

THE CRITICAL KARP–SIPSER CORE OF ERDŐS–RÉNYI RANDOM GRAPHS

Thomas BUDZINSKI* & Alice CONTAT†

Abstract

The Karp–Sipser algorithm consists in removing recursively the leaves as well their unique neighbours and all isolated vertices of a given graph. The remaining graph obtained when there is no leaf left is called the Karp–Sipser core. When the underlying graph is the classical sparse Erdős–Rényi random graph $G[n, \lambda/n]$, it is known to exhibit a phase transition at $\lambda = e$. We show that at criticality, the Karp–Sipser core has size of order $n^{3/5}$, which proves a conjecture of Bauer and Golinelli. We provide the asymptotic law of this renormalized size as well as a description of the distribution of the core as a graph. Our approach relies on the differential equation method, and builds up on a previous work on a configuration model with bounded degrees.

1 Introduction

The Karp–Sipser core. The Karp–Sipser algorithm on a graph \mathbf{g} consists in removing recursively the isolated vertices and the leaves of \mathbf{g} together with their unique neighbours until all vertices have degree at least 2. The subgraph that is obtained once there is no leaf and no isolated vertex left is called the *Karp–Sipser core* of \mathbf{g} and we denote it by $|\text{KSCore}(\mathbf{g})|$.

The initial motivation of Karp and Sipser [11] to study this algorithm is that it provides a computationally efficient way to find a large independent subset in a graph, although finding the largest such set is NP-hard. Indeed, the leaves and isolated vertices removed during the procedure form an independent set, meaning that no pair of these vertices are neighbours. Moreover, this is an optimal strategy in the sense that there exists an independent set of \mathbf{g} with maximal size which contains all these leaves and isolated vertices. Since then, this procedure has been extensively studied on random graph models, with applications to several different domains such as the rank of the adjacency matrix [5] or quantum percolation phase transitions [3].

The Karp–Sipser core of random graphs. In their pioneering paper [11], Karp and Sipser proved that the KS core of an Erdős–Rényi random graph underlies an interesting phase transition in the sparse regime. More precisely, let $G[n, p]$ be the usual Erdős–Rényi random graph, that is a random graph whose set of vertices is $\{1, 2, \dots, n\}$ and containing each of the $\binom{n}{2}$ possible edges with probability p independently. Let us fix $\lambda > 0$. Then

$$\frac{|\text{KSCore}(G[n, \lambda/n])|}{n} \xrightarrow[n \rightarrow \infty]{(d)} C(\lambda),$$

*ENS de Lyon and CNRS.

†Université Sorbonne Paris Nord.

thomas.budzinski@ens-lyon.fr
alice.contat@math.cnrs.fr

where $C(\lambda)$ is a nonnegative constant and $C(\lambda) = 0$ if and only if $\lambda \leq e = 2,7182\dots$

This result was later refined by Aronson, Frieze and Pittel [1] who proved that in the subcritical regime (i.e. when $\lambda < e$), the size of the Karp–Sipser core of $G[n, \lambda/n]$ converges in law towards an (explicit) sum of Poisson random variables as n goes to infinity. Similar results have been obtained for more general configuration models [4, 10]. In another direction, a central limit theorem for the size of the matching created by the algorithm (using an extended version that keeps running even when there is no leaf left) in the supercritical phase was obtained by Kreacic [12]. This was recently extended to the subcritical and critical regimes by $\lambda > e$ Glasgow, Kwan, Sah and Sawhney [8].

The critical regime. Understanding the Karp–Sipser core in the critical regime is even more challenging. Very recently, the analog question was considered by the authors and Curien [6] on a more tractable configuration model of graph with only vertices of degree 1, 2 and 3. It was shown that for this configuration model, the Karp–Sipser core has size of order $n^{3/5}$ in the critical regime, which confirmed a prediction of Bauer and Golinelli [2]. Moreover, the KS core is mostly made of vertices of degree 2 and contains only about $n^{2/5}$ vertices of degree 3. The goal of this paper is to study Erdős–Rényi random graphs $G[n, p]$ in the critical regime, i.e. when $p = e/n$. For $0 \leq m \leq \binom{n}{2}$, we will also denote by $G(n, m)$ a uniform random graph with vertex set $\{1, \dots, n\}$ and exactly m edges.

Theorem 1. *For $i \geq 2$, let $D_i(n)$ be the number of vertices of degree i in the Karp–Sipser core of $G[n, e/n]$. Then we have the joint convergence in distribution*

$$\begin{pmatrix} n^{-3/5} \cdot D_2(n) \\ n^{-2/5} \cdot D_3(n) \\ n^{-1/5} \cdot D_4(n) \\ D_5(n) \end{pmatrix} \xrightarrow[n \rightarrow \infty]{(d)} \begin{pmatrix} \frac{2^{9/5} 3^{4/5}}{e^{3/5}} \vartheta^{-2} \\ \frac{2^{16/5} 3^{1/5}}{e^{2/5}} \vartheta^{-3} \\ \frac{2^{13/5} 3^{3/5}}{e^{1/5}} \vartheta^{-4} \\ \text{Poi}\left(\frac{48}{5} \vartheta^{-5}\right) \end{pmatrix}, \quad (1)$$

where $\vartheta = \inf\{t \geq 0 : W_t = t^{-2}\}$ for a standard Brownian motion W started from 0, and $\text{Poi}(\lambda)$ stands for a Poisson variable with parameter λ . Moreover, with high probability, we have $D_i(n) = 0$ for all $i \geq 6$. Finally, conditionally on $(D_i(n))_{i \geq 2}$, the Karp–Sipser core is a configuration model¹ conditioned to be simple. Moreover, the same result is true if we replace $G[n, e/n]$ by $G(n, m_n)$ where $(m_n)_{n \geq 1}$ is a sequence satisfying $m_n = \frac{e}{2}n + O(\sqrt{n})$.

Ideas of the proof and structure of the paper. To analyze the Karp–Sipser core of this graph, the global strategy is inspired by [6]. The first crucial observation is that the Karp–Sipser algorithm is Abelian, in the sense that the Karp–Sipser core does not depend on the order in which the vertices are removed. In particular, we can remove the leaves (and their neighbours) one by one. In this setting, the execution of the algorithm can be described by an \mathbb{N}^3 -valued Markov chain (X_k, V_k, M_k) already introduced by Karp and Sipser [11]. The three coordinates of this Markov chain correspond respectively to the number X_k of vertices of degree 1, the number V_k of vertices of degree at least 2 and the total number M_k of edges after k steps of the algorithm. Understanding the size of the Karp–Sipser core is then equivalent to understanding the value of this chain at extinction, i.e. when X_k hits 0 for the first time. Rather than the total number of edges, it will be convenient to consider instead the *surplus* of the graph defined by $S_k = 2M_k - X_k - 2V_k$. This is the number of additional

¹We refer to Section 2.1 for a precise definition of the configuration model.

half-edges of the graph compared to a graph where vertices would only have degree 1 or 2. In order to have exactly a Markov chain, it will also be more convenient to consider a multigraph version of the Erdős–Rényi random graph (see Section 2.1 for precise definitions, and Theorem 2 for the analog of Theorem 1 on the multigraph model).

A standard tool to estimate \mathbb{R}^d -valued Markov chains is the *differential equation method*. In our case, this method gives the convergence of the renormalized process $\left((X_{\lfloor tn \rfloor}, V_{\lfloor tn \rfloor}, S_{\lfloor tn \rfloor})/n\right)_{t \geq 0}$ towards a deterministic limit which we call its *fluid limit*. This limit is characterized by a system of differential equations (see Section 3). However, since the size of the core is $o(n)$, this is not accurate enough close to the extinction time, so we will need to make this fluid limit approximation more quantitative. As in [6], the first step is to show that εn steps before the extinction, the three coordinates X_k , V_k and S_k are typically of order respectively $\varepsilon^2 n$, εn and $\varepsilon^{3/2} n$. A difficulty here is that, while the transitions of the Markov chain in [6] are completely explicit, here they depend on an implicit parameter z (defined so that the degrees of non-leaf vertices are well-approximated by Poisson variables with parameter z , conditioned to be at least 2). Together with the fact that vertex degrees are not bounded anymore, this will make the computations more technical. Our estimates on the transitions of the Markov chain will mostly rely on results from [1], which in particular already computed its fluid limit. However, some of the error terms from [1] are a bit too large for our purpose, so we will need to reprove some estimates in Sections 2.2 and 2.3. In Section 3, we will then gather some estimates on the continuous functions appearing in the fluid limit close to their extinction.

After that, as in [6], we will control precisely the drift and variance of the Markov chain in the neighbourhood of the trajectory given by the fluid limit approximation (Section 4). These estimates will allow us to prove that the fluctuations on the number of leaves εn steps before extinction are typically of order $\varepsilon^{3/4} \sqrt{n}$. As in [6], we will argue that the extinction time arises when the fluctuations have the same order of magnitude as the expectation $\approx \varepsilon^2 n$, that is when $\varepsilon \approx n^{-2/5}$. In particular, the size of the Karp–Sipser core is of order $\varepsilon n \approx n^{3/5}$. Moreover, the vertex degrees in the Karp–Sipser core will be well approximated by Poisson variables with parameter $z \approx \sqrt{\varepsilon}$ conditioned to be at least 2, which is where the various orders of magnitude of Theorem 1 come from. Checking that these approximations remain true even when $\varepsilon \approx n^{-2/5}$ requires a careful control of the Markov chain along different scales (Section 5). This step does not introduce new ideas compared to [6].

The last step, which was not needed in [6], consists of coming back from the multigraph model of Theorem 2 to the simple graph models $G[n, e/n]$ and $G(n, m_n)$ of Theorem 1. Since $G(n, m)$ is a random multigraph conditioned to be simple, we will need to carefully study the impact of this conditioning on the Karp–Sipser algorithm (Section 6). We highlight that even though the use of a multigraph model is a classical idea, the most straightforward approaches to deduce Theorem 1 from Theorem 2 fail because the Karp–Sipser core is a very sensitive function of the graph. For this reason, we will actually need a stronger version of Theorem 2 (Theorem 3), which roughly says that conditioning on the value of the Markov chain ηn steps before extinction for some fixed $\eta > 0$ does not affect the law of the core.

Comparison with previous works. Analyzing the Markov chain (X_k, V_k, S_k) to study the Karp–Sipser algorithm on Erdős–Rényi random graphs is far from being a new idea. However, in the previous works [11, 1, 12], it was usually sufficient to estimate this Markov chain up to error terms of order $o(n)$ (fluid limit via the differential equation method) or $o(\sqrt{n})$ (Gaussian fluctuations via

stochastic differential equations). Here, we will need to control X_k at the order $n^{1/5}$ in the end of the process. A key feature that we need to understand here is the “self-correcting” effect of the Markov chain: while the fluctuations on the number X_k of leaves are of order \sqrt{n} during most of the algorithm, they become much smaller when we get close to extinction.

Moreover, many of these works were limited to the subcritical or supercritical regimes, where the limiting drift of the Markov chain behaves pretty nicely around 0. In the critical regime, the fluid limits becomes more degenerate close to its extinction, which explains that the chain is of order $(\varepsilon^2 n, \varepsilon n, \varepsilon^{3/2} n)$ before extinction. The results of [8, Sections 4, 5] can be interpreted as estimates on this Markov chain at time εn before extinction in the regime where $n \rightarrow +\infty$ first, and then $\varepsilon \rightarrow 0$. On the other hand, the estimates that we obtain in the present work (Proposition 5) cover the whole range from ε of order 1 to $n^{-2/5}$. In particular, it seems likely to us that the techniques developed here would provide a more precise estimation of the expected matching number in the central limit theorem of [8].

Contents

1	Introduction	1
2	A three-dimensional Markov chain	4
2.1	Definitions	4
2.2	The asymptotic drift of the Markov chain	7
2.3	General variance estimates	12
3	Fluid limit of the Markov chain	16
3.1	Solution to the differential equations	16
3.2	The fluid limit near the extinction time	17
4	Drift and variance estimates in the neighbourhood of the fluid limit	21
5	Good region and a stochastic differential equation	27
5.1	Rough behaviour of the fluctuations	27
5.2	Reaching the end via a stochastic differential equation	31
5.3	Size and composition of the Karp–Sipser core	33
6	From the configuration model back to Erdős–Rényi	35

2 A three-dimensional Markov chain

2.1 Definitions

As mentioned above, the execution of the Karp–Sipser algorithm on $G[n, e/n]$ can be described by a triple with integer values representing the number of leaves, the number of vertices of degree at least 2 and the number of “surplus” half-edges. However, the multigraph version of the model enjoys a nicer Markov property compared to simple graphs. This motivates the following definitions.

Random (multi)graphs with a fixed number of edges. For $0 \leq m \leq \binom{n}{2}$, we define $\mathbb{G}(n, m)$ (resp. $\mathbb{G}(n, m)$) as the uniform graph (resp. multigraph) on n vertices with m edges. More precisely, we use labels to define this graph to avoid symmetry issues: a *labelled multigraph* with m edges on the set of vertices $\{1, \dots, n\}$ is defined by its edge-sequence $(a_i, b_i)_{1 \leq i \leq m}$, where $a_i, b_i \in \{1, \dots, n\}$ for all $1 \leq i \leq m$. Note that it is possible that $a_i = b_i$, in which case the edge i is a loop. Then the multigraph $\mathbb{G}(n, m)$ is obtained by taking a uniform labelled multigraph with m edges on the set of vertices $\{1, \dots, n\}$ and forgetting its edge-labels, and the graph $\mathbb{G}(n, m)$ has the law of $\mathbb{G}(n, m)$ conditioned to be simple (i.e. without loops and multiple edges). We will first study the Karp–Sipser core of $\mathbb{G}(n, m)$ and then deduce our Theorem 1. For $d_1 + \dots + d_n$ even, a *configuration model* with vertex degrees d_1, \dots, d_n is the random multigraph obtained by drawing d_i half-edges around the vertex i for all $1 \leq i \leq n$ and pairing those half-edges uniformly at random.

