{E}[M_k^2 - M_{k-1}^2 | M_{k-1} < h] \leq \lambda^{-2}\mathbb{E}[J^2] + 2\frac{h}{N}(T_h + T_0 + h) + o(1) \leq b'.$$

Thus, conditioned on the event  $A_{\tau_h}^f \geq h$  the process  $\{M_{k \wedge \tau_0}^2 - b'(k \wedge \tau_0)\}$  is a supermartingale which starts at zero. Noting that  $\{\tau_0 < T_0\} \subseteq \{M_{\tau_0} = h\}$ , upon applying the optional stopping theorem for this process at  $\tau_0$ , we conclude that

$$\mathbb{P}(\tau_0 < T_0 | A_{\tau_h}^f \geq h) \leq h^{-2}\mathbb{E}[M_{\tau_0}^2 | A_{\tau_h}^f \geq h] \leq b'h^{-2}\mathbb{E}[\tau_0 | A_{\tau_h}^f \geq h] \leq b'\frac{T_0}{h^2},$$

as stated.  $\square$

**PROOF OF THEOREM 1.2.** Observe that  $\mathcal{P}(\mathcal{C}_{\max})$  is stochastically dominated by  $\sum_{i=1}^{|\mathcal{C}_{\max}|} \xi_i$ , where  $\{\xi_i\}$  are i.i.d. cut-Gamma random variables. Repeating the arguments set forth in Section 2.2 we can easily conclude the upper bound as in equation (2.1).

Similarly, we shall propose a process using the same undercounting algorithm such that it is stochastically dominated by  $\mathcal{P}(\mathcal{C}_{\max})$ . In particular, instead of accounting only the number of once visited vertices, let us associate a cut gamma

$\Gamma_\theta(2, 1)$  random variable with every such vertex visited. Then consider the sum of all such random variables, which clearly is dominated by  $\mathcal{P}(\mathcal{C}_{\max})$ . Thereafter, again following the same steps as in Section 2.3 we conclude the required result of Theorem 1.2.  $\square$

### 3. Proof of Theorem 1.3.

3.1. *Exploration process and Brownian excursions.* Recall the length of a sampled interval of the QRG being  $J = \min(J_- + J_+, \theta)$  for i.i.d.  $\text{Exp}(1)$  variables  $J_-, J_+$  (a distribution we denote by  $\Gamma_\theta(2, 1)$ ). With  $\mathbb{E}(J) = F(\theta)$  of (1.3), the critical curve has the explicit expression  $\lambda = F(\theta)$ . Further, the critical window around some  $\lambda_\star = F(\theta_\star)$  for  $\theta_\star > 0$ , corresponds to fixing  $a \in \mathbb{R}$  and considering

$$(3.1) \quad \theta_N \rightarrow \theta_\star, \quad F(\theta_N) = \lambda_N(1 + aN^{-1/3}).$$

Let  $\tilde{N} = \sum_{i=1}^N m_i$  denote the total number of steps in the exploration process of Section 2.1, and  $(Y_N^{a, \theta_N}(k), k \leq \tilde{N})$  be the breadth-first walk associated with the QRG on  $\mathcal{G}_{\theta_N}^N$ , where  $(\theta_N, \lambda_N)$  satisfy (3.1). That is,

$$(3.2) \quad Y_N^{a, \theta_N}(k) = Y_N^{a, \theta_N}(k - 1) + \eta_k - 1, \quad Y_N^{a, \theta_N}(0) = 1,$$

for  $\eta_k$  of recursion (2.3). Thus,  $Y_N^{a, \theta_N}(k)$  (which may well become negative as  $k$  grows), counts the number of active points at the end of step  $k$ , minus the number of explored components before step  $k$ .

As in [2], observe that lengths of excursions of the process  $Y_N^{a, \theta_N, +}(k) = Y_N^{a, \theta_N}(k) - \min_{l \leq k} Y_N^{a, \theta_N}(l)$  above zero correspond to size of the connected component containing the vertex where the process  $Y_N^{a, \theta_N}(k)$  started.

Setting

$$(3.3) \quad \sigma_\star^2 = \sigma^2(\theta_\star) = \frac{1}{\lambda_\star^2} \mathbb{E}[(J_- + J_+) \wedge \theta_\star]^2,$$

our goal in this subsection is to prove the following proposition.

PROPOSITION 3.1. *For  $(\theta_N, \lambda_N)$  that satisfy (3.1), as  $N \rightarrow \infty$ , the processes*

$$(3.4) \quad \bar{Y}_N^{a, \theta_N}(s) = N^{-1/3} Y_N^{a, \theta_N}(\lfloor N^{2/3} s \rfloor \wedge \tilde{N}),$$

*converge in law to  $W^{a, \theta_\star}$  of (1.9) (on the space  $D([0, \infty))$  equipped with the topology of uniform convergence on compacts).*

Recall that  $\zeta_k$  links are generated at step  $k$  of the exploration process and let

$$(3.5) \quad Z_N^{a, \theta_N}(l) = Z_N^{a, \theta_N}(l - 1) + \zeta_l - 1, \quad Z_N^{a, \theta_N}(0) = 1,$$

be the corresponding breadth-first walk. Since we sample intervals only when they are to be explored, the walk  $Z_N$  does not distinguish between active points that end as intervals of the QRG and those that are later found to be on surplus edges. Nevertheless, our next proposition controls the number of active points, which as seen in Remark 3.3, yields having at most  $O_{\mathbb{P}}(1)$  surplus edges until step  $sN^{2/3}$ .

PROPOSITION 3.2. *Fixing  $a \in \mathbb{R}$ , recall the count  $A_k$  of active points at the end of step  $k$  of the exploration for  $(\theta_N, \lambda_N)$  satisfying (3.1). Then, for some  $K = K(a, s) < \infty$  and all  $L, N \geq L_0(a, s)$ ,*

$$(3.6) \quad \mathbb{P}\left(N^{-1/3} \max_{k \leq sN^{2/3}} \{A_k\} > L\right) \leq KL^{-2}.$$

The proof of above proposition involves elaborate, but crude, bounds on functionals of  $Z_N$ , and thus we defer it to the [Appendix](#).

REMARK 3.3. Recall  $\text{sur}(l)$  counts the surplus links detected in part (a) of the exploration process during its first  $l$  steps. The order of exploring active points is such that the first active point formed on any given vertex never contributes to  $\text{sur}(l)$ . Further,  $\text{sur}(l)$  is bounded above by the aggregate count  $\Delta_e(l)$  of active points on vertex  $w_k$  at step  $k \leq l$ , beyond the explored point  $(w_k, t)$ . Conditional on the state of the process at the start of its  $k$ th step, the number of active points registered during that step that may contribute to  $\Delta_e(l)$ , is stochastically dominated by a  $\text{Poisson}(A_{k-1}\theta_N(\lambda_N N)^{-1})$  random variable. In particular, for some  $\kappa$  finite and all  $N, l$ ,

$$(3.7) \quad \mathbb{E}[\text{sur}(l)] \leq \mathbb{E}[\Delta_e(l)] \leq \frac{\kappa}{N} \mathbb{E}\left[\sum_{k=0}^{l-1} A_k\right] \leq \frac{\kappa l}{N} \mathbb{E}\left[\max_{k \leq l} \{A_k\}\right].$$

From Proposition 3.2, we have that  $\mathbb{E}[\Delta_e(sN^{2/3})]$ , and hence  $\mathbb{E}[\text{sur}(sN^{2/3})]$ , is uniformly bounded in  $N$ .

The control on number of vertices, which the exploration process visits at least twice by the end of the  $k$ th step (for  $k = sN^{2/3}$ ), is crucial for the success of our analysis. To this end, we define hereafter *the number of visits to vertex  $v \in G_N$  by the end of the  $k$ th exploration step*, as the total number of active points formed on  $S_{\theta_N}^v$  by that time, that is, we count past active points which were removed from the list and also those which are currently active by the end of time  $k$ .

PROPOSITION 3.4. *For the exploration of the QRG at parameters satisfying (3.1), let  $v_{\geq m}^l$  count the total number of visits by the end of its  $l$ th step, to sites (circles), having at least  $m$  such visits each. Then, for some finite  $\kappa$ , all positive  $A, s$  and  $N$  large enough, we have that for  $m = 1, 2, 3$  and any  $l \in [1, sN^{2/3}]$ ,*

$$(3.8) \quad \mathbb{P}(v_{\geq m}^l \geq Al^m N^{1-m}) \leq \frac{\kappa^m}{A}.$$

PROOF. Let  $\mathcal{F}_t$  denote the filtration generated by the state of the exploration process of Section 2.1, namely, the neutral subspace and collection of active points, with  $\mathcal{F}_k$  for integer  $k \geq 0$ , denoting the state at the end of step  $k$ . Further, let  $\tau(v)$  the stage in which it first visits  $v \in G_N$  (so  $\mathcal{F}_{\tau(v)}$  records the state of the exploration process immediately after selecting its first active point on  $\mathbb{S}_\theta^v$ ). For any  $l \geq 1$  and  $v$ , let  $L_v(l)$  count the links whose end points are on the neutral part of  $v^2$  until the end of the  $l$ th step of that process, with  $L'_v(l)$  counting only such links made after  $\tau(v)$  (setting  $L'_v(l) = 0$  in case  $\tau(v) \geq l + 1$ ). Only one interval is explored in each step, hence

$$(3.9) \quad \mathbb{E}[v_{\geq 1}^l] \leq l + \sum_{v=1}^N \mathbb{E}[L_v(l)].$$

Note that  $L_v(l)$  increases when considering the overcounting process, so the  $l$  explored intervals are complete circles (of length  $\theta_N$ ), other than the circle at  $v$ , which remains completely neutral, even after links to it are formed and intervals are sampled around the links. Thus,  $L_v(l)$  is stochastically dominated by a Poisson random variable with parameter

$$(3.10) \quad \frac{\theta_N l}{\lambda_N N} \leq \frac{\kappa' l}{N},$$

for some  $\kappa'$  finite and all  $N$  large enough. In particular, by (3.9)  $\mathbb{E}[v_{\geq 1}^l] \leq (\kappa' + 1)l$ , which in combination with Markov's inequality establishes (3.8) for  $m = 1$  and any  $\kappa \geq \kappa' + 1$ .