Theorem 2. *Let $(m_n)_{n \geq 1}$ be a (deterministic) sequence of nonnegative integers such that $m_n = \frac{e}{2}n + O(\sqrt{n})$. For $i \geq 2$, let $\mathbb{D}_i(n)$ be the number of vertices of degree i in the Karp–Sipser core of $\mathbb{G}(n, m_n)$. Then we have the joint convergence in distribution*

$$\begin{pmatrix} n^{-3/5} \cdot \mathbb{D}_2(n) \\ n^{-2/5} \cdot \mathbb{D}_3(n) \\ n^{-1/5} \cdot \mathbb{D}_4(n) \\ \mathbb{D}_5(n) \end{pmatrix} \xrightarrow[n \rightarrow \infty]{(d)} \begin{pmatrix} \frac{2^{9/5} 3^{4/5}}{e^{3/5}} \vartheta^{-2} \\ \frac{2^{16/5} 3^{1/5}}{e^{2/5}} \vartheta^{-3} \\ \frac{2^{13/5} 3^{3/5}}{e^{1/5}} \vartheta^{-4} \\ \text{Poi}\left(\frac{48}{5} \vartheta^{-5}\right) \end{pmatrix},$$

where $\vartheta = \inf\{t \geq 0 : W_t = t^{-2}\}$ for a standard Brownian motion W started from 0. Moreover, with high probability, we have $\mathbb{D}_i(n) = 0$ for all $i \geq 6$. Finally, conditionally on $(\mathbb{D}_i(n))_{i \geq 2}$, the Karp–Sipser core is a configuration model.

Fluctuations of the initial condition. For the rest of the paper (until Section 6), we fix a deterministic sequence $(m_n)_{n \geq 1}$ of nonnegative integers such that $m_n = \frac{e}{2}n + O(\sqrt{n})$. Two important quantities in our analysis are the number of leaves and of vertices of degree at least 2 in $\mathbb{G}(n, m_n)$, that we denote respectively by X_0^n and V_0^n . We also define $S_0^n = 2m_n - X_0^n - 2V_0^n$, which can be seen as the number of “additional half-edges” compared to the case when all vertices of degree at least 2 have degree exactly 2. We first estimate the triplet (X_0^n, V_0^n, S_0^n) , which will be the initial condition of our Markov chain. Heuristically, the vertex degrees in $\mathbb{G}(n, m_n)$ can be approximated by i.i.d. Poisson variables with parameter e . Thus, at first order, there are $e^{1-e}n$ leaves and $(1 - e^{-e} - e^{1-e})n$ vertices of degree at least 2. The following result is close to [8, Lemma 4.3], except that it is for m_n fixed instead of random.

Lemma 1. *The following vector is tight as $n \rightarrow +\infty$:*

$$\frac{1}{\sqrt{n}} \left(X_0^n - e^{1-e}n, V_0^n - (1 - e^{-e} - e^{1-e})n, S_0^n - (e + e^{1-e} + 2e^{-e} - 2)n \right).$$

Proof. First, we note that the degree distribution of the n vertices in $\mathbb{G}(n, m_n)$ corresponds exactly to the random allocation problem studied by Janson in [9, Example 3.1], where $m = 2m_n$ balls (or half-edges) are thrown uniformly at random in n boxes (or vertices). In particular, as an application of [9, Corollary 2.5], the normalized fluctuations

$$\frac{1}{\sqrt{n}} \left((X_0^n, V_0^n, S_0^n) - \left(\frac{2m_n}{n} e^{-\frac{2m_n}{n}}, 1 - \left(1 + \frac{2m_n}{n}\right) e^{-\frac{2m_n}{n}}, \frac{2m_n}{n} + 2e^{-\frac{2m_n}{n}} + \frac{2m_n}{n} e^{-\frac{2m_n}{n}} - 2 \right) n \right)$$

converge in distribution towards a Gaussian vector. The covariance matrix of the limit is explicit but we do not need it here. Moreover, since $m_n = \frac{en}{2} + O(\sqrt{n})$, we have

$$\begin{aligned} & \left(\frac{2m_n}{n} e^{-\frac{2m_n}{n}}, 1 - \left(1 + \frac{2m_n}{n} \right) e^{-\frac{2m_n}{n}}, \frac{2m_n}{n} + 2e^{-\frac{2m_n}{n}} + \frac{2m_n}{n} e^{-\frac{2m_n}{n}} - 2 \right) \\ &= \left(e^{1-e}, 1 - e^{-e} - e^{1-e}, e + e^{1-e} + 2e^{-e} - 2 \right) + O\left(\frac{1}{\sqrt{n}}\right), \end{aligned}$$

which concludes the proof of the Lemma. \square

Multigraphs with a fixed number of leaves, non-leaf vertices and edges. An important distribution for us will be uniform multigraphs with given numbers of leaves, of vertices of degree at least 2 and of edges. For $x, v, s \geq 0$ such that $x + 2v + s$ is even, we consider a uniform random variable among all the labelled multigraphs on the vertex set $\{1, \dots, x + v\}$ with $\frac{x+2v+s}{2}$ edges and such that the vertices $1, \dots, x$ have degree 1 and the vertices $x + 1, \dots, x + v$ have degree at least 2. We denote by $\mathbf{G}(x, v, s)$ the random multigraph obtained by forgetting the edge-labels of this labelled multigraph. We call s the *surplus* of the multigraph, as it is the additional number of half-edges compared to the case where all vertices have degree either 1 or 2.

Lemma 2. *Let $x, v, s \geq 0$ with $x + 2v + s$ even. Conditionally on the family of vertex degrees $(\deg(j))_{1 \leq j \leq x+v}$, the multigraph $\mathbf{G}(x, v, s)$ is a configuration model.*

Proof. The labelled multigraph $\mathbf{G}(x, v, s)$ is uniform among all the labelled multigraphs with x leaves, with v vertices of degree at least 2 and with surplus s , so it is still uniform if we condition on all the vertex degrees. \square

In particular, the two random multigraphs $\mathbf{G}(n, m_n)$ and $\mathbf{G}(X_0^n, V_0^n, S_0^n)$ have the same law.

One step of the Karp–Sipser algorithm. Equipped with this definition, we now want to apply the Karp–Sipser algorithm on the random multigraphs $\mathbf{G}(x, v, s)$. Given a multigraph G which is a possible realization of $\mathbf{G}(x, v, s)$ for some $x \geq 1$ and $v, s \geq 0$, we denote by G^- the multigraph obtained from G by the following sequence of operations (see Figure 1):

1. remove the leaf with label 1 and its unique neighbour,
2. remove all the edges incident to that neighbour,
3. remove the isolated vertices that might appear in the process,
4. relabel the leaves in increasing order followed by the vertices of degree at least 2 in increasing order.

Our starting point is the following property, already observed in [1, Lemma 2].

Lemma 3. *If G follows the law of $\mathbf{G}(x, v, s)$, then conditionally on the number X^- of leaves, the number V^- of non-leaf vertices and the surplus S^- of G^- , the graph G^- follows the law of $\mathbf{G}(X^-, V^-, S^-)$.*

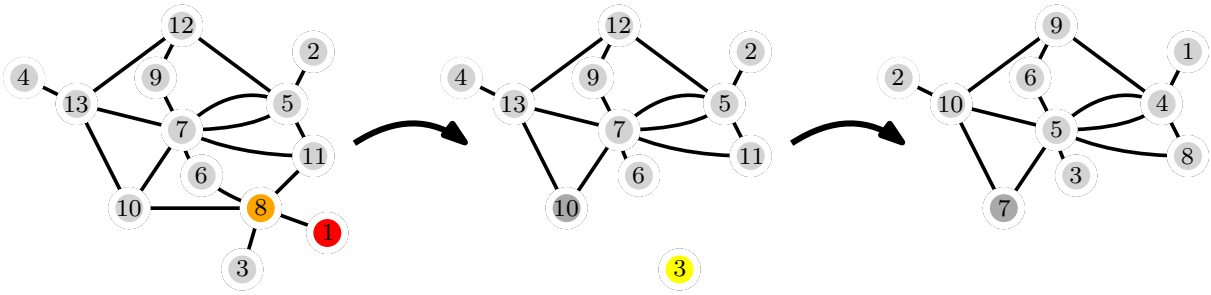


Figure 1: Illustration of one step of the Karp–Sipser algorithm. We first remove the leaf with label 1 (in red) and its unique neighbour (in orange) and all edges incident to them. Then we remove the isolated vertices (in yellow) and relabel the leaves in increasing order followed by the vertices of degree at least 2 in increasing order.

The Markov chain. As a result, the Karp–Sipser algorithm yields a natural three-dimensional random process defined as follows. Let us start from $G_0^n := \mathbf{G}(X_0^n, V_0^n, S_0^n)$, where we recall that (X_0^n, V_0^n, S_0^n) are respectively the number of leaves, vertices of degree at least 2 and surplus of $\mathbf{G}(n, m_n)$. For all $k \geq 0$, if G_k^n has at least one leaf, let $G_{k+1}^n = (G_k^n)^-$. If G_k^n has no leaf, we just set $G_{k+1}^n = G_k^n$, and we denote by θ^n the first time at which this is the case. Then $G_{\theta^n}^n$ is the Karp–Sipser core of G_0^n (up to a relabelling of its vertices). For all $k \geq 0$, we denote by X_k^n (resp. V_k^n , M_k^n) the number of leaves (resp. non-leaf vertices, half-edges) of G_k^n , and by $S_k^n = 2M_k^n - X_k^n - 2V_k^n$ its surplus. We will also denote by $(\mathcal{F}_k^n)_{k \geq 0}$ the associated filtration, i.e. \mathcal{F}_k^n is the σ -algebra generated by $(X_i^n, V_i^n, S_i^n)_{0 \leq i \leq k}$. As observed in [11, Assertion (1), p266] or [1, Lemma 3], this defines a Markov chain.

Lemma 4. *The process $(X_k^n, V_k^n, S_k^n)_{0 \leq k \leq \theta^n}$ is a Markov chain. Moreover, for any stopping time T , conditionally on \mathcal{F}_T^n , the graph G_T^n has the law of $\mathbf{G}(X_T^n, V_T^n, S_T^n)$.*

The proof of the phase transition by Karp and Sipser [11] relies on this Markov chain. Using the *differential equation method* [15] for the first time in the context of random graphs, they showed that the renormalized version of this Markov chain is well approximated by a differential equation on \mathbb{R}^3 for which they find an explicit solution, see Section 3 below.

To simplify notation, the n in the exponent will be implicit for the rest of the paper when there is no ambiguity, even if we will often look at asymptotics as $n \rightarrow +\infty$.

2.2 The asymptotic drift of the Markov chain

The next step is to estimate the local drift of the Markov chain (X, V, S) (Proposition 1 below). Most of the estimates that we need were already proved in [1, Lemma 6]. Unfortunately, in one of those estimates, the error term is too large for our purpose, which is why we will need to redo some of the proofs. As in [1], the first step is to understand the distribution of vertex degrees in the random graph $\mathbf{G}(x, v, s)$. We will see that the degrees of non-leaf vertices are well approximated by i.i.d. Poisson variables conditioned to be at least 2. We will now define the parameter z of these Poisson variables: for all $v, s > 0$, we denote by $z(v, s) > 0$ the unique solution to the equation

$$\frac{z(e^z - 1)}{e^z - z - 1} = 2 + \frac{s}{v}. \quad (2)$$

We also set the convention $z(v, 0) = 0$ for $v > 0$, and we will write $Z_k = z(V_k, S_k)$. Note that in [1], the right-hand side is written in a different form, as it is expressed in terms of x, v and the number of edges. Since the number of edges is equal to $\frac{1}{2}(x + 2v + s)$, the adaptation is immediate. It will be important for us to understand the behaviour of z when s and v go to 0. Although it is not a smooth function of (v, s) as $v \rightarrow 0$, we can still write it as a smooth function of the ratio $\frac{s}{v}$.

Lemma 5. *There is a smooth function $\tilde{z}: [0, +\infty) \rightarrow [0, +\infty)$ such that for all $v > 0$ and $s \geq 0$, we have $z(v, s) = \tilde{z}\left(\frac{s}{v}\right)$. Moreover, we have*

$$z(v, s) = 3\frac{s}{v} + O\left(\left(\frac{s}{v}\right)^2\right). \quad (3)$$

In particular, we will see later that εn steps before extinction, the quantities S_k and V_k are respectively of order $\varepsilon^{3/2}n$ and εn , so Z_k will be of order $\varepsilon^{1/2}$.

Proof. We first verify that the derivative of the left-hand side of (2) with respect to z is positive. Specifically, we have

$$\frac{d}{dz} \frac{z(e^z - 1)}{e^z - z - 1} = \frac{e^{2z} - 2e^z + 1 - z^2e^z}{(e^z - z - 1)^2} = \frac{e^z(2\cosh(z) - 2 - z^2)}{(e^z - z - 1)^2}. \quad (4)$$

The numerator is positive for $z > 0$ by expanding $\cosh(z)$ into a power series. Note that this proves that the solution to (2) is indeed unique. Moreover (2) defines $\tilde{z}\left(\frac{s}{v}\right)$ for $\frac{s}{v} \neq 0$ and can be extended to a continuous function by setting $\tilde{z}(0) = 0$. Since (4) does not vanish, the function \tilde{z} is smooth on $(0, +\infty)$. Therefore, to conclude, all we need to check is that (4) does not vanish as $z \rightarrow 0$. But in this regime, the numerator of (4) is equivalent to $\frac{1}{12}z^4$ and its denominator to $\frac{z^4}{4}$, so $\frac{d}{dz} \frac{z(e^z - 1)}{e^z - z - 1} > 0$ on $[0, +\infty)$.

Finally, the second part (3) immediately follows from the fact that the left-hand side of (2) is $2 + \frac{z}{3} + O(z^2)$ as $z \rightarrow 0$ and is of order z as $z \rightarrow +\infty$. \square

Poisson approximation for vertex degrees. We now show as in [1, Lemmas 4, 5] that vertex degrees are well approximated by i.i.d. Poisson variables conditioned to be at least 2. As it will appear a lot, we define $f(z) = e^z - z - 1$.

Lemma 6. *Let $C > 0$ be a constant and assume that $V_k Z_k \geq \log^3 n$ and $Z_k \leq C$. In what follows, the constants involved in the O notation will only depend on C .*

1. *For all $X_k + 1 \leq j \leq X_k + V_k$ and for all $3 \leq i \leq \log n$, we have*

$$\mathbb{P}\left(\deg_{G_k}(j) = i | \mathcal{F}_k\right) = \frac{Z_k^i}{i!f(Z_k)} \left(1 + O\left(\frac{i^2}{V_k Z_k}\right)\right). \quad (5)$$

Moreover, if $i = 2$, the $O\left(\frac{i^2}{V_k Z_k}\right)$ error term can be replaced by $O\left(\frac{1}{V_k}\right)$. Also, for all $i \geq 2$, we have

$$\mathbb{P}\left(\deg_{G_k}(j) = i | \mathcal{F}_k\right) = O\left(\sqrt{V_k Z_k} \frac{Z_k^i}{i!f(Z_k)}\right). \quad (6)$$

In particular, we have

$$\mathbb{P}\left(\deg_{G_k}(j) \geq \log n | \mathcal{F}_k\right) = O\left(n^{-10}\right). \quad (7)$$

2. For all $X_k + 1 \leq j_1 < j_2 \leq X_k + V_k$ and $2 \leq i_1, i_2 \leq \log n$, we have

$$\mathbb{P}\left(\deg_{G_k}(j_1) = i_1 \text{ and } \deg_{G_k}(j_2) = i_2 | \mathcal{F}_k\right) = \frac{Z_k^{i_1}}{i_1! f(Z_k)} \frac{Z_k^{i_2}}{i_2! f(Z_k)} \left(1 + O\left(\frac{\log^2 n}{V_k Z_k}\right)\right). \quad (8)$$

Moreover, if $i_1 = i_2 = 2$, the error term can be replaced by $O\left(\frac{\log^2 n}{V_k}\right)$.

3. For all $X_k + 1 \leq j_1 < j_2 < j_3 \leq X_k + V_k$ and $2 \leq i_1, i_2, i_3 \leq \log n$, we have

$$\mathbb{P}\left(\forall a \in \{1, 2, 3\}, \deg_{G_k}(j_a) = i_a | \mathcal{F}_k\right) = \left(\prod_{a=1}^3 \frac{Z_k^{i_a}}{i_a! f(Z_k)}\right) \times \left(1 + O\left(\frac{\log^2 n}{V_k Z_k}\right)\right). \quad (9)$$

4. Let $m \geq 1$. Then for all $X_k + 1 \leq j_1 < \dots < j_m \leq X_k + V_k$, we have

$$\mathbb{P}\left(\deg_{G_k}(j_1) = \dots = \deg_{G_k}(j_m) = 5 | \mathcal{F}_k\right) = (1 + o(1)) \left(\frac{Z_k^5}{5! f(Z_k)}\right)^m, \quad (10)$$

where the o only depends on m, C .

Proof. 1. This item is exactly the same as [1, Lemma 5], except for (7) and for the improved error term $O\left(\frac{1}{V_k}\right)$ in (5) for $i = 2$. We first deduce (7) from (6). Using (6), we can write

$$\mathbb{P}\left(\deg_{G_k}(j) \geq \log n | \mathcal{F}_k\right) = O\left(\sqrt{V_k Z_k} \sum_{i > \log n} \frac{Z_k^i}{i! f(Z_k)}\right).$$

Using $f(z) \geq c \times z^2$ on $[0, C]$ for some constant $c > 0$ and $i! \geq i^i e^{-i}$, we deduce

$$\begin{aligned} \mathbb{P}\left(\deg_{G_k}(j) \geq \log n | \mathcal{F}_k\right) &= O\left(n^{1/2} \sum_{i > \log n} \frac{Z_k^{i-2} e^i}{i^i}\right) \\ &= O\left(n^{1/2} \sum_{i > \log n} \left(\frac{eC}{\log n}\right)^i\right) \\ &= O\left(n^{1/2} \frac{(eC)^{\log n}}{(\log n)^{\log n}}\right). \end{aligned}$$

Since the numerator is polynomial in n and the denominator is larger than any polynomial, this is $O(n^{-10})$.

We now move on to the case $i = 2$ of (5). For this, using (5) for $i \geq 3$ and (7), we can write

$$\begin{aligned} \mathbb{P}\left(\deg_{G_k}(j) = 2 | \mathcal{F}_k\right) &= 1 - \sum_{i \geq 3} \mathbb{P}\left(\deg_{G_k}(j) = i | \mathcal{F}_k\right) \\ &= 1 - \sum_{i=3}^{\log n} \frac{Z_k^i}{i! f(Z_k)} + O\left(\frac{1}{V_k Z_k f(Z_k)} \sum_{i=3}^{\log n} \frac{i^2}{i!} Z_k^i\right) + O(n^{-10}). \quad (11) \end{aligned}$$

Since $0 < Z_k \leq C$, the quantity $\frac{Z_k^2}{f(Z_k)}$ is bounded away from 0 and $+\infty$, so what we want to show is that (11) is $\frac{Z_k^2}{2f(Z_k)} + O\left(\frac{1}{V_k}\right)$. First, by definition of f we have $\sum_{i \geq 2} \frac{z^i}{i! f(z)} = 1$ for all z , so

$$1 - \sum_{i=3}^{\log n} \frac{Z_k^i}{i! f(Z_k)} = \frac{Z_k^2}{2f(Z_k)} + O\left(\sum_{i > \log n} \frac{Z_k^i}{i! f(Z_k)}\right) = \frac{Z_k^2}{2f(Z_k)} + O(n^{-10})$$

by the same computation as in the proof of (7). In particular, the error term is $O\left(\frac{1}{V_k}\right)$. We now have to handle the first error term of (11). We rewrite it as

$$O\left(\frac{Z_k^2}{V_k f(Z_k)} \sum_{i=3}^{\log n} \frac{i^2}{i!} Z_k^{i-3}\right) = O\left(\frac{1}{V_k} \sum_{i \geq 3} \frac{i^2}{i!} C^{i-3}\right) = O\left(\frac{1}{V_k}\right).$$

The second error term is $O(n^{-10})$ by (7), which concludes the proof.

2. This is proved in [1, Lemma 5], except for two small modifications:

- in [1, Lemma 5], the corresponding estimate was only stated for $i_1, i_2 \leq \log V_k$ and not for $\log n$, but the error term was $\frac{\log^2 V_k}{V_k Z_k}$ and not $\frac{\log^2 n}{V_k Z_k}$. However, the proof remains exactly the same, relying on [1, Equations (5), (6)];
- again, we claim an improved error term in the case $i_1 = i_2 = 2$. For this improvement, we use the exact same argument as for $i = 2$ in Item 1, writing

$$\begin{aligned} & \mathbb{P}\left(\deg_{G_k}(j_1) = 2 \text{ and } \deg_{G_k}(j_2) = 2 | \mathcal{F}_k\right) \\ &= \mathbb{P}\left(\deg_{G_k}(j_1) = 2 | \mathcal{F}_k\right) - \sum_{i_2 \geq 3} \mathbb{P}\left(\deg_{G_k}(j_1) = 2 \text{ and } \deg_{G_k}(j_2) = i_2 | \mathcal{F}_k\right). \end{aligned}$$

3. The proof of the three-point estimate (9) is exactly the same as the proof of the two-point estimate in [1, Lemma 5], with the same remark as above on replacing $\log V_k$ by $\log n$.

4. Again, the argument is the same as for the two-point estimate. Note that the error $\frac{\log^2 n}{V_k Z_k}$ is indeed $o(1)$ (this is why we assumed $V_k Z_k \geq \log^3 n$, which is slightly stronger than in [1]) and that the error $o(1)$ can depend on m , so we do not need a uniform statement in m . \square

Asymptotic drift. We now define the functions that will describe the drift of our Markov chain (X, V, S) in the fluid limit. In the equations below, we will just write z instead of the $z(v, s)$ given by (2). For $x, v, s > 0$, we write

$$\Phi_A(x, v, s) = -1 - \frac{x}{x + 2v + s} + \frac{v^2 z^4 e^z}{(x + 2v + s)^2 f(z)^2} - \frac{x v z^2 e^z}{(x + 2v + s)^2 f(z)}, \quad (12)$$

$$\Phi_B(x, v, s) = -1 + \frac{x}{x + 2v + s} - \frac{v^2 z^4 e^z}{(x + 2v + s)^2 f(z)^2}, \quad (13)$$

$$\Phi_C(x, v, s) = 1 - \frac{x}{x + 2v + s} - 2 \frac{v z^2 e^z}{(x + 2v + s) f(z)} + \frac{v^2 z^4 e^z}{(x + 2v + s)^2 f(z)^2} + \frac{x v z^2 e^z}{(x + 2v + s)^2 f(z)}. \quad (14)$$

It was proved in [1, Lemma 6] that the fluid limit of the Markov chain (X, V, S) satisfies a differential equation given by these three functions. We write ΔX_k for $X_{k+1} - X_k$, and similarly for the other processes. As we need slightly better error terms than those of [1], we will improve some of those estimates.

Proposition 1. *Let $C > 0$. If $V_k Z_k \geq \log^3 n$ and $Z_k \leq C$ and $X_k > 0$, then we have*

$$\mathbb{E} [\Delta X_k | \mathcal{F}_k] = \Phi_A (X_k, V_k, S_k) + O \left(\frac{\log^2 n}{V_k} \right), \quad (15)$$

$$\mathbb{E} [\Delta V_k | \mathcal{F}_k] = \Phi_B (X_k, V_k, S_k) + O \left(\frac{\log^2 n}{V_k Z_k} \right), \quad (16)$$

$$\mathbb{E} [\Delta S_k | \mathcal{F}_k] = \Phi_C (X_k, V_k, S_k) + O \left(\frac{\log^2 n}{V_k Z_k} \right), \quad (17)$$

where the constants implied by the O notation only depend on C .

Proof. To make the following computations more compact, we introduce the number $H_k := 2V_k + X_k + S_k = 2M_k$ of half-edges of the graph G_k . Note that H_k corresponds to $2m$ in the notations of [1, Lemma 6], so all the error terms $O\left(\frac{1}{m}\right)$ of [1] can be replaced by $O\left(\frac{1}{H_k}\right)$, which is also $O\left(\frac{1}{V_k}\right)$.

Equations (16) and (17) are already proved in [1, Lemma 6] (with the natural substitution of the number $2m$ of half-edges by $x + 2v + s$), so there is nothing to do. On the other hand (15) is only proved in [1] with an error term $O\left(\frac{\log^2 V_k}{V_k Z_k}\right)$, which we need to improve. However, our argument will follow [1] closely.

Let $k \geq 0$ be such that V_k , X_k and Z_k satisfy the assumptions of the Proposition. We recall that to pass from G_k to G_{k+1} , we start by removing the leaf 1 of G_k , and we denote by y its unique neighbour in G_k . All of the vertex degrees in the proof below will be degrees in the graph G_k . We denote by D_k the degree of y and by $D_k^{(1)}$ (resp. $D_k^{(2)}$, $D_k^{(3)}$) the number of neighbours of degree 1 (resp. 2, at least 3) of y in G_k , with the leaf 1 counted in $D_k^{(1)}$. Let also \tilde{D}_k be the total number of loops and multiple edges incident to y . It is proved in [1, Equation (8)] that

$$\mathbb{E} [\tilde{D}_k | \mathcal{F}_k] = O \left(\frac{1}{H_k} \right) = O \left(\frac{1}{V_k} \right). \quad (18)$$

Moreover, we can write

$$\Delta X_k = -D_k^{(1)} - \mathbb{1}_{D_k=1} + D_k^{(2)} + O \left(\tilde{D}_k \right). \quad (19)$$

Indeed, the leaves that are removed are the neighbours of y , plus y itself if it has degree 1, whereas the leaves which are created are the neighbours of y with degree 2 (unless they are linked to y by multiple edges, which is accounted for by the term $O(\tilde{D}_k)$). We now estimate the expectations of these terms one by one. First, by Lemma 2, we have

$$\mathbb{P} (D_k = 1 | \mathcal{F}_k) = \frac{X_k - 1}{H_k - 1}. \quad (20)$$

Moreover, let \mathcal{D}_k be the family of vertex degrees of G_k . By Lemma 2 again, we can write

$$\mathbb{E} [D_k^{(1)} | \mathcal{F}_k, \mathcal{D}_k] = 1 + O(\tilde{D}_k) + \sum_{j: \deg(j) \geq 2} \frac{\deg(j)}{H_k - 1} \cdot (\deg(j) - 1) \cdot \frac{X_k - 1}{H_k - 3},$$

where the sum over j corresponds to the possible values of y and the error term accounts for the possibility of loops and multiple edges around y . We can now integrate over \mathcal{D}_k and use (18) and