Next recall that for  $k \leq N$ , either  $A_k > 0$  so part (a) of the exploration process applies at the  $k$ th step, or else part (b) applies for it (since at most  $k - 1 < N$  vertices have been explored before). Consequently, assuming hereafter that  $sN^{2/3} \leq N$ , part (c) of the exploration does not occur throughout its first  $l$  steps. Further, all active points chosen in part (b) or the initial stage of the process result with a *first* visit of new vertex. Hence we have in analogy with (3.9) that

$$(3.11) \quad \mathbb{E}[v_{\geq 2}^l] \leq 2 \sum_{v=1}^N \mathbb{E}[L'_v(l)].$$

As argued before, conditional on  $\mathcal{F}_{\tau(v)}$  the value of  $L'_v(l)$  increases if from time  $\tau(v)$  onward we modify the process to have all explored intervals be complete circles (of length  $\theta_N$ ), on vertices other than  $v$ , while keeping the circle at  $v$  completely neutral. That is, conditionally on  $\mathcal{F}_{\tau(v)}$  the variable  $L'_v(l)$  is stochastically

---

<sup>2</sup>We note here that  $L_v(l)$  is different from the total number of visits to the vertex  $v$  by time  $l$  because according to our description, the point where we start the exploration process, and the points where we restart our exploration process after  $A_k = 0$  are indeed counted as visits but these are not identified as end points of links.

dominated by a Poisson variable of parameter  $\kappa' l N^{-1}$  times the indicator on the event  $\{\tau(v) < l + 1\}$ . Hence, for any  $l \geq 1$  and  $v \in G_N$ ,

$$(3.12) \quad \mathbb{E}[L'_v(l) | \mathcal{F}_{\tau(v)}] \leq \left(\frac{\kappa' l}{N}\right) \mathbf{1}_{\{\tau(v) < l+1\}}.$$

Summing over  $v$  the expected value of (3.12), we deduce from (3.11) that

$$(3.13) \quad \mathbb{E}[v^l_{\geq 2}] \leq 2 \left(\frac{\kappa' l}{N}\right) \mathbb{E}[v^l_{\geq 1}] \leq 2(\kappa' + 1) \kappa' l^2 N^{-1}$$

from which we recover (3.8) for  $m = 2$  and  $\kappa = 2\kappa' + 1$  (by Markov's inequality).

Finally, repeating this argument, now with  $\tau(v)$  the time at which the second active point on  $v \in G_N$  is selected, we deduce that

$$\mathbb{E}[v^l_{\geq 3}] \leq 3 \left(\frac{\kappa' l}{N}\right) \mathbb{E}[v^l_{\geq 2}],$$

which upon suitably increasing the value of  $\kappa$ , yields (3.8) for  $m = 3$ .  $\square$

**PROOF OF PROPOSITION 3.1.** Equipping  $D([0, \infty))$  with the topology of uniform convergence on compacts, let

$$(3.14) \quad \bar{Z}_N^{a, \theta_N}(s) = N^{-1/3} Z_N^{a, \theta_N}(\lfloor N^{2/3} s \rfloor \wedge \tilde{N}),$$

for the breadth first walk  $Z_N^{a, \theta_N}(\cdot)$  of (3.5). Recall that

$$\bar{Z}_N^{a, \theta_N}(s) - \bar{Y}_N^{a, \theta_N}(s) = N^{-1/3} \text{sur}(\lfloor N^{2/3} s \rfloor \wedge \tilde{N})$$

is nondecreasing in  $s$  and so by Remark 3.3, as  $N \rightarrow \infty$ ,

$$(3.15) \quad \sup_{s \leq s_0} |\bar{Z}_N^{a, \theta_N}(s) - \bar{Y}_N^{a, \theta_N}(s)| \rightarrow 0 \quad \text{in } \mathbb{P}.$$

It thus suffices to prove that  $\bar{Z}_N^{a, \theta_N}$  converges in law to the desired limit  $W^{a, \theta_\star}$ . To this end, by Doob's decomposition with respect to the canonical filtration  $\mathcal{F}_k$  associated with the exploration process, we get that

$$(3.16) \quad Z_N^{a, \theta_N} = M_N^{a, \theta_N} + B_N^{a, \theta_N}, \quad (M_N^{a, \theta_N})^2 = Q_N^{a, \theta_N} + D_N^{a, \theta_N},$$

with martingales  $M_N^{a, \theta_N}$ ,  $Q_N^{a, \theta_N}$  (null at  $k = 0$ ), and predictable processes  $B_N^{a, \theta_N}$  and  $D_N^{a, \theta_N}$ . Adopting the notation  $\bar{M}_N$ ,  $\bar{B}_N$ , in accordance with (3.14), and  $\bar{Q}_N$ ,  $\bar{D}_N$  similarly scaled by extra factor  $N^{-1/3}$  in accordance to the RHS of (3.16), we show in Lemmas 3.5 and 3.6, respectively, that for  $\sigma_\star$  of (3.3) and any finite  $s_0$ , as  $N \rightarrow \infty$ ,

$$(3.17) \quad \bar{D}_N^{a, \theta_N}(s_0) \xrightarrow{\mathbb{P}} \sigma_\star^2 s_0,$$

$$(3.18) \quad \mathbb{E} \left[ \sup_{s \leq s_0} |\bar{M}_N^{a, \theta_N}(s) - \bar{M}_N^{a, \theta_N}(s^-)|^2 \right] \rightarrow 0.$$

Combining (3.17) and (3.18), it then follows from [9], Theorem 7.1.4, that the martingales  $\{\bar{M}_N^{a,\theta_N}\}$  converge weakly in  $D([0, \infty))$  to  $\sigma_\star W$  for a standard Brownian motion  $W$ . Further, we show in Proposition 3.7 that

$$(3.19) \quad \sup_{s \leq s_0} |\bar{B}_N^{a,\theta_N}(s) - \rho^{a,\theta_\star}(s)| \rightarrow_{\mathbb{P}} 0.$$

That is, the sequence of predictable processes  $\bar{B}_N^{a,\theta_N}$  converges in probability in  $D([0, \infty))$  to the nonrandom  $\rho^{a,\theta_\star}$  of (1.10), hence  $\bar{Z}_N^{a,\theta_N}$  converges in law to  $W^{a,\theta_\star}$ .  $\square$

Proceeding with the proof of (3.17)–(3.19), we often drop the indices  $(a, \theta_N)$  from  $Z_N^{a,\theta_N}$  and related random variables. We start by establishing (3.17). That is, we have the following.

LEMMA 3.5. *For  $(\theta_N, \lambda_N)$  that satisfy (3.1),  $\sigma_\star$  of (3.3), any  $\delta > 0$  and  $s_0$  finite,*

$$(3.20) \quad \lim_{N \rightarrow \infty} \sup_{l \leq N^{2/3} s_0} \mathbb{P}(|D_N^{a,\theta_N}(l) - \sigma_\star^2 l| \geq 2\delta N^{2/3}) = 0.$$

PROOF. Recall that  $D_N(k) - D_N(k - 1) = \text{var}(\zeta_k | \mathcal{F}_{k-1})$  (starting at  $D_N(0) = 0$ ). Hereafter,  $\mathcal{F}_{k-}$  denotes  $\mathcal{F}_{k-1}$  augmented by the active point  $(w_k, X_k)$  of step  $k$  and the interval  $I_k = \{w_k\} \times \tilde{I}_k$  around it. The law of  $\zeta_k$  given  $\mathcal{F}_{k-}$  is Poisson( $\varphi_k$ ), with  $\varphi_k$  denoting the aggregate over circles other than  $\mathbb{S}_{\theta_N}^{w_k}$ , of the length of their neutral space restricted to  $\tilde{I}_k$  and divided by  $\lambda_N N$ . As such, both its conditional mean and conditional variance are given by  $\varphi_k$ , hence by the variance conditioning decomposition (at  $\mathcal{F}_{k-}$ ),

$$(3.21) \quad \text{var}(\zeta_k | \mathcal{F}_{k-1}) = \mathbb{E}[\varphi_k | \mathcal{F}_{k-1}] + \text{var}[\varphi_k | \mathcal{F}_{k-1}].$$

Further, at the start of the  $k$ th step there are at least  $N - k$  completely neutral circles beyond the vertex on which the  $k$ th explored interval lies. Hence, for  $J_k = |\tilde{I}_k|$  and any  $k \leq s_0 N^{2/3}$ ,

$$0 \leq \frac{J_k}{\lambda_N} - \varphi_k \leq \frac{J_k k}{\lambda_N N} \leq \kappa' N^{-1/3}$$

(for  $\kappa' = s_0 \sup_N \{\theta_N / \lambda_N\}$  finite). It thus suffices to prove (3.20) for  $\hat{D}_N(l)$  instead of  $D_N(l)$ , where

$$\hat{D}_N(k) - \hat{D}_N(k - 1) = \mathbb{E}[J_k / \lambda_N | \mathcal{F}_{k-1}] + \text{var}[J_k / \lambda_N | \mathcal{F}_{k-1}] =: \Delta_k.$$

The nonnegative  $\Delta_k$  are uniformly bounded by  $\bar{\Delta} := \sup_N \{(\theta_N / \lambda_N) + (\theta_N / \lambda_N)^2\}$ . Moreover, whenever  $\mathcal{F}_{k-1}$  dictates that the  $k$ th step explores the first interval on a given vertex  $w \in G_N$ , it yields a conditionally independent  $J_k$  that follow the

$\Gamma_{\theta_N}(2, 1)$  distribution. We consequently have that  $\Delta_k = F_2(\theta_N, \lambda_N)$  in any such step, where

$$F_2(\theta, \lambda) := \mathbb{E}[(J/\lambda)] + \text{var}(J/\lambda).$$

This applies to all but at most  $v_{\geq 2}^l$  of the first  $l$  steps, hence

$$|\widehat{D}_N(l) - F_2(\beta_N, \lambda_N)l| \leq \overline{\Delta} v_{\geq 2}^l.$$

From Proposition 3.4, we know that  $\mathbb{P}(v_{\geq 2}^l \geq \delta N^{2/3}) \rightarrow 0$  for  $l = s_0 N^{2/3}$  and we thus get (3.20) for  $\widehat{D}_N(l)$  upon noting that  $F_2(\theta_*, \lambda_*) = \sigma_*^2$  of (3.3).  $\square$

We next establish (3.18), thereby moving closer to completing the proof of Proposition 3.1.