Lemma 6. By decomposing according to the possible values d of $\deg(j)$, we get

$$\begin{aligned} \mathbb{E} \left[D_k^{(1)} | \mathcal{F}_k \right] &= 1 + O \left(\frac{1}{V_k} \right) + \frac{X_k - 1}{(H_k - 1)(H_k - 3)} \sum_{j=X_k+1}^{X_k+V_k} \mathbb{E} [\deg(j) (\deg(j) - 1) | \mathcal{F}_k] \\ &= 1 + O \left(\frac{1}{V_k} \right) + \frac{(X_k - 1)V_k}{(H_k - 1)(H_k - 3)} \cdot \left(\frac{2Z_k^2}{2f(Z_k)} \left(1 + O \left(\frac{1}{V_k} \right) \right) \right. \\ &\quad \left. + \sum_{d=3}^{\log n} \frac{d(d-1)Z_k^d}{d!f(Z_k)} \left(1 + O \left(\frac{\log^2 n}{V_k Z_k} \right) \right) + O(n^{-8}) \right), \end{aligned} \quad (21)$$

where the last term is the contribution of $\deg(j) \geq \log n$ and comes from (7). The second error term is $O \left(\frac{X_k Z_k^2}{H_k^2 f(Z_k)} \right)$, which is $O \left(\frac{1}{V_k} \right)$ since $Z_k^2 = O(f(Z_k))$ and $X_k, V_k \leq H_k$ (we are here slightly more accurate than [1]). On the other hand, for $3 \leq d \leq \log n$, we have $\frac{Z_k^d}{f(Z_k)} = \frac{Z_k^2}{f(Z_k)} \cdot Z_k \cdot Z_k^{d-3} = O(C^{d-3} Z_k)$, where C is an upper bound on Z_k as in the statement of the Proposition. Hence, the corresponding error term is $O \left(\frac{C^d \log^2 n}{(d-2)! V_k} \right)$. Therefore, the sum of the error terms over $3 \leq d \leq \log n$ is $O \left(\frac{\log^2 n}{V_k} \right)$. Finally, the last error term is $O \left(\frac{\log^2 n}{V_k} \right)$. Summing the main terms as in [1, Equation (17)], we obtain

$$\mathbb{E} \left[D_k^{(1)} | \mathcal{F}_k \right] = 1 + \frac{X_k V_k Z_k^2 e^{Z_k}}{H_k^2 f(Z_k)} + O \left(\frac{\log^2 n}{V_k} \right). \quad (22)$$

We now estimate the number of neighbours of y with degree 2 using similar ideas. We first write

$$\mathbb{E} \left[D_k^{(2)} | \mathcal{F}_k, \mathcal{D}_k \right] = O(\tilde{D}_k) + \sum_{j_1; \deg(j_1) \geq 2} \sum_{j_2; \deg(j_2) = 2} \frac{\deg(j_1)}{H_k - 1} \times (\deg(j_1) - 1) \times \frac{2}{H_k - 3},$$

where j_1 corresponds to possible values of y , and j_2 corresponds to possible values of neighbours of y other than 1. Again, we integrate over \mathcal{D}_k and use (18) to bound the error term:

$$\mathbb{E} \left[D_k^{(2)} | \mathcal{F}_k \right] = O \left(\frac{1}{V_k} \right) + \frac{2}{(H_k - 1)(H_k - 3)} \sum_{j_1=X_k+1}^{X_k+V_k} \sum_{j_2=X_k+1}^{X_k+V_k} \mathbb{E} \left[\deg(j_1) (\deg(j_1) - 1) \mathbb{1}_{\deg(j_2)=2} \right].$$

We then decompose each term according to the value of $\deg(j_1)$ as in (21) and apply the estimate (8) from Lemma 6. The error terms are the same as in the computation for $D_1^{(k)}$ so, using our improved error in (8) for $i_1 = i_2 = 2$, we find

$$\mathbb{E} \left[D_k^{(2)} | \mathcal{F}_k \right] = \frac{V_k^2 Z_k^4 e^{Z_k}}{H_k^2 f(Z_k)^2} + O \left(\frac{\log^2 n}{V_k} \right), \quad (23)$$

which again is a slight improvement compared to [1, Equation (18)]. Finally, we obtain (15) with the right error term by plugging (18), (20), (22) and (23) in (19). \square

2.3 General variance estimates

In order to show that the Markov chain stays close to its fluid limit, we will also need to control its variance. As we will see later in Section 5.2, it will be particularly important for us to have a precise estimate on $\text{Var}(\Delta X_k)$ in the end of the process. To fix ideas, we mention right now that εn steps before the end of the process, the ratio $\frac{X_k}{V_k}$ will be of order ε and Z_k of order $\varepsilon^{1/2}$.

Proposition 2. *Let $C > 0$. If $V_k Z_k \geq \log^3 n$ and $Z_k \leq C$ and $X_k > 0$, then we have*

$$\text{Var}(\Delta V_k | \mathcal{F}_k) = O(1), \quad (24)$$

$$\text{Var}(\Delta S_k | \mathcal{F}_k) = O(Z_k) + O\left(\frac{X_k}{V_k}\right), \quad (25)$$

$$\text{Var}(\Delta X_k | \mathcal{F}_k) = Z_k + O(Z_k^2) + O\left(\frac{X_k}{V_k}\right) + O\left(\frac{\log^2 n}{V_k}\right), \quad (26)$$

where the constants implied by the O notation only depend on C .

Proof. We keep the notation of the proof of Proposition 1. We start by proving (24). Note that $0 \geq \Delta V_k \geq -D_k$, so it is enough to prove $\mathbb{E}[D_k^2 | \mathcal{F}_k] = O(1)$. For this, we use the same ideas as in the proof of Proposition 1. We first note that $D_k \leq n$ so by (7), the contribution of the event $\{D_k > \log n\}$ is $O\left(\frac{1}{n}\right)$. Summing over all possible values of the neighbour y of the removed leaf and over all possible values of D_k between 3 and $\log n$ and using Lemma 6, we find

$$\mathbb{E}\left[D_k^2 \mathbb{1}_{D_k \geq 3} | X_k, V_k, S_k\right] = O\left(\frac{1}{n}\right) + O\left(\frac{V_k}{H_k - 1} \sum_{d=2}^{\log n} d^3 \frac{Z_k^d}{d! f(Z_k)}\right) = O(1)$$

using $V_k \leq H_k - 1$ and $\frac{Z_k^d}{d! f(Z_k)} = O(Z_k^{d-2}) = O(1)$.

The argument for (25) is similar, but we also need to notice that if both Z_k and $\frac{X_k}{V_k}$ are very small, then the most likely case is that y has degree 2 and its second neighbour as well, i.e. $D_k = 2$ and $D_k^{(1)} = D_k^{(2)} = 1$. In this case, we have $\Delta X_k = 0$ but $\Delta V_k = -2$ and $\Delta H_k = -4$, so $\Delta S_k = 0$. Therefore, we can write the bound

$$\text{Var}(\Delta S_k | \mathcal{F}_k) = O\left(\mathbb{E}\left[D_k^2 \left(\mathbb{1}_{D_k=1} + \mathbb{1}_{D_k \geq 3} + \mathbb{1}_{D_k=2, D_k^{(1)}=2} + \mathbb{1}_{D_k=2, D_k^{(3)}=1}\right) | \mathcal{F}_k\right]\right).$$

We can then bound each of the four terms one by one. We find that the contributions of the first and third terms are $O\left(\frac{X_k}{V_k}\right)$, whereas the second and fourth are $O(Z_k)$. Let us illustrate this with the proof for the fourth term, as it is the most complicated one. We sum over all possible values of y, w, d , where y is the unique neighbour of the removed leaf 1 and w is the second neighbour of y , and where $d \geq 3$ is the degree of w .

$$\mathbb{E}\left[D_k^2 \mathbb{1}_{D_k=2, D_k^{(3)}=1} | \mathcal{F}_k\right] = \sum_{d \geq 3} \sum_{y=X_k+1}^{X_k+V_k} \sum_{\substack{w=X_k+1 \\ w \neq y}}^{X_k+V_k} 4\mathbb{P}(\deg(y) = 2 \text{ and } \deg(w) = d) \times \frac{2}{H_k - 1} \times \frac{d}{H_k - 3}.$$

We now use Lemma 6. Using (7) for $d \geq \log n$ and (8) for $3 \leq d \leq \log n$, we find

$$\begin{aligned} \mathbb{E}\left[D_k^2 \mathbb{1}_{D_k=2, D_k^{(3)}=1} | \mathcal{F}_k\right] &= O(n^{-8}) + O\left(V_k^2 \sum_{d=3}^{\log n} \frac{d}{H_k^2} \frac{Z_k^{d+2}}{d! f(Z_k)^2}\right) \\ &= O(n^{-8}) + O\left(Z_k \sum_{d \geq 3} \frac{d}{d!} C^{d-3}\right) \\ &= O(Z_k), \end{aligned}$$

where we have used that $f(Z_k)$ is comparable to Z_k^2 , and the assumption $V_k Z_k \geq \log^3 n$ to guarantee that $O(n^{-8})$ is $O(Z_k)$ in the end.

Finally, let us prove (26), for which we need to be more accurate. Roughly speaking, for Z_k and $\frac{X_k}{V_k}$ small, the dominant case is again $D_k = 2$ and $D_k^{(1)} = D_k^{(2)} = 1$, in which case $\Delta X_k = 0$. We will see that the second most likely cases are the case where $D_k = 2$ and the second neighbour of y has degree 3, and the case where $D_k = 3$ and the second and third neighbours of y have degree 2 each. In these two cases, we have respectively $\Delta X_k = -1$ and $\Delta X_k = 1$.

More precisely, let us write $\text{Var}(\Delta X_k | \mathcal{F}_k) = \mathbb{E}[(\Delta X_k)^2 | \mathcal{F}_k] - \mathbb{E}[\Delta X_k | \mathcal{F}_k]^2$. We first handle the second term using Proposition 1. Using $f(z) = \frac{z^2}{2} + O(z^3)$ and $\frac{v}{x+2v+s} = \frac{1}{2} + O(\frac{x}{v}) + O(\frac{s}{v})$, we have

$$\Phi_A(x, v, s) = O(z(v, s)) + O\left(\frac{x}{v}\right) + O\left(\frac{s}{v}\right) = O\left(\frac{x}{v}\right) + O(z(v, s))$$

by Lemma 5. Therefore, by Proposition 1, we have

$$\mathbb{E}[\Delta X_k | \mathcal{F}_k] = O\left(\frac{\log^2 n}{V_k}\right) + O\left(\frac{X_k}{V_k}\right) + O(Z_k),$$

so

$$\mathbb{E}[\Delta X_k | \mathcal{F}_k]^2 = O\left(\frac{\log^4 n}{V_k^2}\right) + O\left(\frac{X_k^2}{V_k^2}\right) + O(Z_k^2).$$

On the other hand, we denote by A_k^0 the event where y has degree 2 and its second neighbour as well (which implies $\Delta X_k = 0$). We also denote by A_k^1 the event where y has degree 3 and its neighbours which are not 1 are distinct and both have degree 2, which implies $\Delta X_k = 1$. Finally, let A_k^{-1} be the event where y has degree 2 and its second neighbour has degree 3, which implies $\Delta X_k = -1$. We have $\Delta X_k = O(D_k)$, so we can write

$$\mathbb{E}[(\Delta X_k)^2 | \mathcal{F}_k] = \mathbb{P}(A_k^1 | \mathcal{F}_k) + \mathbb{P}(A_k^{-1} | \mathcal{F}_k) + O\left(\mathbb{E}\left[D_k^2 \mathbb{1}_{(A_k^0 \cup A_k^1 \cup A_k^{-1})^c} | \mathcal{F}_k\right]\right). \quad (27)$$

As before, we can decompose according to the values of y and of its neighbours w_1, w_2 other than 1 to get

$$\begin{aligned} \mathbb{P}(A_k^1 | \mathcal{F}_k) &= \sum_{y, w_1, w_2 = X_k + 1}^{X_k + V_k} \mathbb{P}(\deg(y) = 3, \deg(w_1) = 2, \deg(w_2) = 2 | \mathcal{F}_k) \\ &\quad \times \frac{3}{H_k - 1} \frac{2}{H_k - 3} \frac{2}{H_k - 5}, \end{aligned}$$

where the sum is over pairwise distinct y, w_1, w_2 . Using Lemma 6 (more precisely (9) for $i_1 = 3$ and $i_2 = i_3 = 2$), this becomes

$$\begin{aligned} \mathbb{P}(A_k^1 | \mathcal{F}_k) &= \frac{12V_k(V_k - 1)(V_k - 2)}{(H_k - 1)(H_k - 3)(H_k - 5)} \frac{Z_k^7}{24f(Z_k)^3} \left(1 + O\left(\frac{\log^2 n}{V_k Z_k}\right)\right) \\ &= \frac{1}{2} Z_k \left(1 + O\left(\frac{1}{V_k}\right) + O(Z_k) + O\left(\frac{X_k}{H_k}\right) + O\left(\frac{\log^2 n}{V_k Z_k}\right)\right), \end{aligned}$$

where the second equality uses $\frac{V_k}{H_k} = \frac{1}{2} + O\left(\frac{X_k}{H_k}\right) + O(Z_k)$ and $f(Z_k) = \frac{1}{2} Z_k^2 + O(Z_k^3)$.

Similarly, using Lemma (6) for $i_1 = 2$ and $i_2 = 3$, we obtain

$$\begin{aligned}\mathbb{P}\left(A_k^{-1}|\mathcal{F}_k\right) &= \sum_{\substack{y,w=X_k+1 \\ y \neq w}}^{X_k+V_k} \mathbb{P}(\deg(y) = 2, \deg(w) = 3|\mathcal{F}_k) \times \frac{2}{H_k-1} \frac{3}{H_k-3} \\ &= \frac{6V_k(V_k-1)}{(H_k-1)(H_k-3)} \frac{Z_k^5}{12f(Z_k)^2} \left(1 + O\left(\frac{\log^2 n}{V_k Z_k}\right)\right) \\ &= \frac{1}{2}Z_k \left(1 + O\left(\frac{1}{V_k}\right) + O(Z_k) + O\left(\frac{X_k}{V_k}\right) + O\left(\frac{\log^2 n}{V_k Z_k}\right)\right).\end{aligned}$$

Finally, it remains to show that the last term of (27) is small. We decompose this term according to the value of D_k :

- the case $D_k = 1$ has probability $\frac{X_k-1}{H_k-1}$ and implies $\Delta X_k = -2$, so its contribution is $O\left(\frac{X_k}{V_k}\right)$.
- If $D_k = 2$, then the second neighbour of y has degree either 1, or at least 4 (if not, either A_k^0 or A_k^{-1} occurs). The first case has probability $O\left(\frac{X_k}{H_k}\right)$ and implies $\Delta X_k = -2$, so its contribution is $O\left(\frac{X_k}{V_k}\right)$. In the second case we have $\Delta X_k = -1$, so its contribution is the probability that the second neighbour w of y has degree at least 4. By summing over the possible values of w and its degree, this is

$$\begin{aligned}\sum_{d \geq 4} \sum_{y,w=X_k+1}^{X_k+V_k} \mathbb{P}(\deg(y) = 2 \text{ and } \deg(w) = d|\mathcal{F}_k) \times \frac{2}{H_k-1} \frac{d}{H_k-3} &= O(n^{-8}) + O\left(\sum_{d=4}^{\log n} \frac{dZ_k^d}{d!f(Z_k)}\right) \\ &= O(n^{-8}) + O(Z_k^2)\end{aligned}$$

by bounding Z_k^d by $Z_k^4 C^{d-4}$.