LEMMA 3.6. *For  $(\theta_N, \lambda_N)$  that satisfy (3.1) and any  $s_0$  finite,*

$$(3.22) \quad \lim_{N \rightarrow \infty} N^{-2/3} \mathbb{E} \left[ \max_{1 \leq l \leq s_0 N^{2/3}} |M_N^{a, \theta_N}(l) - M_N^{a, \theta_N}(l-1)|^2 \right] = 0.$$

PROOF. Recall that  $M_N(\cdot)$  is the martingale part of  $Z_N(\cdot)$ . Hence, from (3.5)

$$M_N(l) - M_N(l-1) = \zeta_l - \mathbb{E}(\zeta_l | \mathcal{F}_{l-1})$$

and (3.22) amounts to showing that

$$N^{-2/3} \mathbb{E} \left[ \max_{l \leq s_0 N^{2/3}} (\zeta_l - \mathbb{E}(\zeta_l | \mathcal{F}_{l-1}))^2 \right] \rightarrow 0.$$

Clearly,  $(\zeta_l - \mathbb{E}(\zeta_l | \mathcal{F}_{l-1}))^2 \leq 2\zeta_l^2 + 2\mathbb{E}(\zeta_l | \mathcal{F}_{l-1})^2$ . Further, in Section 2.2 we saw that conditionally on  $\mathcal{F}_{l-1}$  the variable  $\zeta_l$  is stochastically dominated by the independent  $\xi_l \geq 0$  whose mean  $F(\beta_N, \lambda_N)$  is uniformly bounded (in  $N$ ). Hence,  $\sup_l \mathbb{E}(\zeta_l | \mathcal{F}_{l-1})^2 \leq \sup_N F(\beta_N, \lambda_N)^2$  is finite and it suffices to show that for i.i.d.  $(\xi_l)$ ,

$$(3.23) \quad N^{-2/3} \mathbb{E} \left[ \max_{l \leq s_0 N^{2/3}} \xi_l^2 \right] \leq N^{-2/3} \mathbb{E} \left[ \max_{l \leq s_0 N^{2/3}} \xi_l^2 \right] \rightarrow 0.$$

Finally, recall [11], equation (6'), that the expected maximum of  $n$  i.i.d. variables of zero-mean and unit variance is at most  $(n-1)/\sqrt{2n-1}$ . Consequently, the expectation on the right-hand side of (3.23) grows at most at rate  $O(N^{1/3})$ , which proves (3.22) (and thereby (3.18) as well).  $\square$

For the remainder of Section 3.1, we complete the proof of Proposition 3.1 by establishing (3.19). Indeed, upon rearranging the expression (1.10) for  $\rho^{a, \theta_*}$ , this is precisely the statement of our next proposition.

PROPOSITION 3.7. For  $(\theta_N, \lambda_N)$  that satisfy (3.1) and any  $s_0$  finite, as  $N \rightarrow \infty$ ,

$$(3.24) \quad \sup_{s \leq s_0} \left| \bar{B}_N^{a, \theta_N}(s) - as + \frac{s^2}{2} \left( 1 - \frac{F(\theta_\star)}{\theta_\star} - \gamma(\theta_\star) \right) + \frac{s^2}{2} \frac{F(\theta_\star)}{\theta_\star} \right| \xrightarrow{\mathbb{P}} 0.$$

As the starting point for Proposition 3.7, we provide the geometric quantities behind the coefficients of  $s^2$  in (3.24).

LEMMA 3.8. For any given interval  $\mathcal{I} \subseteq \mathbb{S}_\theta$ , let

$$(3.25) \quad H(\mathcal{I}) := \mathbb{E}[|\mathcal{I}' \cap \mathcal{I}|], \quad H_0(\mathcal{I}) := \mathbb{E}[|\mathcal{I}' \cap_0 \mathcal{I}|],$$

denote the expectation over an independent interval  $\mathcal{I}' \subseteq \mathbb{S}_\theta$  of length law  $\Gamma_\theta(2, 1)$ , built around 0, where

$$(3.26) \quad \mathcal{I}' \cap_0 \mathcal{I} := \begin{cases} \{x \in \mathcal{I}' : x \leftrightarrow 0 \text{ within } \mathcal{I}' \setminus \mathcal{I}\}, & 0 \notin \mathcal{I}, \\ \emptyset, & 0 \in \mathcal{I}. \end{cases}$$

For uniform  $U \in \mathbb{S}_\theta$ , independent of  $\mathcal{I}$ , let

$$(3.27) \quad \text{UH}(|\mathcal{I}|) := \mathbb{E}[H(\mathcal{I} - U)|\mathcal{I}], \quad \text{UH}_0(|\mathcal{I}|) := \mathbb{E}[H_0(\mathcal{I} - U)|\mathcal{I}].$$

Then, for  $J$  of  $\Gamma_\theta(2, 1)$  law

$$(3.28) \quad \mathbb{E}[\text{UH}(J)] = \frac{F(\theta)^2}{\theta}, \quad \mathbb{E}[\text{UH}_0(J)] = F(\theta) \left( 1 - \frac{F(\theta)}{\theta} - \gamma(\theta) \right).$$

PROOF. Recall that for two arcs  $\mathcal{I}$  and  $\mathcal{I}'$  in  $\mathbb{S}_\theta$  of uniformly chosen relative shift  $U$ , the expected length of  $\mathcal{I}' \cap (\mathcal{I} - U)$  is the product of arc lengths divided by  $\theta$ . In particular,  $\text{UH}(J) = \frac{F(\theta)}{\theta} J$ , from which the LHS of (3.28) follows. Similarly, per arc  $\mathcal{I}' = [-J_-, J_+]$  of length  $J$  around 0 in  $\mathbb{S}_\theta$  and  $x = \theta - |\mathcal{I}|$ , the expectation of  $\theta|\mathcal{I}' \cap_0 (\mathcal{I} - U)|$  over the uniform shift  $U$ , is  $x(J - x)$  for  $J = \theta$ , while for  $J < \theta$  it is

$$\int_{(J_- - x)_+}^{J_-} u \, du + \int_{(J_+ - x)_+}^{J_+} u \, du = x(J - x) + \frac{1}{2} [((x - J_+)_+)^2 + ((x - J_-)_+)^2],$$

by elementary geometric considerations. Computing the expectation of this expression for  $J_-$  and  $J_+$  independent  $\text{Exp}(1)$  variables, yields

$$(3.29) \quad \text{UH}_0(\theta - x) = \frac{1}{\theta} [x F(\theta) - \widehat{F}(x)],$$

for  $\widehat{F}(\cdot)$  as given in (1.8). Finally, note that  $\gamma(\theta)$  of (1.8) was set so the RHS of (3.28) be the expectation of (3.29) when  $(\theta - x) \sim \Gamma_\theta(2, 1)$ .  $\square$

Proceeding with the proof of Proposition 3.7, we next express  $B_N(l)$  as the sum of the terms (3.30)–(3.32), which for  $l = sN^{2/3}$  upon further scaling by  $N^{-1/3}$  converge to the three limit expressions in (3.24), respectively.

LEMMA 3.9. *Let  $X_k$  denote the position on  $\mathbb{S}_{\theta_N}^{w_k}$  around which  $\tilde{I}_k$  is carved (and if there are no active points by the end of exploration step  $(k - 1)$ , then any function of  $X_k$  is replaced by its expectation over a uniform  $U \in \mathbb{S}_{\theta_N}^{w_k}$ ). Similarly, for  $1 \leq k < l$  let  $\{X_{(k,i)}, 1 \leq i \leq r(k, l)\}$  be the collection of links formed on  $\mathbb{S}_{\theta_N}^{w_k}$  during steps  $(k, l]$  (some may be in  $\tilde{I}_k$ ), with the convention that  $r(k, l) = 0$  if vertex  $w_k$  has been explored before. Then, for  $(\theta_N, \lambda_N)$  that satisfy (3.1), any fixed  $s_0$  and all  $N$  large enough, we have uniformly over  $l \in [1, s_0 N^{2/3}]$ ,*

$$(3.30) \quad B_N(l) = O_{\mathbb{P}}(1) + l \left( \frac{F(\theta_N)}{\lambda_N} - 1 \right)$$

$$(3.31) \quad - \frac{1}{\lambda_N} \sum_{k=1}^l \sum_{i=1}^{r(k,l)} H_0(\tilde{I}_k - X_{(k,i)})$$

$$(3.32) \quad - \frac{1}{\lambda_N N} \sum_{1 \leq k < k' \leq l} H(\tilde{I}_k - X_{k'}).$$

PROOF. Recall from proof of Lemma 3.5 that conditional on  $\mathcal{F}_{k'}$  the number of links  $\zeta_{k'}$  formed during the  $k'$ th exploration step is Poisson distributed, whose parameter

$$\varphi_{k'} := \frac{1}{\lambda_N} |\tilde{I}_{k'}| - \frac{1}{\lambda_N N} \sum_{k=1}^{k'} |\tilde{I}_{k'} \cap \tilde{I}_k|,$$

is bounded by the nonrandom  $\theta_N/\lambda_N$ . Since  $B_N(l)$  is the predictable process in Doob’s decomposition of  $Z_N(l)$ , we get from (3.5) that

$$(3.33) \quad \begin{aligned} B_N(l) + (l - 1) &= \sum_{k'=1}^l \mathbb{E}(\zeta_{k'} | \mathcal{F}_{k'-1}) = \sum_{k'=1}^l \mathbb{E}(\varphi_{k'} | \mathcal{F}_{k'-1}) \\ &= \frac{1}{\lambda_N} \left( l F(\theta_N) - \sum_{k'=1}^l (F(\theta_N) - \mathbb{E}(|\tilde{I}_{k'}| | \mathcal{F}_{k'-1})) \right) \\ &\quad - \frac{1}{N} \sum_{1 \leq k \leq k' \leq l} \mathbb{E}(|\tilde{I}_{k'} \cap \tilde{I}_k| | \mathcal{F}_{k'-1}). \end{aligned}$$