- By a similar computation as in the previous cases, the contribution of the case $D_k \geq 4$ is bounded by

$$\begin{aligned}\mathbb{E}\left[D_k^2 \mathbb{1}_{D_k \geq 4}|\mathcal{F}_k\right] &= O(n^{-8}) + \sum_{d=4}^{\log n} \sum_{y=X_k+1}^{X_k+V_k} d^2 \mathbb{P}(\deg(y) = d|\mathcal{F}_k) \times \frac{d}{H_k-1} \\ &= O(n^{-8}) + O\left(\frac{V_k}{H_k} \sum_{d=4}^{\log n} d^3 \frac{Z_k^d}{d!f(Z_k)}\right) \\ &= O(n^{-8}) + O\left(\sum_{d=4}^{\log n} \frac{d^3}{d!} C^{d-4} Z_k^2\right) \\ &= O(n^{-8}) + O(Z_k^2).\end{aligned}$$

- Finally, if $D_k = 3$, then either y has a neighbour other than 1 and y with degree $\neq 2$, or it is linked twice to the same vertex of degree 2, or it is linked to itself.
 - The contribution of the case where y has a neighbour w other than 1 and y with degree 1 is bounded (up to a multiplicative constant) by

$$\mathbb{P}\left(D_k = 3 \text{ and } D_k^{(1)} \geq 2|\mathcal{F}_k\right) \leq \sum_{y=X_k+1}^{X_k+V_k} \sum_{w=2}^{X_k} \mathbb{P}(\deg(y) = 3|\mathcal{F}_k) \times \frac{3}{H_k-1} \times \frac{2}{H_k-3},$$

where the factor 2 in the numerator accounts for the 2 half-edges going out from y that are not directed toward 1. Using Lemma 6, this is $O\left(\frac{X_k V_k}{H_k^2} \frac{Z_k^3}{f(Z_k)}\right) = O\left(\frac{X_k}{V_k}\right)$.

- Similarly, the contribution of the case where y has a neighbour w other than 1 and y with degree $d \geq 3$ is bounded (up to a multiplicative constant) by

$$\begin{aligned} & \mathbb{E} \left[D_k^2 \mathbb{1}_{D_k=3 \text{ and } D_k^{(3)} \geq 1} | \mathcal{F}_k \right] \\ & \leq \sum_{d \geq 3} \sum_{\substack{y, w = X_k + 1 \\ y \neq w}}^{X_k + V_k} \mathbb{P}(\deg(y) = 3 \text{ and } \deg(w) = d | \mathcal{F}_k) \times \frac{3}{H_k - 1} \times \frac{2d}{H_k - 3}, \end{aligned}$$

where the factor 2 in the numerator again accounts for the 2 half-edges going out from y that are not directed to 1. Using Lemma 6 in the same way as before, we find that this contribution is $O(Z_k^2)$ (with one factor Z_k coming from y and another one from w).

- If y is linked twice to the same vertex of degree 2 or to itself, we have $\Delta X_k = O(1)$ so the corresponding contribution is bounded (up to a multiplicative constant) by the probability that y is incident to a loop or a multiple edge. By (18), this is $O\left(\frac{1}{V_k}\right)$.

□

3 Fluid limit of the Markov chain

3.1 Solution to the differential equations

Following [1], let us investigate the solution to the differential equations which appear as the limit in Proposition 1. We recall from 12, 13 and 14 the definition of Φ_A , Φ_B and Φ_C . Let $(\mathcal{X}(t), \mathcal{V}(t), \mathcal{S}(t))$ be the solution of the system

$$\begin{cases} \mathcal{X}' &= \Phi_A(\mathcal{X}, \mathcal{V}, \mathcal{S}), \\ \mathcal{V}' &= \Phi_B(\mathcal{X}, \mathcal{V}, \mathcal{S}), \\ \mathcal{S}' &= \Phi_C(\mathcal{X}, \mathcal{V}, \mathcal{S}), \end{cases}$$

with the initial conditions (coming from Lemma 1)

$$\begin{cases} \mathcal{X}(0) &= e^{1-e}, \\ \mathcal{V}(0) &= 1 - e^{-e} - e^{1-e}, \\ \mathcal{S}(0) &= e + e^{1-e} + 2e^{-e} - 2. \end{cases} \quad (28)$$

The solution is given by [1, Lemma 8] with $c = e$, where v_1 , v and $2m$ in [1] stand respectively for \mathcal{X} , \mathcal{V} and $\mathcal{X} + 2\mathcal{V} + \mathcal{S}$. We write $\mathcal{Z}(t) = z(\mathcal{V}(t), \mathcal{S}(t))$, where $z(v, s)$ is given by (2), and we recall that $f(z) = e^z - z - 1$. For $z \geq 0$, let $\beta(z)$ be the unique solution in $[e^{-1}, +\infty)$ to the equation

$$\beta(z)e^{e\beta(z)} = e^z, \quad (29)$$

and note that $\beta(0) = e^{-1}$. Parametrized by \mathcal{Z} , the solutions can be written

$$\mathcal{V} = \beta(\mathcal{Z})e^{-\mathcal{Z}}f(\mathcal{Z}), \quad (30)$$

$$\mathcal{X} = e^{-1}\mathcal{Z}^2 - \mathcal{Z}\beta(\mathcal{Z})(1 - e^{-\mathcal{Z}}), \quad (31)$$

$$\mathcal{S} = \beta(\mathcal{Z})\left(\mathcal{Z} + \mathcal{Z}e^{-\mathcal{Z}} + 2e^{-\mathcal{Z}} - 2\right), \quad (32)$$

$$t = 1 - \beta(\mathcal{Z}) - \frac{1}{2e}\log^2\beta(\mathcal{Z}). \quad (33)$$

In [1], it is proved in the subcritical and supercritical cases that the Markov chain (X_k, V_k, S_k) , once properly rescaled, converges to the process $(\mathcal{X}, \mathcal{V}, \mathcal{S})$. In the critical case, such a result will follow from the results of Section 5.

3.2 The fluid limit near the extinction time

As in [6], we will now focus on carefully studying the solutions $\mathcal{X}, \mathcal{V}, \mathcal{S}$ near the extinction time $t^* = \inf\{t \geq 0 \mid \mathcal{X}(t) = 0\}$. We first notice that such a time exists, since plugging $\mathcal{Z} = 0$ in (33) and (31) gives $t = 1 - \frac{3}{2e} > 0$ and $\mathcal{X} = 0$.

We now argue that we indeed have $t^* = 1 - \frac{3}{2e}$. For this, let $z^* = \mathcal{Z}(t^*)$. We will show that $z^* = 0$. Indeed, since $\mathcal{X}(t^*) = 0$, the formula (31) gives

$$e^{-1}(z^*)^2 - z^* \beta^* (1 - e^{-z^*}) = 0,$$

where $\beta^* = \beta(z^*)$. Hence, we have either $z^* = 0$, or $e^{-1}z^* - \beta^*(1 - e^{-z^*}) = 0$. In the second case, we replace z^* by $e\beta^* + \log \beta^*$ using (29). The condition becomes

$$\frac{1}{e} \log \beta^* + e^{-e\beta^*} = 0.$$

We note that $\beta^* = e^{-1}$ (which implies $z^* = 0$) is a solution, and claim that the left-hand side is increasing in β , so the solution is unique. Indeed, differentiating the left-hand side with respect to β , we obtain $\frac{1}{e\beta} - e^{1-e\beta}$. We will prove that this is positive for all $\beta \neq e^{-1}$, which is equivalent (via the change of variable $y = e\beta$) to showing $ye^{1-y} < 1$ for all $y \neq 1$. This last assertion is immediate by differentiating again with respect to y , which proves $\beta^* = e^{-1}$ and $z^* = 0$. Using (33), we get $t^* = 1 - \frac{3}{2e}$.

Moreover, by differentiating (33) twice, we find that t is a smooth function of β with

$$\left. \frac{dt}{d\beta} \right|_{\beta=\beta^*} = 0, \quad \left. \frac{d^2t}{d\beta^2} \right|_{\beta=\beta^*} = -2e \quad \text{and} \quad \left. \frac{d^3t}{d\beta^3} \right|_{\beta=\beta^*} = 5e^2, \quad (34)$$

so $t^* - t = e(\beta - \beta^*)^2 - \frac{5}{6}e^2(\beta - \beta^*)^3 + O((\beta - \beta^*)^4)$ as $t \rightarrow t^*$ (or equivalently $\beta \rightarrow \beta^*$). That is, if we write $t = t^* - \varepsilon$, we have

$$\beta = e^{-1} + e^{-1/2}\varepsilon^{1/2} + \frac{5}{12}\varepsilon + O(\varepsilon^{3/2}). \quad (35)$$

By definition of β , we have $\left. \frac{dz}{d\beta} \right|_{\beta=\beta^*} = 2e$, so

$$\mathcal{Z}(t^* - \varepsilon) \underset{\varepsilon \rightarrow 0}{\sim} 2e(\beta(\mathcal{Z}(t^* - \varepsilon)) - \beta^*) \underset{\varepsilon \rightarrow 0}{\sim} 2e^{1/2}\varepsilon^{1/2}. \quad (36)$$

We will need to estimate precisely the processes $\mathcal{X}, \mathcal{V}, \mathcal{Z}$ and their derivatives near time t^* . For this, we note that by replacing z by $e\beta + \log \beta$ in (30), (31), (32), these processes are explicit functions of β . In particular \mathcal{X} is a smooth function of β and we can compute

$$\left. \frac{d\mathcal{X}}{d\beta} \right|_{\beta=\beta^*} = \left. \frac{d^2\mathcal{X}}{d\beta^2} \right|_{\beta=\beta^*} = \left. \frac{d^3\mathcal{X}}{d\beta^3} \right|_{\beta=\beta^*} = 0 \quad \text{and} \quad \left. \frac{d^4\mathcal{X}}{d\beta^4} \right|_{\beta=\beta^*} = 8e^3, \quad (37)$$

so $\mathcal{X}(t^* - \varepsilon) \sim \frac{e^3}{3}(\beta - \beta^*)^4$, i.e.

$$\mathcal{X}(t^* - \varepsilon) \underset{\varepsilon \rightarrow 0}{\sim} \frac{e}{3}\varepsilon^2. \quad (38)$$

Similarly \mathcal{V} is a smooth function of β with

$$\left. \frac{d\mathcal{V}}{d\beta} \right|_{\beta=\beta^*} = 0 \text{ and } \left. \frac{d^2\mathcal{V}}{d\beta^2} \right|_{\beta=\beta^*} = 4e \text{ and } \left. \frac{d^3\mathcal{V}}{d\beta^3} \right|_{\beta=\beta^*} = -10e^2, \quad (39)$$

so, using also (35):

$$\mathcal{V}(t^* - \varepsilon) = 2e(\beta - \beta^*)^2 - \frac{5}{3}e^2(\beta - \beta^*)^3 + O((\beta - \beta^*)^4) = 2\varepsilon + O(\varepsilon^2). \quad (40)$$

We finally do the same computation for \mathcal{S} : this is also a smooth function of β with

$$\left. \frac{d\mathcal{S}}{d\beta} \right|_{\beta=\beta^*} = \left. \frac{d^2\mathcal{S}}{d\beta^2} \right|_{\beta=\beta^*} = 0 \text{ and } \left. \frac{d^3\mathcal{S}}{d\beta^3} \right|_{\beta=\beta^*} = 8e^2, \quad (41)$$

so

$$\mathcal{S}(t^* - \varepsilon) = \frac{4}{3}e^2(\beta - \beta^*)^3 + O((\beta - \beta^*)^4) = \frac{4e^{1/2}}{3}\varepsilon^{3/2} + O(\varepsilon^2) \quad (42)$$

by using (35).

Remark. As mentioned before, the exponents governing the number of vertices of degrees 1, 2 and 3 are the same as in the configuration model of [6]. Moreover, the multiplicative constant 2 appearing in (40) is the same as in [6] and means that in the end, most steps simply consist of removing two vertices of degree 2. However, we note that the multiplicative constants for \mathcal{X} and \mathcal{S} are significantly different from their analogs in [6]. This is due to the impact of vertices of degree 4. For example, a step where the removed leaf has a neighbour of degree 4 attached to three other vertices of degree 2 will create three new leaves. By (38), the (normalized) number of leaves at time $t^* - \varepsilon$ behaves like ε^2 . On the other hand, by combining the Poisson approximation of Lemma 6 and (36), the number of vertices of degree 4 behaves like $\mathcal{V} \times \mathcal{L}^2 \approx \varepsilon^2$. Hence, the number of vertices of degree 4 is comparable to the number of leaves in the end of the process, which is why their contribution cannot be neglected.

When we will estimate the drift of the Markov chain (X_k, V_k, S_k) , it will be important to understand the functions Φ_A, Φ_B, Φ_C of (12), (13), (14) in the neighbourhood of the fluid limit $(\mathcal{X}, \mathcal{V}, \mathcal{S})$. For this, we will rely on the following estimates on the partial derivatives of Φ_A, Φ_B, Φ_C . For the sake of brevity, we will write $(\mathcal{X}, \mathcal{V}, \mathcal{S}(t))$ for $(\mathcal{X}(t), \mathcal{V}(t), \mathcal{S}(t))$.