Hereafter, for  $m \geq 1$  let  $E_m^k$  and  $E_{\geq m}^k$ , denote the event that upon exploring the interval  $\tilde{I}_k$  during the  $k$ th step, the site  $w_k$  on which it lies has been visited precisely  $m$  times, or at least  $m$  times, respectively. Equipped with these notation, we next show that the difference between the left sum in (3.33) and the sum in (3.31) is merely part of the  $O_{\mathbb{P}}(1)$  in (3.30). Specifically, all terms are at most  $F(\theta_N)$ , which in turn is uniformly bounded over  $\theta_N \rightarrow \theta_*$ , so it suffices to show that there are only  $O_{\mathbb{P}}(1)$  differing terms between these sums. To this end, since  $H_0(\tilde{I}_k - X_{(k,i)}) = 0$

whenever  $X_{(k,i)} \in \tilde{I}_k$  (see (3.26)), such superfluous links never contribute to (3.31). Similarly, there is no contribution to the left sum in (3.33) when  $E_1^{k'}$  occurs (i.e., when  $\tilde{I}_{k'}$  is the first explored interval on its vertex). Moreover, the contribution to the latter sum from  $\bigcup_{k' \leq l} E_2^{k'}$  is by intervals  $\tilde{I}_{k'}$  of second exploration of some  $w_k$ , such that  $k' \in (k, l]$  and

$$F(\theta_N) - \mathbb{E}(|\tilde{I}_{k'}| | \mathcal{F}_{k'-1}) = H_0(\tilde{I}_k - X_{k'})$$

measures for an independent interval  $\mathcal{I}' \subseteq \mathbb{S}_{\theta_N}$  of length law  $\Gamma_{\theta_N}(2, 1)$  around  $X_{k'}$ , the expected length of all but the part of  $\mathcal{I}'$  built within  $(\tilde{I}_k)^c$  (note that for  $k' \leq N$ , a second exploration of the vertex requires an a priori active point on it, with  $X_{k'} \notin \tilde{I}_k$  then  $\mathcal{F}_{k'-1}$ -measurable). These precise terms appear also in (3.31) unless the link  $X_{k'}$  has been formed before step  $k$ , and we recall Remark 3.3 that at most  $\Delta_e(l) = O_{\mathbb{P}}(1)$  such terms may be missing from (3.31). In contrast, the  $\Delta_a(l)$  active points on explored vertices after step  $l$  do not contribute to the left sum of (3.33), while some may participate in (3.31). However, as we show in the sequel,

$$(3.34) \quad \mathbb{E}[\Delta_a(l)] \leq \frac{l}{N} \mathbb{E}[A_l], \quad \forall l \geq 1,$$

hence  $\Delta_a(l) = O_{\mathbb{P}}(1)$  uniformly in  $l \leq s_0 N^{2/3}$  (thanks to Proposition 3.2). Thus, as claimed, we have at most

$$(3.35) \quad \Delta_e(l) + \Delta_a(l) + 2v_{\geq 3}^l = O_{\mathbb{P}}(1),$$

differing terms between these two sums (recall Proposition 3.4 that  $v_{\geq 3}^l = O_{\mathbb{P}}(l^3/N^2) = O_{\mathbb{P}}(1)$  uniformly over  $l \leq s_0 N^{2/3}$ ).

Returning to establish (3.34), note that the fraction  $q_l$  of explored vertices after step  $l$ , never exceeds  $l/N$  and the average over vertices of the mean number of active points per vertex is  $\mathbb{E}[A_l]/N$ . Thus, the bound (3.34) holds if these mean numbers are the same across all vertices, or more generally, if they are tilted in favor of the nonexplored vertices. Further, it suffices to consider  $l \leq N$ , whereby only parts (a) and (b) of the rules for choosing the explored points are ever used; see Section 2.1. Exploring there the connections of  $I_k$ , one uses the same rate to each vertex  $i \neq w_k$ , before erasing some of the connections to explored vertices, while erasing none of those to nonexplored vertices. Beyond this reduction in  $\mathbb{E}[\Delta_a(l)]$  relative to  $\mathbb{E}[A_l]$ , and possibly having  $q_l < l/N$  (due to events  $\{E_{\geq 2}^k\}$ ), the only other deviation from uniformity is due to the nonrandom preference of choosing  $w_k$  according to its index. The latter can only cause the probability that  $w_k$  is an already explored vertex (which thereby reduces  $\Delta_a(k)$  by one), to exceed  $q_k$ . In conclusion, each of these effects merely tilts the mean number of active points per vertex, toward the nonexplored vertices, hence collectively they merely reinforce the inequality (3.34).

The other contribution to the  $O_{\mathbb{P}}(1)$  of (3.30) comes from the case of

$$(3.36) \quad \mathbf{1}_{(E_1^k \cap E_1^{k'} \cap \{k' \neq k\})^c} \leq \mathbf{1}_{\{k'=k\}} + \mathbf{1}_{E_{\geq 2}^k} + \mathbf{1}_{E_{\geq 2}^{k'}},$$

in the right-most sum of (3.33). Indeed, uniformly over  $l \leq s_0 N^{2/3}$ ,

$$\begin{aligned}
 \frac{1}{N} \sum_{1 \leq k \leq k' \leq l} (\mathbf{1}_{k'=k} + \mathbf{1}_{E_{\geq 2}^k} + \mathbf{1}_{E_{\geq 2}^{k'}}) &\leq \frac{2l}{N} (1 + \nu_{\geq 2}^l) \\
 (3.37) \qquad \qquad \qquad &= \frac{2l}{N} + O_{\mathbb{P}}(l^3 N^{-2}) = O_{\mathbb{P}}(1).
 \end{aligned}$$

Further, the event  $E_1^k \cap E_1^{k'}$  results for  $k' > k$  with  $w_{k'} \neq w_k$  and  $\tilde{I}_{k'}$  lying on a circle which is completely neutral (other than active points), after step  $(k' - 1)$ . So following our convention in making  $X_{k'}$  a  $\mathcal{F}_{k'-1}$ -measurable variable, this case contributes

$$\mathbb{E}(\tilde{I}_{k'} \cap \tilde{I}_k | \mathcal{F}_{k'-1}) = H(\tilde{I}_k - X_{k'}),$$

to the right-most sum of (3.33), thereby completing the proof.  $\square$

**PROOF OF PROPOSITION 3.7.** Plugging  $F(\theta_N) = \lambda_N(1 + aN^{-1/3})$ , we find that the RHS of (3.30) is

$$O_{\mathbb{P}}(1) + a l N^{-1/3},$$

which upon setting  $l = s N^{2/3}$  and scaling by  $N^{-1/3}$ , converges to  $as$  when  $N \rightarrow \infty$ , uniformly over  $s \leq s_0$ . Since  $\lambda_N \rightarrow \lambda_{\star} = F(\theta_{\star})$  (see (3.1)), we complete the proof of the proposition by way of Lemmas 3.10 and 3.11, which show that upon scaling by  $N^{-1/3}$ , for  $l = s N^{2/3}$  the terms in (3.32) and (3.31) converge in probability as  $N \rightarrow \infty$ , uniformly over  $s \leq s_0$ , to the appropriate nonrandom limits, respectively. Indeed, by the preceding we conclude that the processes  $s \mapsto \bar{B}_N(s)$  converge to the deterministic path  $s \mapsto \rho^{a, \theta_{\star}}(s)$  of (1.10), uniformly over  $s \leq s_0$ .  $\square$

**LEMMA 3.10.** *For  $(\theta_N, \lambda_N)$  satisfying (3.1) and  $l_N = s N^{2/3}$ , we have*

$$(3.38) \qquad \lim_{N \rightarrow \infty} \frac{1}{\binom{l_N}{2}} \sum_{k=1}^{l_N} \sum_{k'=k+1}^{l_N} H(\tilde{I}_k - X_{k'}) = \frac{F^2(\theta_{\star})}{\theta_{\star}},$$

*in probability, uniformly over  $s \in [\epsilon, s_0]$ .*

**LEMMA 3.11.** *For  $(\theta_N, \lambda_N)$  satisfying (3.1),  $l_N = s N^{2/3}$*

$$(3.39) \qquad \lim_{N \rightarrow \infty} \frac{2N}{l_N^2} \sum_{k=1}^{l_N} \sum_{i=1}^{r(k, l_N)} H_0(\tilde{I}_k - X_{(k, i)}) = F(\theta_{\star}) \left( 1 - \frac{F(\theta_{\star})}{\theta_{\star}} - \gamma(\theta_{\star}) \right),$$

*in probability, uniformly over  $s \in [\epsilon, s_0]$ .*

REMARK 3.12. In view of Lemma 3.8, the heuristic behind (3.38) is that for most pairs  $k' > k$  the conditional law of  $X_{k'}$  given  $\mathcal{F}_{k-}$  is nearly uniform. To see why we expect (3.39) to hold, let

$$(3.40) \quad R_l := \sum_{k=1}^l r(k, l),$$

be the aggregate number of links by the end of the  $l$ th step, so  $R_k - R_{k-1}$  merely counts the connections made during the  $k$ th step to all explored vertices, other than  $w_k$ . As such, conditional on  $\mathcal{F}_{k-}$ , the law of  $R_k - R_{k-1}$  is Poisson of rate  $|\tilde{I}_k|q_{k-}/\lambda_N$ , where  $q_{k-}$  denotes the fraction of explored vertices, other than  $w_k$ , by step  $k$ . With  $q_{k-} = k/N(1 + o(1))$  and  $F(\theta_N)/\lambda_N \rightarrow 1$ , the totality of these Poisson rates

$$(3.41) \quad V_l := \frac{1}{\lambda_N} \sum_{k=1}^l |\tilde{I}_k|q_{k-},$$

should be about  $l^2/(2N)$ . We further expect most positions  $X_{(k,i)}$  to be nearly uniform on  $\mathbb{S}_{\theta_N}^{w_k}$  and approximately independent of the first explored interval  $\tilde{I}_k$  on that vertex. Upon justifying these two approximations, we get (3.39) from the LLN for the empirical average of  $\text{UH}_0(\cdot)$  at the nearly i.i.d.  $|\tilde{I}_k|$ .

PROOF OF LEMMA 3.10. While proving (3.28), we saw that at  $\theta_N \rightarrow \theta_\star$ ,

$$\text{UH}(J) = \frac{F(\theta_N)}{\theta_N} J \rightarrow \frac{F(\theta_\star)}{\theta_\star} J.$$

Thus, by the LLN, if  $J_k := |\tilde{I}_k|$  are i.i.d.  $\Gamma_{2,1}(\theta_N)$  variables, then

$$(3.42) \quad \frac{1}{\binom{l_N}{2}} \sum_{k=1}^{l_N} \sum_{k'=k+1}^{l_N} \text{UH}(J_k) = \frac{2F(\theta_N)}{\theta_N(l_N - 1)} \sum_{k=1}^{l_N} \left(1 - \frac{k}{l_N}\right) J_k \rightarrow \frac{F^2(\theta_\star)}{\theta_\star},$$

in probability, uniformly over  $N^{-2/3}l_N \in [\epsilon, s_0]$ . This indeed is the joint law of  $\{J_k\}$  for the QRG, apart from possibly at most

$$v_{\geq 2}^{l_N} = O_{\mathbb{P}}(l_N^2/N) = o_{\mathbb{P}}(l_N)$$

values of  $k$ . Hence, with  $H(\cdot)$  uniformly bounded, the uniform convergence in probability of (3.42) extends to our setting. Recalling (3.27) and using the notation  $l_N := s_0 N^{2/3}$ ,

$$\hat{H}(\mathcal{I}) := H(\mathcal{I}) - \mathbb{E}[H(\mathcal{I} - U)|\mathcal{I}],$$

with  $U \in \mathbb{S}_{\theta_N}$  uniform, it suffices for proving the lemma, to show that

$$(3.43) \quad \frac{1}{l_N^4} \sum_{1 \leq k < k' \leq l_N} \sum_{1 \leq j < j' \leq l_N} |\mathbb{E}[\hat{H}(\tilde{I}_k - X_{k'}) \hat{H}(\tilde{I}_j - X_{j'})]| \rightarrow 0$$