Lemma 7. *We have the following estimates as $\varepsilon \rightarrow 0$:*

$$\begin{aligned}
\frac{\partial \Phi_A}{\partial x}(\mathcal{X}, \mathcal{V}, \mathcal{S}(t^* - \varepsilon)) &= -\varepsilon^{-1} + O(\varepsilon^{-1/2}), \\
\frac{\partial \Phi_A}{\partial v}(\mathcal{X}, \mathcal{V}, \mathcal{S}(t^* - \varepsilon)) &= O(1), \\
\frac{\partial \Phi_A}{\partial s}(\mathcal{X}, \mathcal{V}, \mathcal{S}(t^* - \varepsilon)) &= O(\varepsilon^{-1/2}), \\
\frac{\partial \Phi_B}{\partial x}(\mathcal{X}, \mathcal{V}, \mathcal{S}(t^* - \varepsilon)) &= O(\varepsilon^{-1}), \\
\frac{\partial \Phi_B}{\partial v}(\mathcal{X}, \mathcal{V}, \mathcal{S}(t^* - \varepsilon)) &= O(1), \\
\frac{\partial \Phi_B}{\partial s}(\mathcal{X}, \mathcal{V}, \mathcal{S}(t^* - \varepsilon)) &= O(\varepsilon^{-1/2}), \\
\frac{\partial \Phi_C}{\partial x}(\mathcal{X}, \mathcal{V}, \mathcal{S}(t^* - \varepsilon)) &= O(\varepsilon^{-1/2}), \\
\frac{\partial \Phi_C}{\partial v}(\mathcal{X}, \mathcal{V}, \mathcal{S}(t^* - \varepsilon)) &= e^{1/2} \varepsilon^{-1/2} + O(1), \\
\frac{\partial \Phi_C}{\partial s}(\mathcal{X}, \mathcal{V}, \mathcal{S}(t^* - \varepsilon)) &= -\frac{3}{2} \varepsilon^{-1} + O(\varepsilon^{-1/2}).
\end{aligned}$$

In particular, the fact that $\frac{\partial \Phi_A}{\partial x}$ is negative means that (X_k) enjoys a strong “restoring force” towards its fluid limit, which will play an important role later in the paper.

Proof. The proof is a calculation that we perform using Mathematica (see the attached computation sheet). More precisely, as an example, let us sketch the proof of the estimate on $\frac{\partial \Phi_A}{\partial s}$. We recall from (12) the formula

$$\Phi_A(x, v, s) = -1 - \frac{x}{x + 2v + s} + \frac{v^2 z(v, s)^4 e^{z(v, s)}}{(x + 2v + s)^2 f(z(v, s))^2} - \frac{xvz(v, s)^2 e^{z(v, s)}}{(x + 2v + s)^2 f(z(v, s))},$$

where we recall that $z(v, s)$ is given by (2). We can rewrite Φ_A as $\Phi_A = \Psi_A \circ \varphi$, where $\varphi(x, v, s) = (x, v, s, z(v, s))$ and

$$\Psi_A(x, v, s, z) = -1 - \frac{x}{x + 2v + s} + \frac{v^2 z^4 e^z}{(x + 2v + s)^2 f(z)^2} - \frac{xvz^2 e^z}{(x + 2v + s)^2 f(z)}.$$

We then have

$$\frac{\partial \Phi_A}{\partial s} = \frac{\partial \Psi_A}{\partial s} + \frac{\partial \Psi_A}{\partial z} \times \frac{\partial z}{\partial s}. \tag{43}$$

Moreover, by (2), we can write $s = v \left(\frac{z(e^z - 1)}{e^z - z - 1} - 2 \right)$, so

$$\frac{\partial z}{\partial s} = \left(\frac{\partial}{\partial z} \left(v \left(\frac{z(e^z - 1)}{e^z - z - 1} - 2 \right) \right) \right)^{-1}.$$

Therefore (43) can be expressed as a completely explicit function of x, v, s and $z(v, s)$. We can finally replace $(x, v, s, z(v, s))$ in this expression by $(\mathcal{X}(t^* - \varepsilon), \dots, \mathcal{Z}(t^* - \varepsilon))$. By the explicit expressions (30) to (32), this is an explicit function of $\beta(\mathcal{Z})$. We can then expand this as a power series in $\beta - \beta^*$, and finally translate the result to an expansion in ε using (35), which gives the third equation of Lemma 7. The other eight estimates are proved in the exact same way. We have attached a Mathematica script that performs these calculations. \square

Lemma 8. We denote by $H\Phi(x, v, s)$ the Hessian matrix of a function Φ at the point (x, v, s) . Let $C > 0$ be a constant and assume $x, s \leq Cv$. Then we have

$$\|H\Phi_A(x, v, s)\|, \|H\Phi_B(x, v, s)\|, \|H\Phi_C(x, v, s)\| = O\left(\frac{1}{v^2}\right),$$

where the implied constant only depends on C . Moreover, we have

$$\frac{\partial^2 \Phi_A}{\partial v^2} = O\left(\frac{x}{v^3}\right) + O\left(\frac{s^2}{v^4}\right), \quad (44)$$

$$\frac{\partial^2 \Phi_A}{\partial v \partial s} = O\left(\frac{x}{v^3}\right) + O\left(\frac{s}{v^3}\right), \quad (45)$$

$$\frac{\partial^2 \Phi_C}{\partial v^2} = O\left(\frac{x}{v^3}\right) + O\left(\frac{s}{v^3}\right). \quad (46)$$

Later, as suggested by (38), (40) and (42), we will typically apply these estimates in a regime where x , v and s are respectively of order ε^2 , ε and $\varepsilon^{3/2}$. In this regime (44) becomes $O(\varepsilon^{-1})$, whereas (45) and (46) become $O(\varepsilon^{-3/2})$ and all the other second order partial derivatives are $O(\varepsilon^{-2})$.

Proof. Let us first prove that all the second order partial derivatives are $O\left(\frac{1}{v^2}\right)$. We write $\bar{s} = \frac{s}{v}$ and $\bar{x} = \frac{x}{v}$, and we recall from Lemma 5 that z is a smooth function of \bar{s} . It follows from the definition (12) of Φ_A that we can write $\Phi_A(x, v, s) = \bar{\Phi}_A\left(\frac{x}{v}, \frac{s}{v}\right)$ where $\bar{\Phi}_A$ is smooth on $[0, +\infty)^2$, and similarly for Φ_B and Φ_C . Therefore, we can write down

$$\frac{\partial^2 \Phi_A}{\partial x^2}(x, v, s) = \frac{1}{v^2} \frac{\partial^2 \bar{\Phi}_A}{\partial \bar{x}^2}\left(\frac{x}{v}, \frac{s}{v}\right) = O\left(\frac{1}{v^2}\right)$$

since $\bar{\Phi}_A$ is smooth and $x, s = O(v)$, so the partial derivatives of $\bar{\Phi}_A$ are $O(1)$. The same argument applies to the partial derivatives $\frac{\partial^2 \Phi_A}{\partial x \partial s}$ and $\frac{\partial^2 \Phi_A}{\partial s^2}$. Similarly, we have

$$\frac{\partial^2 \Phi_A}{\partial v \partial s}(x, v, s) = -\frac{1}{v^2} \frac{\partial \bar{\Phi}_A}{\partial \bar{s}}\left(\frac{x}{v}, \frac{s}{v}\right) - \frac{s}{v^3} \frac{\partial^2 \bar{\Phi}_A}{\partial \bar{x}^2}\left(\frac{x}{v}, \frac{s}{v}\right) - \frac{x}{v^3} \frac{\partial^2 \bar{\Phi}_A}{\partial \bar{x} \partial \bar{s}}\left(\frac{x}{v}, \frac{s}{v}\right) = O\left(\frac{1}{v^2}\right) \quad (47)$$

using the assumption $x, s = O(v)$, and the same computation applies to $\frac{\partial^2 \Phi_A}{\partial x \partial v}$. Finally, we have

$$\frac{\partial^2 \Phi_A}{\partial v^2} = 2 \frac{x}{v^3} \frac{\partial \bar{\Phi}_A}{\partial \bar{x}} + 2 \frac{s}{v^3} \frac{\partial \bar{\Phi}_A}{\partial \bar{s}} + \frac{x^2}{v^4} \frac{\partial^2 \bar{\Phi}_A}{\partial \bar{x}^2} + 2 \frac{xs}{v^4} \frac{\partial^2 \bar{\Phi}_A}{\partial \bar{x} \partial \bar{s}} + \frac{s^2}{v^4} \frac{\partial^2 \bar{\Phi}_A}{\partial \bar{s}^2}, \quad (48)$$

where all the partial derivatives of $\bar{\Phi}_A$ are taken at $\left(\frac{x}{v}, \frac{s}{v}\right)$. Again, the partial derivatives are all $O(1)$, so this is $O\left(\frac{1}{v^2}\right)$, and the same argument applies to Φ_B and Φ_C .

We now prove the three estimates (44), (45) and (46). To show (46), we just need to write down the analog of (48) for Φ_C instead of Φ_A . As before, all the partial derivatives of $\bar{\Phi}_C$ are $O(1)$, so using $x, s = O(v)$, all terms are either $O\left(\frac{x}{v^3}\right)$ or $O\left(\frac{s}{v^3}\right)$.

Moreover, we notice that all the terms of (47) are $O\left(\frac{x}{v^3}\right)$ or $O\left(\frac{s}{v^3}\right)$, except perhaps $\frac{1}{v^2} \frac{\partial \bar{\Phi}_A}{\partial \bar{s}}\left(\frac{x}{v}, \frac{s}{v}\right)$. Similarly, all the terms of (48) are $O\left(\frac{x}{v^3}\right)$ or $O\left(\frac{s^2}{v^4}\right)$, except perhaps $2 \frac{s}{v^3} \frac{\partial \bar{\Phi}_A}{\partial \bar{s}}\left(\frac{x}{v}, \frac{s}{v}\right)$. Therefore, to prove both (44) and (45), it is sufficient to prove that for $\bar{s}, \bar{x} = O(1)$, we have

$$\frac{\partial \bar{\Phi}_A}{\partial \bar{s}}(\bar{x}, \bar{s}) = O(\bar{s} + \bar{x}).$$

Since $\bar{\Phi}_A$ is smooth, the partial derivative $\frac{\partial \bar{\Phi}_A}{\partial \bar{s}}$ is Lipschitz, so it is sufficient to show $\frac{\partial \bar{\Phi}_A}{\partial \bar{s}}(0, 0) = 0$, which is a direct computation that we do using Mathematica (see the second part of the attached computation sheet). \square

Finally, the last purely continuous estimate that we will need is on the second derivative of the fluid limit of our Markov chain.

Lemma 9. *As $\varepsilon \rightarrow 0$, we have*

$$\mathcal{X}''(t^* - \varepsilon) = O(1), \quad (49)$$

$$\mathcal{Y}''(t^* - \varepsilon) = O(\varepsilon^{-1/2}), \quad (50)$$

$$\mathcal{S}''(t^* - \varepsilon) = O(\varepsilon^{-1/2}). \quad (51)$$

Proof. By (31) (combined with the equation $z = e\beta(z) + \log \beta(z)$) and (33), both $\mathcal{X}(t)$ and t can be written as explicit functions of $\beta(\mathcal{X}(t))$. Therefore, we can write

$$\mathcal{X}'(t) = \left(\frac{dt}{d\beta} \right)^{-1} \times \frac{d\mathcal{X}}{d\beta}$$

and

$$\mathcal{X}''(t) = \left(\frac{dt}{d\beta} \right)^{-2} \times \frac{d^2\mathcal{X}}{d\beta^2} - \left(\frac{dt}{d\beta} \right)^{-3} \times \frac{d^2t}{d\beta^2} \times \frac{d\mathcal{X}}{d\beta}. \quad (52)$$

By (37), we have $\frac{d\mathcal{X}}{d\beta} = O((\beta - \beta^*)^3)$ and $\frac{d^2\mathcal{X}}{d\beta^2} = O((\beta - \beta^*)^2)$. On the other hand, by (34), we have $\frac{dt}{d\beta} \sim 2e(\beta - \beta^*)$ as $\beta \rightarrow \beta^*$ and $\frac{d^2t}{d\beta^2} = O(1)$. Combining all these estimates, we find that (52) is $O(1)$.

The proof of (50) and (51) is similar: we can write the same equation as (52) for \mathcal{Y} and \mathcal{S} . Using (34) and (39), we obtain

$$\begin{aligned} \left(\frac{dt}{d\beta} \right)^{-2} \times \frac{d^2\mathcal{Y}}{d\beta^2} &= e^{-1}(\beta - \beta^*)^{-2} + O((\beta - \beta^*)^{-1}), \\ \left(\frac{dt}{d\beta} \right)^{-3} \times \frac{d^2t}{d\beta^2} \times \frac{d\mathcal{Y}}{d\beta} &= e^{-1}(\beta - \beta^*)^{-2} + O((\beta - \beta^*)^{-1}), \end{aligned}$$

so $\mathcal{Y}''(t^* - \varepsilon) = O((\beta - \beta^*)^{-1}) = O(\varepsilon^{-1/2})$ by (35). For the estimate on \mathcal{S} , using (41), we find that both terms are $O((\beta - \beta^*)^{-1}) = O(\varepsilon^{-1/2})$, which proves (51). \square

4 Drift and variance estimates in the neighbourhood of the fluid limit

Our goal is now to combine the drift and variance estimates of Section 2 with the fluid limit computations of Section 3 to estimate the drift and variance of the Markov chain (X, V, S) in the neighbourhood of the fluid limit. While Propositions 1 and 2 consisted mostly of refining result from [1], the proofs here will be very similar to those of [6, Section 4.1].