Assuming WLOG that  $j' > r := j \vee k'$  (there are only  $O(\bar{l}_N^3)$  terms with  $j' = k'$ ), and noting that the uniformly bounded  $\widehat{H}(\tilde{I}_k - X_{k'})$  is  $\mathcal{F}_{r-}$ -measurable, it suffices for (3.43) to show that

$$(3.44) \quad \frac{1}{\bar{l}_N^2} \sum_{1 \leq r < r' \leq \bar{l}_N} h(r, r') \rightarrow 0,$$

where

$$(3.45) \quad h(r, r') := \max_{j \leq r} \{ \mathbb{E}[|\mathbb{E}[\widehat{H}(\tilde{I}_j - X_{r'}) | \mathcal{F}_{r-}]|] \}.$$

Further, with  $\pi_N$  denoting the uniform law on  $\mathbb{S}_{\theta_N}$  we have that for some  $C < \infty$  and all  $r, r', N$ ,

$$(3.46) \quad h(r, r') \leq C \mathbb{E}[d_{TV}(\pi_{r,r'}, \pi_N)], \quad \pi_{r,r'}(\cdot) := \mathbb{P}(X_{r'} \in \cdot | \mathcal{F}_{r-}).$$

In view of (3.53) of Lemma 3.13, plugging this bound into (3.44) completes the proof of Lemma 3.10.  $\square$

**PROOF OF LEMMA 3.11.** Recalling that if  $J_k := |\tilde{I}_k|$  is  $\Gamma_{2,1}(\theta_N)$  variable then  $\mathbb{E}J_k = F(\theta_N) = \lambda_N(1 + o(1))$ , it follows from the LLN by the same argument as in our proof of Lemma 3.10, that the following proxy

$$V_l^* := \frac{1}{\lambda_N} \sum_{k=1}^l J_k \frac{k}{N},$$

for  $V_l$  of (3.41) is such that uniformly over  $s \in [\epsilon, s_0]$ ,

$$\mathbb{E} \left[ \left| \frac{2N}{l_N^2} V_{l_N}^* - 1 \right| \right] \rightarrow 0.$$

Moreover, for any  $k \leq l_N$ ,

$$\frac{k}{N} - q_{k-} \in [0, N^{-1} \nu_{\geq 2}^{l_N}] = O_{\mathbb{P}}(N^{-2/3}).$$

Hence,  $\mathbb{E}[|V_{l_N}^* - V_{l_N}|] \leq O_{\mathbb{P}}(1)$  and consequently, uniformly over  $s \in [\epsilon, s_0]$ ,

$$(3.47) \quad \mathbb{E} \left[ \left| \frac{2N}{l_N^2} V_{l_N} - 1 \right| \right] \rightarrow 0.$$

As explained in Remark 3.12,  $M_k := R_{k-1} - V_{k-1}$  is an  $\mathcal{F}_{k-}$  martingale and the corresponding predictable part in Doob's decomposition for  $M_k^2$  is precisely  $\langle M \rangle_k = V_{k-1}$ . Thus, by (3.47), uniformly over  $s$  as above,

$$\mathbb{E} \left[ \left( \frac{2N}{l_N^2} M_{l_N} \right)^2 \right] \rightarrow 0,$$

hence by yet another application of (3.47), also

$$(3.48) \quad \mathbb{E} \left[ \left| \frac{2N}{l_N^2} R_{l_N} - 1 \right| \right] \rightarrow 0.$$

Consider the successive steps  $\varpi_1 \leq \varpi_2 \leq \dots \leq \varpi_j \leq \dots$  in which links of type  $\{X_{(k,i)}, k, i \geq 1\}$  to previously explored vertices, say  $X_j^* := X_{(k,i)}$ , are formed, with the induced stopped-filtration  $\{\mathcal{F}_{\varpi_j}\}$ . We can ignore here ties (namely,  $\varpi_{j+1} = \varpi_j$ ), since the expected number of such is bounded uniformly in  $N$  and  $l \leq s_0 N^{2/3}$  (indeed,  $R_k - R_{k-1}$  which is at most a Poisson of rate  $\theta_N k / (\lambda_N N)$ , yields  $(R_k - R_{k-1} - 1)_+$  ties). Thanks to (3.48), it suffices for (3.39) to show that for  $\ell_N = tN^{1/3}$ ,

$$(3.49) \quad \lim_{N \rightarrow \infty} \frac{1}{\ell_N} \sum_{j=1}^{\ell_N} H_0(\tilde{I}_{k(j)} - X_j^*) = F(\theta_\star) \left( 1 - \frac{F(\theta_\star)}{\theta_\star} - \gamma(\theta_\star) \right) \quad \text{in } \mathbb{P},$$

uniformly over  $t \in [\epsilon', t_0]$ , where  $k(j)$  denotes the first exploration step of the vertex on which  $X_j^*$  lies. Further, for at most  $v_{\geq 3}^{\tilde{I}_N} = O_{\mathbb{P}}(1)$  of the links  $\{X_j^*, j \leq \ell_N\}$ , the vertex  $w_{k(j)}$  appears more than once in this collection. Consequently, appealing to the RHS of (3.28), as in the proof of Lemma 3.10 we get analogously to (3.42) that as  $N \rightarrow \infty$ , uniformly over  $t$  as above,

$$\frac{1}{\ell_N} \sum_{j=1}^{\ell_N} \text{UH}_0(|\tilde{I}_{k(j)}|) \rightarrow F(\theta_\star) \left( 1 - \frac{F(\theta_\star)}{\theta_\star} - \gamma(\theta_\star) \right) \quad \text{in } \mathbb{P}.$$

Thus, similar to the proof of Lemma 3.10, setting  $\bar{\ell}_N = t_0 N^{1/3}$  and

$$\hat{H}_0(\mathcal{I}) := H_0(\mathcal{I}) - \text{UH}_0(|\mathcal{I}|),$$

we get (3.49) as soon as we show that

$$(3.50) \quad \frac{1}{\bar{\ell}_N^2} \sum_{1 \leq j < j' \leq \bar{\ell}_N} |\mathbb{E}[\hat{H}_0(\tilde{I}_{k(j)} - X_j^*) \hat{H}_0(\tilde{I}_{k(j')} - X_{j'}^*)]| \rightarrow 0.$$

With  $\hat{H}_0(\tilde{I}_{k(j)} - X_j^*)$  uniformly bounded and  $\mathcal{F}_{\varpi_j}$ -measurable, this in turn follows from

$$(3.51) \quad \frac{1}{\bar{\ell}_N} \sum_{j=1}^{\bar{\ell}_N} \sum_{j'=j+1}^{\bar{\ell}_N} \mathbb{E}[|\mathbb{E}[\hat{H}_0(\tilde{I}_{k(j')} - X_{j'}^*) | \mathcal{F}_{\varpi_j}]|] \rightarrow 0.$$

Further, setting  $r = r(j, j') := \varpi_j \vee k(j')$ ,

$$(3.52) \quad \pi_{j,j'}^*(\cdot) := \mathbb{P}(X_{j'}^* \in \cdot | \mathcal{F}_{r(j,j')}),$$

and recalling the definition (3.27) of  $\text{UH}_0(\cdot)$  with  $U \sim \pi_N$  on  $\mathbb{S}_{\theta_N}$ , we have that for some  $C < \infty$  and all  $j, j', N$ ,

$$\mathbb{E}[|\mathbb{E}[\hat{H}_0(\tilde{I}_{k(j')} - X_{j'}^*) | \mathcal{F}_{\varpi_j}]|] \leq C \mathbb{E}[d_{\text{TV}}(\pi_{j,j'}^*, \pi_N)].$$

In view of Lemma 3.14, plugging this into (3.51) completes the proof.  $\square$

We next complete the proof of Lemma 3.10 by showing that for most  $r < r' \leq \bar{l}_N$ , the law of  $X_{r'}$  given  $\mathcal{F}_{r-}$  is nearly uniform (in total-variation distance).

LEMMA 3.13. *Setting  $\bar{l}_N = s_0 N^{2/3}$  we have for  $\pi_{r,r'}$  of (3.46), that*

$$(3.53) \quad \lim_{N \rightarrow \infty} \frac{1}{\bar{l}_N^2} \sum_{r=1}^{\bar{l}_N} \sum_{r'=r+1}^{\bar{l}_N} \mathbb{E}[d_{\text{TV}}(\pi_{r,r'}, \pi_N)] = 0.$$

PROOF. Fixing  $r \geq 1$  we assign to each active point at the end of the  $r$ th step, and to any link formed thereafter (including a link onto the previously explored part of the QRG), a proxy counter for its uniformity and  $\mathcal{F}_r$ -independence, as follows. First, the counter of each active point at the end of the  $r$ th step is set to 0 if it is on an unexplored vertex and to  $-1$  otherwise. Then, sequentially in  $k > r$ , if  $A_{k-1} > 0$  and the vertex  $w_k$  was not previously explored, we set as  $n_{r,k}$  the counter of the active point  $(w_k, X_k)$ , if  $A_{k-1} = 0$  we set  $n_{r,k} = \infty$ , and otherwise let  $n_{r,k} = -1$ . Thereafter, each link formed during the  $k$ th step gets a counter value  $n_{r,k} + 1$ . We claim that for  $\rho := \sup_N \{\mathbb{P}(\Gamma_{\theta_N}(2, 1) < \theta_N)\} < 1$  and any  $r' \in (r, N)$ ,

$$(3.54) \quad d_{\text{TV}}(\mathbb{P}(X_{r'} \in \cdot | \mathcal{F}_{r-}, n_{r,r'} \geq c), \pi_N) \mathbb{P}(n_{r,r'} \geq c | \mathcal{F}_{r-}) \leq \rho^c.$$

Indeed, at each step we choose the value of  $X_k$  independently of the positions of active points within their respective circles. Having  $A_{k-1} = 0$  yields  $X_k \sim \pi_N$  independently of  $\mathcal{F}_{k-1}$  and this property is inherited by any point to which the path from  $X_k$  involves only first explorations of the relevant vertices. Further, the  $\mathcal{F}_{k-1}$ -measurable event  $n_{r,k} \geq 0$  (namely  $E_1^k$ ), results with  $\mathbb{P}(\tilde{I}_k = \mathbb{S}_{\theta_N} | \mathcal{F}_{k-1}) \geq 1 - \rho$  and  $\tilde{I}_k = \mathbb{S}_{\theta_N}$  yields by  $\mathcal{L}_{w_k,i}$  uniformly distributed links to all vertices (prior to the erasures on previously explored space). In case  $n_{r,r'} \geq c$  is finite, one has during  $(r + 1, r')$  at least  $c$  consecutive forefathers of  $(w_{r'}, X_{r'})$  in our exploration tree, all of whom were explored on circles which are neutral (apart from active points). The chance that none of these forefathers forced a uniform conditional law of  $X_{r'}$ , is by the preceding at most  $\rho^c$ , thereby establishing (3.54).