Just like in [6], we will spend most of the rest of the paper studying the fluctuations of the three processes (X_k, V_k, S_k) around their fluid limit described in Section 3. We recall that θ^n is the first

time k such that $X_k = 0$ and that t^* is the first time t where $\mathcal{X}(t) = 0$. For $0 \leq k \leq \theta^n$, we define

$$\begin{cases} A_k = X_k - n\mathcal{X}\left(\frac{k}{n}\right), \\ B_k = V_k - n\mathcal{V}\left(\frac{k}{n}\right), \\ C_k = S_k - n\mathcal{S}\left(\frac{k}{n}\right). \end{cases}$$

Just like in [6], it will be crucial for us to understand the order of magnitude of the fluctuations A, B, C near the extinction time θ^n . For technical reasons, we will work with the stopping time

$$\tilde{\theta}^n := \theta^n \wedge \left(t^*n - n^{3/5-1/100}\right),$$

and we will show later that $\tilde{\theta}^n = \theta^n$ with high probability. To measure how close we are to the end, for all $0 \leq k \leq \tilde{\theta}^n$, we write $\varepsilon_k = \frac{t^*n-k}{n}$ so that $k = (t^* - \varepsilon_k)n$, and we note that $\varepsilon_k \geq n^{-2/5-1/100}$. For $0 \leq k \leq \tilde{\theta}^n$, we also define the rescaled fluctuations

$$\begin{cases} \tilde{A}_k = \frac{A_k}{\varepsilon_k^{3/4}\sqrt{n}}, \\ \tilde{B}_k = \frac{B_k}{\sqrt{n}}, \\ \tilde{C}_k = \frac{C_k}{\varepsilon_k^{1/2}\sqrt{n}}. \end{cases}$$

Note that the normalizations are the same as in [6], with C playing the same role as the fluctuations on the number of vertices of degree 3. One of the goals of Section 5 will be to prove that these rescaled fluctuations stay of order at most $\log n$ all along the process. For now, our goal is to show the following drift and variance estimates, which are the natural analogs of [6, Propositions 4 and 5].

Proposition 3 (Drift estimates). *For any $0 \leq k < \tilde{\theta}^n$, if we have $|\tilde{A}_k|, |\tilde{B}_k|, |\tilde{C}_k| < 1000 \log n$, then*

$$\mathbb{E} \left[\Delta \tilde{A}_k | \mathcal{F}_k \right] = -\frac{1}{4} \frac{1}{\varepsilon_k n} \tilde{A}_k + O \left(\frac{\varepsilon_k^{1/2}}{\varepsilon_k n} |\tilde{A}_k| \right) + O \left(\frac{\varepsilon_k^{1/4}}{\varepsilon_k n} \max(|\tilde{B}_k|, |\tilde{C}_k|) \right) + O \left(\frac{n^{-1/30}}{\varepsilon_k n} \right), \quad (53)$$

$$\mathbb{E} \left[\Delta \tilde{B}_k | \mathcal{F}_k \right] = O \left(\frac{\varepsilon_k^{3/4}}{\varepsilon_k n} \max(|\tilde{A}_k|, |\tilde{B}_k|, |\tilde{C}_k|) \right) + O \left(\frac{n^{-1/30}}{\varepsilon_k n} \right), \quad (54)$$

$$\mathbb{E} \left[\Delta \tilde{C}_k | \mathcal{F}_k \right] = \frac{1}{\varepsilon_k n} \left(\sqrt{\varepsilon_k} \tilde{B}_k - \tilde{C}_k \right) + O \left(\frac{\varepsilon_k^{1/2}}{\varepsilon_k n} \max(|\tilde{B}_k|, |\tilde{C}_k|) \right) + O \left(\frac{\varepsilon_k^{3/4}}{\varepsilon_k n} |\tilde{A}_k| \right) + O \left(\frac{n^{-1/30}}{\varepsilon_k n} \right), \quad (55)$$

where the constants implied by the O notation are absolute.

Note that we have displayed all the results with a denominator $\varepsilon_k n$. This will be convenient later to sum up the contributions of values of k at the same ‘‘scale’’, i.e. where k ranges from $(t^* - \varepsilon)n$ to $(t^* - \frac{\varepsilon}{2})n$ for some $\varepsilon > 0$. We also highlight that \tilde{A} exhibits a strong negative feedback, which will be very useful to guarantee that it stays small.

Proposition 4 (Variance estimates). *For any $0 \leq k < \tilde{\theta}^n$, if we have $|\tilde{A}_k|, |\tilde{B}_k|, |\tilde{C}_k| < 1000 \log n$,*

then

$$\text{Var} \left(\Delta \tilde{A}_k | \mathcal{F}_k \right) = \frac{2\sqrt{e}}{\varepsilon_k n} + O \left(\frac{n^{-1/30}}{\varepsilon_k n} \right) + O \left(\frac{\varepsilon_k^{1/2} n^{1/100}}{\varepsilon_k n} \right), \quad (56)$$

$$\text{Var} \left(\Delta \tilde{A}_k | \mathcal{F}_k \right) = O \left(\frac{1}{\varepsilon_k n} \right), \quad (57)$$

$$\text{Var} \left(\Delta \tilde{B}_k | \mathcal{F}_k \right) = O \left(\frac{\varepsilon_k}{\varepsilon_k n} \right), \quad (58)$$

$$\text{Var} \left(\Delta \tilde{C}_k | \mathcal{F}_k \right) = O \left(\frac{\varepsilon_k^{1/2}}{\varepsilon_k n} \right), \quad (59)$$

where the constants implied by the O notation are absolute.

During most of the process, we will only rely on the ‘‘crude’’ estimates (57), (58) and (59). We will only need the more accurate estimate (56) in the final phase when ε_k is of order $n^{-2/5}$ and the fluctuations A_k become comparable to X_k . In particular, in this regime, the error term $\frac{\varepsilon_k^{1/2} n^{1/100}}{\varepsilon_k n}$ will be much smaller than $\frac{1}{\varepsilon_k n}$.

As in [6, Section 4.1], we first state an easy lemma which will allow us to roughly estimate some denominators and error terms in the computations to prove Propositions 3 and 4.

Lemma 10. *There are absolute constants $C, c > 0$ such that for n large enough, if $0 \leq k < \tilde{\theta}^n$ satisfies $|\tilde{A}_k|, |\tilde{B}_k|, |\tilde{C}_k| < 1000 \log n$, then*

$$X_k \leq C \min \left(\varepsilon_k^2 n \times n^{1/100}, \varepsilon_k^{3/2} n \right), \quad c \varepsilon_k n \leq V_k \leq C \varepsilon_k n, \quad S_k \leq C \varepsilon_k^{3/2} n \quad (60)$$

and

$$c \varepsilon_k^{1/2} \leq Z_k \leq C \varepsilon_k^{1/2}. \quad (61)$$

The reason why we needed to write down two different upper bounds on X_k is that the first one is too crude for our purpose when ε_k is of order 1, whereas the second one is too crude when ε_k becomes very small. We note that, although the fluid limit approximation gives $X_k \approx \varepsilon_k^2 n$, it is not possible to write down $X_k = O(\varepsilon_k^2 n)$ in general. Indeed, the final regime that we will be interested in is precisely the regime where the fluctuations of X_k start to overwhelm the fluid limit.

Proof. Note that $k < \tilde{\theta}^n$ guarantees $\varepsilon_k \geq n^{-2/5-1/100}$. The proofs of the estimates on V_k, S_k and of $X_k \leq C \varepsilon_k^2 n \times n^{1/100}$ are exactly the same as for [6, Lemma 1], with V_k (resp. S_k) playing the same role as the Y_k (resp. Z_k) of [6]. Moreover, we can write

$$X_k = n \mathcal{X} \left(\frac{k}{n} \right) + \varepsilon_k^{3/4} \sqrt{n} \tilde{A}_k = O \left(\varepsilon_k^2 n \right) + O \left(\varepsilon_k^{3/4} \sqrt{n} \log n \right),$$

and the second term is $O(\varepsilon_k^{3/2} n)$ because $\varepsilon_k \geq n^{-2/5-1/100}$.

The only bounds left to prove are those on Z_k . We first recall from (2) the definition of Z_k :

$$\frac{Z_k(e^{Z_k} - 1)}{e^{Z_k} - Z_k - 1} = 2 + \frac{S_k}{V_k}.$$

By (60), the right-hand side is bounded by an absolute constant, and so is Z_k . Moreover, the left-hand side is $2 + \frac{1}{3} Z_k + O(Z_k^2)$, so there are two absolute constants c, C such that

$$c \frac{S_k}{V_k} \leq Z_k \leq C \frac{S_k}{V_k}. \quad (62)$$

By definition of C_k , we have $S_k = C_k + n\mathcal{S}\left(\frac{k}{n}\right)$, which is of order $\varepsilon_k^{3/2}n$ by (42) and the assumption on $|\tilde{C}_k|$. On the other hand, by (60) the denominator of (62) is of order $\varepsilon_k n$ and the bounds on Z_k follow. \square

Proof of Proposition 3. Before starting the proof, we notice that the functions Φ_A, Φ_B, Φ_C are homogeneous, so we can rewrite Proposition 1 as

$$\mathbb{E}[\Delta X_k | \mathcal{F}_k] = \Phi_A\left(\frac{X_k}{n}, \frac{V_k}{n}, \frac{S_k}{n}\right) + O\left(\frac{\log^2 n}{V_k}\right), \quad (63)$$

$$\mathbb{E}[\Delta V_k | \mathcal{F}_k] = \Phi_B\left(\frac{X_k}{n}, \frac{V_k}{n}, \frac{S_k}{n}\right) + O\left(\frac{\log^2 n}{V_k Z_k}\right),$$

$$\mathbb{E}[\Delta S_k | \mathcal{F}_k] = \Phi_C\left(\frac{X_k}{n}, \frac{V_k}{n}, \frac{S_k}{n}\right) + O\left(\frac{\log^2 n}{V_k Z_k}\right).$$

We also notice that under the assumptions of Proposition 3, the assumptions of Proposition 1 are satisfied for some absolute constant $C > 0$. Indeed, we have $X_k > 0$ because $k < \tilde{\theta}^n$ and Lemma 10 gives $Z_k = O(1)$ and $V_k Z_k \geq c\varepsilon_k^{3/2}n \geq \log^3 n$ if n is large enough because $\varepsilon_k \geq n^{-2/5-1/100}$.

Let us now start with the drift estimate on \tilde{A} , which is the most complex one. As in [6], the left-hand side of (53) can be decomposed as

$$\begin{aligned} & \left| \mathbb{E}[\Delta \tilde{A}_k | \mathcal{F}_k] + \frac{1}{\varepsilon_k n} \frac{1}{4} \tilde{A}_k \right| \\ & \leq \left| \left(\frac{\varepsilon_k^{3/4}}{\varepsilon_{k+1}^{3/4}} - 1 \right) \tilde{A}_k - \frac{3}{4} \frac{1}{\varepsilon_k n} \tilde{A}_k \right| \end{aligned} \quad (64)$$

$$+ \frac{1}{\varepsilon_{k+1}^{3/4} \sqrt{n}} \left| \mathbb{E}[\Delta X_k | \mathcal{F}_k] - \Phi_A\left(\frac{X_k}{n}, \frac{V_k}{n}, \frac{S_k}{n}\right) \right| \quad (65)$$

$$+ \frac{1}{\varepsilon_{k+1}^{3/4} \sqrt{n}} \left| \Phi_A\left(\frac{X_k}{n}, \frac{V_k}{n}, \frac{S_k}{n}\right) - \Phi_A\left(\mathcal{X}, \mathcal{V}, \mathcal{S}\left(\frac{k}{n}\right)\right) - \left(\frac{A_k}{n}, \frac{B_k}{n}, \frac{C_k}{n}\right) \cdot \nabla \Phi_A\left(\mathcal{X}, \mathcal{V}, \mathcal{S}\left(\frac{k}{n}\right)\right) \right| \quad (66)$$

$$+ \frac{1}{\varepsilon_{k+1}^{3/4} \sqrt{n}} \left| \left(\frac{A_k}{n}, \frac{B_k}{n}, \frac{C_k}{n}\right) \cdot \nabla \Phi_A\left(\mathcal{X}, \mathcal{V}, \mathcal{S}\left(\frac{k}{n}\right)\right) - \left(-\frac{A_k}{\varepsilon_k n}\right) \right| \quad (67)$$

$$+ \frac{1}{\varepsilon_{k+1}^{3/4} \sqrt{n}} \left| \Phi_A\left(\mathcal{X}, \mathcal{V}, \mathcal{S}\left(\frac{k}{n}\right)\right) - n \left(\mathcal{X}\left(\frac{k+1}{n}\right) - \mathcal{X}\left(\frac{k}{n}\right) \right) \right|. \quad (68)$$

We will handle each of those terms one by one. Just like in [6], the first term (64) is $O\left(\frac{\tilde{A}_k}{(\varepsilon_k n)^2}\right)$, so it is $O\left(\frac{n^{-1/30}}{\varepsilon_k n}\right)$. By² Proposition 1, the second term (65) is $O\left(\frac{1}{\varepsilon_k^{3/4} \sqrt{n}} \times \frac{\log^2 n}{V_k}\right)$. Using Lemma 10 and $\varepsilon_k \geq n^{-2/5-1/100}$, this is $O\left(\frac{n^{-1/30}}{\varepsilon_k n}\right)$.

We move on to the third term (66): as in [6], this is a linear approximation of Φ_A near $\left(\mathcal{X}, \mathcal{V}, \mathcal{S}\left(\frac{k}{n}\right)\right)$, so the bound relies on bounding second derivatives of Φ_A , which is done in

²Note that this is where we need the improved error term that we obtained in (15). The bound provided by [1] would not be sufficient for our purpose here, but it will be enough for the analog estimates on the drifts of \tilde{B} and \tilde{C} .