Now, thanks to (3.54) and the convexity of the  $[0, 1]$ -valued  $d_{\text{TV}}(\cdot, \pi_N)$ , it suffices for (3.53) to show that for any  $c < \infty$ ,

$$(3.55) \quad \lim_{N \rightarrow \infty} \mathbb{E}[\Gamma_c(\bar{l}_N)] = 0, \quad \text{where } \Gamma_c(l) := l^{-2} \sum_{1 \leq r < r' \leq l} \mathbf{1}_{\{n_{r,r'} < c\}}.$$

To this end, with  $\mathcal{Z}_c(k')$  denoting the size of the exploration subtree of depth at most  $c$ , rooted at the active point  $X_{k'}$  and  $\Theta(r)$  enumerating those  $k' > r$  for which the link to  $(w_{k'}, X_{k'})$  has been formed before the end of the  $r$ th step, we claim that

$$(3.56) \quad \Gamma_c(l) \leq l^{-2} \sum_{1 \leq r < k' \leq l} \mathcal{Z}_c(k') \mathbf{1}_{\{k' \in \Theta(r)\}} + l^{-1} \sum_{k'=1}^l \mathcal{Z}_c(k') \mathbf{1}_{E_{\geq 2}^{k'}}.$$

Indeed, to have  $n_{r,r'} < c$ , one of the  $c$  consecutive exploration forefathers of  $(w_{r'}, X_{r'})$ , say  $(w_{k'}, X_{k'})$ , must have been an event  $E_{\geq 2}^{k'}$  (namely, not a first exploration), or alternatively, be an active point formed before the end of the  $r$ th step. In the latter case, considering the *last such step* (i.e., along the path to  $(w_{r'}, X_{r'})$  and among active points formed by the end of the  $r$ th step), guarantees that  $r < k'$  in the first sum of (3.56). Thereafter,  $\mathcal{Z}_c(k')$  bounds the number of possible pairs  $(k', r')$  having path distance at most  $c$  and should no such  $(k', r')$  exist, the second sum on the RHS of (3.56) bounds the number of  $r' \leq l$  with some previously explored vertex among the last  $c$  steps on the exploration path to  $(w_{r'}, X_{r'})$ .

Now, for any  $k'$ , conditional on  $\mathcal{F}_{k'^-}$  the variable  $\mathcal{Z}_c(k')$  is stochastically dominated by the size of a Galton–Watson tree of depth  $c$  and a Poisson( $\mu$ ) off-spring law, for  $\mu := \max_N \{\theta_N / \lambda_N\}$  finite. It thus follows that for some  $\kappa_c = \kappa_c(\mu)$  finite and all  $N$ ,

$$\max_{k' \geq 1} \mathbb{E}[\mathcal{Z}_c(k') | \mathcal{F}_{k'^-}] \leq \kappa_c.$$

Equipped with the latter bound, upon considering the expected values in (3.56), since both  $E_{\geq 2}^{k'}$  and  $\{k' \in \Theta(r)\}$  are in  $\mathcal{F}_{k'^-}$ , while  $|\Theta(r)| \leq A_r$ , it follows by the tower property of the conditional expectation, that

$$(3.57) \quad \mathbb{E}[\Gamma_c(l)] \leq \frac{\kappa_c}{l} \left( \mathbb{E}[\max_{r < l} \{A_r\}] + \mathbb{E} \left[ \sum_{k'=1}^l \mathbf{1}_{E_{\geq 2}^{k'}} \right] \right).$$

Further, recall that  $\sum_{k' \leq l} \mathbf{1}_{E_{\geq 2}^{k'}} \leq v_{\geq 2}^l$ . Thus, with both  $l^{-1} \mathbb{E}[\max_{r < l} A_r]$  and  $l^{-1} \mathbb{E}[v_{\geq 2}^l]$  decaying to zero at  $l = \bar{l}_N$  and  $N \rightarrow \infty$  (due to Proposition 3.2 and (3.13), resp.), the bound (3.57) yields that (3.55) holds.  $\square$

Similar to Lemma 3.13 we complete the proof of Lemma 3.11 by showing that for most  $j < j' \leq \bar{\ell}_N$  the conditional law of  $X_{j'}^*$  is nearly uniform.

LEMMA 3.14. *Setting  $\bar{\ell}_N = t_0 N^{1/3}$ , we have for  $\pi_{j,j'}^*$  of (3.52), that*

$$(3.58) \quad \lim_{N \rightarrow \infty} \frac{1}{\bar{\ell}_N^2} \sum_{j=1}^{\bar{\ell}_N} \sum_{j'=j+1}^{\bar{\ell}_N} \mathbb{E}[d_{\text{TV}}(\pi_{j,j'}^*, \pi_N)] = 0.$$

PROOF. We record the first exploration step  $\text{ex}(w) \geq 1$  of each vertex  $w \in [1, N]$ , so upon forming a link of type  $X_{j'}^*$  onto a vertex  $v_{j'}$  one has that  $k(j') = \text{ex}(v_{j'})$ . Utilizing the proxy counters of Lemma 3.13, let  $n_{j,j'}^*$  denote the value of the counter for the link  $X_{j'}^*$  starting at step

$$r = r(j, j') := \varpi_j \vee \text{ex}(v_{j'}).$$

By the same reasoning as in the derivation of (3.54), we have that

$$(3.59) \quad d_{TV}(\mathbb{P}(X_{j'}^* \in \cdot | \mathcal{F}_r, n_{j,j'}^* \geq c), \pi_N) \mathbb{P}(n_{j,j'}^* \geq c | \mathcal{F}_r) \leq \rho^c.$$

We thus get (3.58) by establishing the analog of (3.55). That is, upon showing that for  $l := 2\sqrt{\ell N}$  and any  $c < \infty$ ,

$$(3.60) \quad \lim_{N \rightarrow \infty} \mathbb{E}[\Gamma_c^*(\bar{\ell}_N)] = 0, \quad \text{for } \Gamma_c^*(\ell) := \ell^{-2} \mathbf{1}_{\{\varpi_\ell \leq l\}} \sum_{1 \leq j < j' \leq \ell} \mathbf{1}_{\{n_{j,j'}^* < c\}}$$

(by (3.48) it suffices to consider  $2N\ell/\varpi_\ell^2 \rightarrow 1$ , hence the restriction here to  $\varpi_\ell \leq 2\sqrt{\ell N}$ ). Next, for  $k < k'$ , let  $\mathcal{Z}_c(k'; k) = \mathcal{Z}_c(k')$  if the exploration subtree of depth at most  $c$  rooted at the active point  $X_{k'}$ , has a link to  $\mathbb{S}_{\theta_N}^{w_k}$ , otherwise setting  $\mathcal{Z}_c(k'; k) \equiv 0$ . Likewise,  $\mathcal{Z}_c^*(k') := \max_{k < k'} \{\mathcal{Z}_c(k'; k)\}$  is the nonzero  $\mathcal{Z}_c(k')$  iff the relevant subtree produces a link to some previously explored vertex. Setting hereafter  $k = k(j')$ , recall that  $n_{r,r'} < c$  requires that one of the  $c$  consecutive exploration forefathers of the link  $X_{j'}^*$ , say  $(w_{k'}, X_{k'})$ , must have been an event  $E_{\geq 2}^{k'}$  (namely, not a first exploration), or alternatively, be an active point formed before the end of the  $r(j, j')$ th step. We thus claim, similar to (3.56), that

$$(3.61) \quad \begin{aligned} \Gamma_c^*(\ell) &\leq \ell^{-2} \sum_{j < \ell, \varpi_j < k' \leq l} \mathcal{Z}_c^*(k') \mathbf{1}_{\{k' \in \Theta(\varpi_j)\}} \\ &\quad + \ell^{-1} \sum_{1 \leq k < k' \leq l} \mathcal{Z}_c(k'; k) \mathbf{1}_{\{k' \in \Theta(k)\}} \\ &\quad + \ell^{-1} \sum_{k'=1}^l \mathcal{Z}_c^*(k') \mathbf{1}_{E_{\geq 2}^{k'}}. \end{aligned}$$

The first two expressions on the RHS distinguish having  $k < \varpi_j$  from the case of  $k \geq \varpi_j$ , where we sum over  $j \leq \ell$  and cover all choices of  $j'$  by the additional sum over  $k = k(j') < l$ . As done on the RHS of (3.56), in both expressions we guarantee that  $r < k'$  by having  $(w_{k'}, X_{k'})$  stand for the last active point on the path to  $X_{j'}^*$  among those formed by the end of the  $r$ th step. The link  $X_{j'}^*$  must lie on the vertex  $w_{k(j)}$ , yielding the bound  $\mathcal{Z}_c(k'; k)$  in case  $k = r < k'$  with an exploration path distance at most  $c$  from  $X_{k'}$  to  $X_{j'}^*$ . However, in case  $k < \varpi_j = r$  we do not keep track of  $k$ , hence must replace  $\mathcal{Z}_c(k'; k)$  by the larger  $\mathcal{Z}_c^*(k')$  which only indicates the existence of a point of exploration path distance at most  $c$  from  $X_{k'}$  which is on a vertex that was first explored prior to step  $k'$ . Finally, should no active point  $(w_{k'}, X_{k'})$  with  $k' > k$  of exploration path distance at most  $c$  from  $X_{j'}^*$  be formed by the end of the  $r$ th step, the last sum on the RHS of (3.61) bounds the number of  $j' \leq \ell$  (with  $\varpi_{j'} \leq l$ ), having a nonneutral circle (explored at some step  $k' > k$ ), among the last  $c$  points on the path to  $X_{j'}^*$ .

Next recall that  $\mathcal{Z}_c(k')$  is, conditionally on  $\mathcal{F}_{k'-}$ , stochastically dominated by the size of a depth  $c$  Galton–Watson tree of a  $\text{Poisson}(\mu)$  off-spring law. The latter size

variable has finite moments of all order, whereas  $\mathcal{Z}_c(k'; k)$  further demands having at least one tree vertex corresponding to the prescribed  $w_k$ . With the production of the specified vertex  $w_k$  stochastically dominated by a Poisson of rate  $\mu/N$ , we have for some  $\kappa_c^*$  finite and all  $N$ ,

$$\max_{k < k'} \{ \mathbb{E}[\mathcal{Z}_c(k'; k) | \mathcal{F}_{k'-}] \} \leq \frac{\kappa_c^*}{N}, \quad \max_{k' \leq l} \{ \mathbb{E}[\mathcal{Z}_c^*(k') | \mathcal{F}_{k'-}] \} \leq \frac{\kappa_c^* l}{N}.$$

Now, analogously to the derivation of (3.57), upon taking the expectation on both sides of (3.61), we get by the tower property and the preceding estimates that

$$\mathbb{E}[\Gamma_c^*(\ell)] \leq \frac{\kappa_c^* l}{N\ell} \left( \mathbb{E} \left[ \ell^{-1} \sum_{j < \ell} A_{w_j} \mathbf{1}_{\{w_j \leq l\}} \right] + \mathbb{E} \left[ l^{-1} \sum_{k' < l} A_{k'} \right] + \mathbb{E} \left[ \sum_{k'=1}^l \mathbf{1}_{E_{\geq 2}^{k'}} \right] \right).$$

Utilizing the fact that  $l/(N\ell) = 4/l$ , we arrive at the same bound as in the RHS of (3.57). Setting  $\ell = \bar{\ell}_N$  corresponds to having  $l = \bar{l}_N$ , thus yielding (3.60) by the reasoning provided at the end of the proof of Lemma 3.13.  $\square$

*3.2. Joint convergence of component sizes.* Recall the statement of Theorem 1.3. In this section, we shall conclude the proof of this theorem using results from previous sections.

As pointed out in [2], Theorem 1.3 primarily has two parts:

1. First is to prove that the excursions of the limit process are matched by the excursions of the breadth first random walk.
2. The second is to arrange these excursions in the decreasing order. This can be achieved if one can ascertain that there exists a random point after which one is sure (with high probability) not to see large excursions.

In order to settle the first issue, we shall invoke [2], Lemmas 7 and 8, which can be applied verbatim to our case, together with Proposition 3.1 proved in a previous subsection.

Thus, we only need to be concerned about the second issue, for which we shall need to prove an appropriate version of [2], Lemma 9, suited to our case.

Like in [2], let us define

$$T(y) = \min\{s : W^{a, \theta^*}(s) = -y\},$$

$$T_N(y) = \min\{i : Y_N(i) = -\lfloor yn^{1/3} \rfloor\}.$$

Notice that as a consequence of Proposition 3.1

$$N^{-2/3} T_N(y) \rightarrow_d T(y).$$

Therefore, the following lemma completes the proof.

LEMMA 3.15. *Let us denote by  $p(N, y, \delta)$  the probability that the QRG with the parameters  $(\theta_N, \lambda_N)$  that satisfy (3.1), contains a component of size at least  $\delta N^{2/3}$  which does not contain any vertex  $i$  with  $1 \leq i \leq yN^{1/3}$ . Then*

$$\lim_{y \rightarrow \infty} \limsup_{N \rightarrow \infty} p(N, y, \delta) = 0 \quad \text{for all } \delta > 0.$$

PROOF. Fix  $\delta > 0$ . Let  $v_{C_i}$  be the minimal vertex of the component  $C_i$ , then for an interval  $\mathcal{J} \subset \mathbb{R}_+$ , define

$$(3.62) \quad q(N, \mathcal{J}) = \mathbb{E} \left( \sum_{i \geq 1} \mathbf{1}_{(|C_i| \geq \delta N^{2/3}; v_{C_i} \in N^{1/3} \mathcal{J})} \right).$$

Conditioned on arranging the components in a decreasing order of their sizes, the labels of the vertices of any given fixed component  $C_i$  are going to be uniformly randomly ordered. Given such components ordering, define

$$\chi_N(C_i) = N^{-1/3} v_{C_i}$$

and

$$\mathcal{U}_{C_i} = N^{-2/3} \quad (\text{number of vertices in the component } C_i).$$

Then note that for any  $x \geq 0$

$$\mathbb{P}(v_{C_i} > N^{1/3} x | \mathcal{U}_{C_i}) = \left( 1 - \frac{\mathcal{U}_{C_i} N^{2/3}}{N} \right)^{N^{1/3} x},$$

implying

$$(3.63) \quad \mathbb{P}(\chi_N(C_i) > y | \mathcal{U}_{C_i}) \leq \frac{e^{-\mathcal{U}_{C_i} y}}{1 - e^{-\mathcal{U}_{C_i}}} \mathbb{P}(\chi_N(C_i) \leq 1 | \mathcal{U}_{C_i}).$$

Further

$$(3.64) \quad \mathbb{P}(v_{C_i} \in [yN^{1/3}, \infty)) = \mathbb{E}(\mathbb{P}(v_{C_i} \in [yN^{1/3}, \infty) | \mathcal{U}_{C_i})).$$

At this point, conditional on component sizes being  $|C_i| = bN^{2/3}$ , we note that one can adopt the proof of Proposition 3.4 to the original exploration process restricted to the construction of  $C_i$  in order to derive similar results for  $v_{\geq 3}^{C_i}$ , the number of explored intervals belonging to  $C_i$  sampled by the end of the construction of  $C_i$ , which belong to vertices (circles), having at least three such intervals each. Then, observing that  $\mathcal{U}_{C_i} N^{2/3} \geq \frac{1}{2}(bN^{2/3} - v_{\geq 3}^{C_i})$ , we have for  $\varepsilon > 0$ ,

$$\mathbb{P} \left( \mathcal{U}_{C_i} \geq \frac{1}{2}b - \frac{1}{2} \frac{N^{1/3+\varepsilon}}{N^{2/3}} \right) \geq \mathbb{P}(v_{\geq 3}^{C_i} \leq N^{1/3+\varepsilon}) = 1 - o(N^{-1/3}),$$

implying that  $\mathcal{U}_{C_i} \in (\frac{b}{3}, b)$  with probability  $(1 - o(N^{-1/3}))$ . Consequently,

$$\mathbb{E}(\mathbb{P}(v_{C_i} \in [yN^{1/3}, \infty) | \mathcal{U}_{C_i})) \leq \frac{e^{-by/3}}{1 - e^{-b/3}} \mathbb{P}(v_{C_i} \in [0, N^{1/3}]) + o(N^{-1/3}).$$

Recalling the definition of  $q(N, \mathfrak{J})$  from (3.62), and conditioning on the number  $M_b$  of components of size  $bN^{2/3}$ , while observing that given the sizes of components the minimal vertices of various different components are identically distributed, we get

$$q(N, [y, \infty)) = \mathbb{E} \left[ \sum_{b=\delta}^{\infty} M_b \mathbb{P}(v_{\mathcal{C}_i} \in [yN^{1/3}, \infty) \mid |\mathcal{C}_i| = bN^{2/3}) \right].$$

With  $\sum_{b \geq \delta} M_b \leq \delta^{-1} N^{1/3}$ , by the preceding, this and (3.63) imply that

$$q(N, [y, \infty)) \leq \frac{e^{-\delta y/3}}{1 - e^{-\delta/3}} q(N, [0, 1]) + o(1).$$

Since  $p(N, y, \delta) \leq q(N, [y, \infty))$ , to prove the theorem, it suffices to show that

$$(3.65) \quad \sup_N q(N, [0, 1]) < \infty,$$

Writing  $t_i(v)$  as points on the  $v$ th vertex around which intervals are constructed and explored, and denoting  $\mathcal{N}(v)$  as the number of such points we observe that

$$q(N, [0, 1]) \leq \sum_{v=1}^{N^{1/3}} \mathbb{E} \left( \sum_{i=1}^{\mathcal{N}(v)} \mathbf{1}_{\{|\mathcal{C}(t_i(v))| > \delta N^{2/3}\}} \right),$$

where  $\mathcal{C}(t_i(v))$  is the maximal connected component containing  $t_i(v)$ .

Clearly, the collection  $\{t_1(v), \dots, t_{\mathcal{N}(v)}(v)\}$  is independent and identically distributed for different  $v \in G_N$ . We replace the exploration by the overcounting process of Section 2.2 which is coupled with the exploration process until the exploration process hits zero. Then we restart an independent (and identical) overcounting process together with restarting the exploration process. We repeat this process until the end of exploration of the complete graph. Subsequently, setting  $\{t_1^*(v), \dots, t_{\mathcal{N}^*(v)}^*(v)\}$ ,  $\mathcal{N}^*(v)$  and  $\mathcal{C}^*(t_i^*(v))$  as the corresponding elements of the overcounting process, we observe that since  $|\mathcal{C}(t_i^*(v))|$  are i.i.d. we have

$$\begin{aligned} q(N, [0, 1]) &\leq \sum_{v=1}^{N^{1/3}} \mathbb{E} \left( \sum_{i=1}^{\mathcal{N}^*(v)} \mathbf{1}_{\{|\mathcal{C}(t_i^*(v))| > \delta N^{2/3}\}} \right) \\ &= N^{1/3} \mathbb{E}(\mathcal{N}^*(v)) \mathbb{P}(|\mathcal{C}(t_i^*(v))| > \delta N^{2/3}), \end{aligned}$$

where we have used Wald's equality.

Therefore, it suffices to prove that  $N^{1/3} \mathbb{P}(|\mathcal{C}_0^*| \geq \delta N^{2/3})$  is bounded by a constant where  $\mathcal{C}_0^*$  is a typical component of the overcounting process.

We now define the coupled overcounting process via i.i.d. random variables  $\xi_k^w$ , where each  $\xi_k^w$  represents the number of links generated at  $k$ th time step by the overcounting process with the parameters  $\lambda_N$  and  $\theta_N$  lying in the critical window (3.1), Unlike Section 2.2, here  $\xi_i^w \sim \text{Poisson}(\frac{\theta_N}{\lambda_N})$ . Then define  $S_k^w =$

$S_{k-1}^w + (\xi_k^w - 1)$ , with  $S_0^w = 1$ . Setting  $\tau^w = \min\{k \geq 1 : S_k^w = 0\}$ , it suffices to show that  $N^{1/3}\mathbb{P}(\tau^w > \delta N^{2/3})$  is bounded by a universal constant. Using the same arguments as used in Section 2.2, we conclude that

$$(3.66) \quad \mathbb{P}(\tau = n + 1) \leq \frac{\mathbb{E}(S_1^w)}{n} \sup_{\ell} \{\mathbb{P}(S_{n+1}^w - S_1^w = -\ell)\}$$

Using [16], Proposition 2.4.4, observe that

$$\mathbb{P}(S_{n+1}^w - S_1^w = -\ell) \leq \frac{c}{n^{1/2}}.$$

Therefore,

$$\mathbb{P}(\tau = n + 1) \leq cn^{-3/2}.$$

Subsequently, following the same arguments as in Section 2.2, we conclude that

$$N^{1/3}\mathbb{P}(|\mathcal{C}_0^*| \geq \delta N^{2/3}) \leq c,$$

for some finite  $c = c(\delta)$ , thereby proving the statement of the lemma.  $\square$

### APPENDIX: PROOF OF PROPOSITION 3.2

With  $Y_N = Y_N^{a, \theta_N}$ ,  $Z_N = Z_N^{a, \theta_N}$  and writing  $\iota(l)$  for the number of maximal connected components in the corresponding graph completely explored before step  $l$ , we use the relations

$$(A.1) \quad A_l = Y_N(l) + \iota(l) = Z_N(l) - \text{sur}(l) + \iota(l),$$

$$(A.2) \quad \iota(l) = 1 - \min_{0 \leq k \leq (l-1)} \{Z_N(k) - \text{sur}(k)\},$$

and the fact that  $k \mapsto \text{sur}(k)$  is nondecreasing, to find that

$$\begin{aligned} A_l &= 1 + Z_N(l) - \text{sur}(l) + \max_{k \leq (l-1)} \{\text{sur}(k) - Z_N(k)\} \\ &\leq 1 + \max_{k \leq l} \{Z_N(l) - Z_N(k)\}, \end{aligned}$$

which can further be simplified to write

$$A_l \leq 1 + 2 \max_{k \leq l} |Z_N(k)|$$

Recall the martingale decomposition,  $Z_N(k) = M_N(k) + B_N(k)$ , where  $M_N$  is a martingale and  $B_N$  is the predictable process. Then, for any fixed positive  $K$ , set

$$\Upsilon_N = \min\{k : |Z_N(k)| > KN^{1/3}\} \wedge (sN^{2/3}).$$

By Markov's inequality, it thus suffices for Proposition 3.2 to show that

$$(A.3) \quad \mathbb{E}[|Z_N(\Upsilon_N)|^2] = O(N^{2/3}).$$

To this end, using the notation introduced in (3.16), clearly

$$(A.4) \quad \mathbb{E}[|Z_N(\Upsilon_N)|^2] \leq 2\mathbb{E}[|M_N(\Upsilon_N)|^2] + 2\mathbb{E}[|B_N(\Upsilon_N)|^2].$$

Further, by Doob’s optional sampling theorem,

$$\mathbb{E}(M_N(\Upsilon_N)^2) = \mathbb{E}(D_N(\Upsilon_N)) \leq \mathbb{E}(D_N(sN^{2/3})),$$

since  $M_N^2$  is a submartingale and  $\Upsilon_N \leq sN^{2/3}$ . Now, recall that

$$D_N(sN^{2/3}) = \sum_{k=1}^{sN^{2/3}} \text{var}(\zeta_k | \mathcal{F}_{k-1})$$

with by way of (3.21) for uniformly bounded  $\varphi_k \leq \theta_N/\lambda_N$ , has expected value bounded by  $cN^{2/3}$ , for some finite  $c = c(a, s)$ . Turning to show the same for  $\mathbb{E}|B_N(\Upsilon_N)|^2$ , recall that the right sum of (3.33) has  $l(l+1)/2$  terms, each bounded by  $\theta_N/N$ , whereas to the left sum only the at most  $v_{\geq 2}^l$  events  $E_{\geq 2}^{k'}$  contribute (no more than  $2\theta_N$  each). Thus, in view of (3.1) and (3.33),

$$(A.5) \quad B_N(l) \leq 1 + \frac{al}{N^{1/3}} + \frac{\theta_N l^2}{\lambda_N N} + \frac{2\theta_N}{\lambda_N} v_{\geq 2}^l.$$

For  $l = \Upsilon_N \leq sN^{2/3}$ , the nonrandom part of the RHS of (A.5) is at most  $cN^{1/3}$ . Next, upon examining the argument leading to (3.13), we deduce that  $v_{\geq 2}^l$  is stochastically dominated by the sum of at most  $v_{\geq 1}^l$  i.i.d. Poisson variables of rate  $\kappa' l N^{-1}$  each. Hence, for some  $C$  finite and all  $l, N$ ,

$$\mathbb{E}[(v_{\geq 2}^l)^2] \leq C + C \frac{l^2}{N^2} \mathbb{E}[(v_{\geq 1}^l)^2].$$

Similar refinement in the argument leading to (3.9), yields that

$$\mathbb{E}[(v_{\geq 1}^l)^2] \leq Cl^2,$$

hence  $\mathbb{E}[(v_{\geq 2}^l)^2] \leq cN^{2/3}$  for  $l \leq sN^{2/3}$ . Such bound holds for  $\mathbb{E}|B_N(\Upsilon_N)|^2$  and the decomposition (A.4) yields (A.3) (thereby completing the proof).

**Acknowledgments.** This work benefited from helpful discussions of A. L. with Omer Angel and from the hosting of A. L. and A. D. by the Mathematical Sciences Research Institute (as part of its program on random spatial processes). The authors would also like to thank Júlia Komjáthy for useful comments.

### REFERENCES

[1] AIZENMAN, M., KLEIN, A. and NEWMAN, C. M. (1993). Percolation methods for disordered quantum Ising models, phase transitions. In *Mathematics, Physics, Biology...* (R. Kotecky, ed.). **124** 1–26. World Scientific, Singapore.

- [2] ALDOUS, D. (1997). Brownian excursions, critical random graphs and the multiplicative coalescent. *Ann. Probab.* **25** 812–854. [MR1434128](#)
- [3] BHAMIDI, S., VAN DER HOFSTAD, R. and VAN LEEUWAARDEN, J. S. H. (2010). Scaling limits for critical inhomogeneous random graphs with finite third moments. *Electron. J. Probab.* **15** 1682–1703. [MR2735378](#)
- [4] BHAMIDI, S., VAN DER HOFSTAD, R. and VAN LEEUWAARDEN, J. S. H. (2012). Novel scaling limits for critical inhomogeneous random graphs. *Ann. Probab.* **40** 2299–2361. [MR3050505](#)
- [5] BOLLOBÁS, B. (1984). The evolution of random graphs. *Trans. Amer. Math. Soc.* **286** 257–274. [MR0756039](#)
- [6] BOLLOBÁS, B., JANSON, S. and RIORDAN, O. (2007). The phase transition in inhomogeneous random graphs. *Random Structures Algorithms* **31** 3–122. [MR2337396](#)
- [7] CAMPANINO, M., KLEIN, A. and PEREZ, J. F. (1991). Localization in the ground state of the Ising model with a random transverse field. *Comm. Math. Phys.* **135** 499–515. [MR1091575](#)
- [8] ERDŐS, P. and RÉNYI, A. (1960). On the evolution of random graphs. *Magy. Tud. Akad. Mat. Kut. Intéz. Közl.* **5** 17–61. [MR0125031](#)
- [9] ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes: Characterization and Convergence*. *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. Wiley, New York. [MR0838085](#)
- [10] GRIMMETT, G. R. and STIRZAKER, D. R. (2001). *Probability and Random Processes*, 3rd ed. Oxford Univ. Press, New York. [MR2059709](#)
- [11] HARTLEY, H. O. and DAVID, H. A. (1954). Universal bounds for mean range and extreme observation. *Ann. Math. Stat.* **25** 85–99. [MR0060775](#)
- [12] IOFFE, D. (2009). Stochastic geometry of classical and quantum Ising models. In *Methods of Contemporary Mathematical Statistical Physics. Lecture Notes in Math.* **1970** 87–127. Springer, Berlin. [MR2581610](#)
- [13] IOFFE, D. and LEVIT, A. (2007). Long range order and giant components of quantum random graphs. *Markov Process. Related Fields* **13** 469–492. [MR2357384](#)
- [14] JANSON, S. (2007). On a random graph related to quantum theory. *Combin. Probab. Comput.* **16** 757–766. [MR2346812](#)
- [15] JOSEPH, A. (2014). The component sizes of a critical random graph with given degree sequence. *Ann. Appl. Probab.* **24** 2560–2594. [MR3262511](#)
- [16] LAWLER, G. F. and LIMIC, V. (2010). *Random Walk: A Modern Introduction*. *Cambridge Studies in Advanced Mathematics* **123**. Cambridge Univ. Press, Cambridge. [MR2677157](#)
- [17] ŁUCZAK, T. (1990). Component behavior near the critical point of the random graph process. *Random Structures Algorithms* **1** 287–310. [MR1099794](#)
- [18] NACHMIAS, A. and PERES, Y. (2010). The critical random graph, with martingales. *Israel J. Math.* **176** 29–41. [MR2653185](#)
- [19] NACHMIAS, A. and PERES, Y. (2010). Critical percolation on random regular graphs. *Random Structures Algorithms* **36** 111–148. [MR2583058](#)
- [20] RIORDAN, O. (2012). The phase transition in the configuration model. *Combin. Probab. Comput.* **21** 265–299. [MR2900063](#)
- [21] TUROVA, T. S. (2011). Survey of scalings for the largest connected component in inhomogeneous random graphs. In *Random Walks, Boundaries and Spectra. Progress in Probability* **64** 259–275. Birkhäuser, Basel. [MR3051703](#)
- [22] TUROVA, T. S. (2013). Diffusion approximation for the components in critical inhomogeneous random graphs of rank 1. *Random Structures Algorithms* **43** 486–539. [MR3124693](#)

DEPARTMENTS OF STATISTICS  
AND MATHEMATICS  
STANFORD UNIVERSITY  
STANFORD, CALIFORNIA 94305  
USA  
E-MAIL: [amir@math.stanford.edu](mailto:amir@math.stanford.edu)

DEPARTMENT OF MATHEMATICS  
UNIVERSITY OF BRITISH COLUMBIA  
#121-1984 MATHEMATICS ROAD  
VANCOUVER, BRITISH COLOMBIA  
CANADA V6T 1Z2  
E-MAIL: [anna.levit.14@gmail.com](mailto:anna.levit.14@gmail.com)

TIFR-CENTER FOR APPLICABLE MATHEMATICS  
POST BAG 6503, GKVK POST OFFICE  
BANGALORE 560065  
INDIA  
E-MAIL: [sreekar@tifrbng.res.in](mailto:sreekar@tifrbng.res.in)