Lemma 8. More precisely (66) is bounded by

$$\frac{1}{\varepsilon_{k+1}^{3/4} \sqrt{n}} \sum_{1 \leq i, j \leq 3} |w_i| \times |w_j| \times \max_{\substack{|u_1 - \mathcal{X}(k/n)| \leq |w_1| \\ |u_2 - \mathcal{Y}(k/n)| \leq |w_2| \\ |u_3 - \mathcal{S}(k/n)| \leq |w_3|}} \left| \frac{\partial^2 \Phi_A}{\partial x_i \partial x_j} (u_1, u_2, u_3) \right|, \quad (69)$$

where we wrote $(w_1, w_2, w_3) = \left(\frac{A_k}{n}, \frac{B_k}{n}, \frac{C_k}{n} \right)$ for the sake of brevity and u_1, u_2, u_3 stand respectively for the variables x, v, s . By the assumption $|\tilde{A}|, |\tilde{B}|, |\tilde{C}| < 1000 \log n$, we have the bounds

$$|w_1| \leq 1000 \varepsilon_k^{3/4} \frac{\log n}{\sqrt{n}}, \quad |w_2| \leq 1000 \frac{\log n}{\sqrt{n}}, \quad |w_3| \leq 1000 \varepsilon_k^{1/2} \frac{\log n}{\sqrt{n}}. \quad (70)$$

On the other hand, we can use Lemma 8 to bound one by one the 9 partial derivatives, and handle the estimates given by Lemma 8 using Lemma 10. More precisely, by Lemma 10, we know that u_1, u_2, u_3 are respectively $O\left(\varepsilon_k^2 \times n^{1/100}\right)$, of order $\varepsilon_k n$ and $O\left(\varepsilon_k^{3/2} \times n^{1/100}\right)$. Therefore, the terms $O\left(\frac{1}{v^2}\right)$ in the bounds of Lemma 8 become $O(\varepsilon_k^{-2} \times n^{1/50})$. Similarly, the terms $O\left(\frac{s}{v^3}\right)$, $O\left(\frac{x}{v^3}\right)$ and $O\left(\frac{s^2}{v^4}\right)$ are respectively $O(\varepsilon_k^{-3/2} \times n^{1/50})$, $O(\varepsilon_k^{-1} \times n^{1/50})$ and $O(\varepsilon_k^{-1} \times n^{1/50})$. Combining these bounds with (70), we can bound one by one the terms of (69). For example, the term $i = j = 2$ is

$$\frac{1}{\varepsilon_{k+1}^{3/4} \sqrt{n}} \times O\left(\frac{\log n}{\sqrt{n}}\right) \times O\left(\frac{\log n}{\sqrt{n}}\right) \times O\left(\varepsilon_k^{-1} \times n^{1/50}\right) = O\left(\frac{\varepsilon_k^{-3/4} n^{-1/2} n^{1/50} \log^2 n}{\varepsilon_k n}\right).$$

Using the assumption $\varepsilon_k \geq n^{-2/5-1/100}$, we find that the numerator is $O\left(n^{-1/30}\right)$, so this term is $O\left(\frac{n^{-1/30}}{\varepsilon_k n}\right)$. By a similar computation, we prove the same thing for all terms of (69).

For the fourth term (67), we first write

$$\left(\frac{A_k}{n}, \frac{B_k}{n}, \frac{C_k}{n} \right) \cdot \nabla \Phi_A \left(\mathcal{X}, \mathcal{Y}, \mathcal{S} \left(\frac{k}{n} \right) \right) = \frac{A_k}{n} \frac{\partial \Phi_A}{\partial x} + \frac{B_k}{n} \frac{\partial \Phi_A}{\partial v} + \frac{C_k}{n} \frac{\partial \Phi_A}{\partial s}, \quad (71)$$

where the partial derivatives are always considered at the point $\left(\mathcal{X}, \mathcal{Y}, \mathcal{S} \left(\frac{k}{n} \right) \right)$. We then use the first three equations of Lemma 7 to estimate the partial derivatives of Φ_A . In particular, by the first equation of Lemma 7, the first term is $-\frac{A_k}{\varepsilon_k n} + O\left(\frac{\varepsilon_k^{1/2} |A_k|}{\varepsilon_k n}\right)$. Similarly, by the second equation of Lemma 7, the second term of (71) is $O\left(\frac{B_k}{n}\right) = O\left(\frac{|\tilde{B}_k|}{\sqrt{n}}\right)$. Finally, by the third equation of Lemma 7 and the definition of \tilde{C}_k , the third term of (71) is

$$O\left(\frac{|C_k|}{n} \times \varepsilon_k^{-1/2}\right) = O\left(\frac{|\tilde{C}_k|}{\sqrt{n}}\right).$$

Combining our estimates on the three terms of (71), we finally obtain that (67) is

$$\begin{aligned} O\left(\frac{1}{\varepsilon_k^{3/4} \sqrt{n}} \times \frac{\varepsilon_k^{1/2} |A_k|}{\varepsilon_k n}\right) + O\left(\frac{1}{\varepsilon_k^{3/4} \sqrt{n}} \times \frac{\max(|\tilde{B}_k|, |\tilde{C}_k|)}{\sqrt{n}}\right) \\ = O\left(\frac{\varepsilon_k^{1/2} |\tilde{A}_k|}{\varepsilon_k n}\right) + O\left(\frac{\varepsilon_k^{1/4}}{\varepsilon_k n} \max(|\tilde{B}_k|, |\tilde{C}_k|)\right). \end{aligned}$$

Finally, since $\mathcal{X}' = \Phi_A(\mathcal{X}, \mathcal{V}, \mathcal{S})$, the fifth term (68) is just a linear approximation of the function \mathcal{X}' , so it is bounded by

$$\frac{n}{\varepsilon_{k+1}^{3/4} \sqrt{n}} \times \left(\frac{1}{n}\right)^2 \times \max_{\left[\frac{k}{n}, \frac{k+1}{n}\right]} |\mathcal{X}''|. \quad (72)$$

Using Lemma 9, this is $O\left(n^{-3/2} \varepsilon_{k+1}^{-3/4}\right)$, which is $O\left(\frac{n^{-1/30}}{\varepsilon_k n}\right)$ because $\varepsilon_k \geq n^{-2/5-1/100}$.

Just like in [6], the proofs of the drift estimates on \tilde{B}_k and \tilde{C}_k follow the same lines. More precisely, the decomposition of the error into five terms is the same up to the following changes:

- in the first term (64), the constant $\frac{3}{4}$ becomes $\frac{1}{2}$ for \tilde{C} , and the term disappears completely for \tilde{B} ;
- in the terms (65), (66), (67) and (68), the factors $\frac{1}{\varepsilon_{k+1}^{3/4} \sqrt{n}}$ become $\frac{1}{\sqrt{n}}$ for \tilde{B} and $\frac{1}{\varepsilon_{k+1}^{1/2} \sqrt{n}}$ for \tilde{C} ;
- in the fourth term (67), the drift $-\frac{A_k}{\varepsilon_k n}$ becomes 0 for \tilde{B} and $e^{1/2} \frac{\varepsilon_k^{1/2} B_k}{\varepsilon_k n} - \frac{3}{2} \frac{C_k}{\varepsilon_k n}$ for \tilde{C} .

The analog of the first term (64) is then $O\left(\frac{n^{-1/30}}{\varepsilon_k n}\right)$ for the same reason as for \tilde{A} .

For the second term (65), we also rely on Proposition 1. We note that the bound given by Proposition 1 is slightly weaker because of the factor Z_k in the denominator, which is compensated by the substitution of the factor $\frac{1}{\varepsilon_{k+1}^{3/4} \sqrt{n}}$ by $\frac{1}{\varepsilon_{k+1}^{1/2} \sqrt{n}}$ or $\frac{1}{\sqrt{n}}$ in the analog of (65). More precisely, by Proposition 1, the analog of (65) for \tilde{C} is

$$O\left(\frac{1}{\varepsilon_{k+1}^{1/2} \sqrt{n}} \times \frac{\log^2 n}{V_k Z_k}\right) = O\left(\frac{\log^2 n}{\varepsilon_k^2 n^{3/2}}\right) = O\left(\frac{n^{-1/30}}{\varepsilon_k n}\right)$$

by using Lemma 10 and then $\varepsilon_k \geq n^{-2/5-1/100}$.

For the third term (66), the proof is the same as for \tilde{A} : we decompose the error using all the second-order partial derivatives of Φ_B and Φ_C and estimate the corresponding errors one by one. We find that all error terms are $O\left(\frac{n^{-1/30}}{\varepsilon_k n}\right)$.

For the fourth term (67), we write down the natural analog of (71) for Φ_B and Φ_C and use Lemma 7 for each of the three partial derivatives. For Φ_C , we find

$$\frac{1}{\varepsilon_{k+1}^{1/2} \sqrt{n}} \left(\frac{A_k}{n}, \frac{B_k}{n}, \frac{C_k}{n}\right) \cdot \nabla \Phi_C \left(\mathcal{X}, \mathcal{V}, \mathcal{S} \left(\frac{k}{n}\right)\right) = \frac{1}{\varepsilon_{k+1}^{1/2} \sqrt{n}} \left(\frac{A_k}{n} \frac{\partial \Phi_C}{\partial x} + \frac{B_k}{n} \frac{\partial \Phi_C}{\partial v} + \frac{C_k}{n} \frac{\partial \Phi_C}{\partial s}\right),$$

where all the partial derivatives are taken at the point $\left(\mathcal{X}, \mathcal{V}, \mathcal{S} \left(\frac{k}{n}\right)\right)$. Using Lemma 7 and replacing A_k , B_k and C_k by $\varepsilon_k^{3/4} \sqrt{n} \tilde{A}_k$, $\sqrt{n} \tilde{B}_k$ and $\varepsilon_k^{1/2} \sqrt{n} \tilde{C}_k$, we find that the first term is $O\left(\frac{\varepsilon_k^{3/4} |\tilde{A}_k|}{\varepsilon_k n}\right)$, the second term is $e^{1/2} \frac{\tilde{B}_k}{\varepsilon_k n} + O\left(\frac{\varepsilon_k^{1/2} |\tilde{B}_k|}{\varepsilon_k n}\right)$ and the third term is $-\frac{3}{2} \frac{\tilde{C}_k}{\varepsilon_k n} + O\left(\frac{\varepsilon_k^{1/2} |\tilde{C}_k|}{\varepsilon_k n}\right)$. The argument for Φ_B is similar, and we find that the three terms are respectively $O\left(\frac{\varepsilon_k^{3/4} |\tilde{A}_k|}{\varepsilon_k n}\right)$, $O\left(\frac{\varepsilon_k |\tilde{B}_k|}{\varepsilon_k n}\right)$ and $O\left(\frac{\varepsilon_k |\tilde{C}_k|}{\varepsilon_k n}\right)$.

Finally, for the fifth term (68), we quantify the linear approximation using the second derivatives of \mathcal{V} and \mathcal{S} , which we bound using Lemma 9. We find again that the analog of this term is always $O\left(\frac{n^{-1/30}}{\varepsilon_k n}\right)$. \square

We now move on to the proof of the variance estimates of Proposition 4, which relies mostly on Proposition 2.

Proof of Proposition 4. Just like in the proof of Proposition 3, we first notice that Lemma 10 guarantees that the assumptions of Proposition 2 are satisfied under the assumptions of Proposition 4. Therefore, by Proposition 2, we can write

$$\begin{aligned}\text{Var}(\Delta\tilde{A}_k|\mathcal{F}_k) &= \frac{1}{\varepsilon_k^{3/2}n} \text{Var}(\Delta X_k|\mathcal{F}_k) \\ &= \frac{Z_k}{\varepsilon_k^{3/2}n} + O\left(\frac{Z_k^2}{\varepsilon_k^{3/2}n}\right) + O\left(\frac{1}{\varepsilon_k^{3/2}n} \frac{X_k}{V_k}\right) + O\left(\frac{1}{\varepsilon_k^{3/2}n} \frac{\log^2 n}{V_k}\right).\end{aligned}\quad (73)$$

By Lemma 10 all three terms are $O\left(\frac{1}{\varepsilon_k n}\right)$, which proves (57). The proofs of (58) and (59) from Proposition 2 are the same with simpler terms, so we only have the more accurate estimate (56) left to prove.

For this, we also start from (73). Let us first handle the error terms one by one. The first error term $O\left(\frac{Z_k^2}{\varepsilon_k^{3/2}n}\right)$ is $O\left(\frac{\varepsilon_k^{1/2}}{\varepsilon_k n}\right)$ by Lemma 10. Similarly, by Lemma 10 we have $X_k = O(\varepsilon_k^2 n \times n^{1/100})$ so the second error term $O\left(\frac{X_k}{V_k \varepsilon_k^{3/2}n}\right)$ is $O\left(\frac{\varepsilon_k^{1/2} n^{1/100}}{\varepsilon_k n}\right)$. Finally, also by Lemma 10, the last term of (73) is $O\left(\frac{\log^2 n}{\varepsilon_k n} \times \frac{1}{\varepsilon_k n}\right)$, which is $O\left(\frac{n^{-1/30}}{\varepsilon_k n}\right)$ because $\varepsilon_k \geq n^{-2/5-1/100}$.

We now move on to the main term $\frac{Z_k}{\varepsilon_k^{3/2}n}$: by Lemma 5, we have $Z_k = 3\frac{S_k}{V_k} + O(Z_k^2)$, and the $O(Z_k^2)$ part gets absorbed in the first error term of (73). Moreover, using (42) we can write

$$S_k = \mathcal{S}\left(\frac{k}{n}\right)n + \varepsilon_k^{1/2}\sqrt{n}\tilde{C}_k = \frac{4e^{1/2}}{3}\varepsilon_k^{3/2}n\left(1 + O\left(\varepsilon_k^{1/2}\right) + O\left(\varepsilon_k^{-1}n^{-1/2}|\tilde{C}_k|\right)\right)$$

and similarly using (40)

$$V_k = \mathcal{V}\left(\frac{k}{n}\right)n + \sqrt{n}\tilde{B}_k = 2\varepsilon_k n\left(1 + O\left(\varepsilon_k\right) + O\left(\varepsilon_k^{-1}n^{-1/2}|\tilde{B}_k|\right)\right).$$

Since $\varepsilon_k \geq n^{-2/5-1/100}$ and $\tilde{B}_k, \tilde{C}_k = O(\log n)$, we have

$$\varepsilon_k^{-1}n^{-1/2}|\tilde{B}_k| \leq n^{-1/30} \quad \text{and} \quad \varepsilon_k^{-1}n^{-1/2}|\tilde{C}_k| \leq n^{-1/30}.$$

It follows that

$$\frac{S_k}{V_k} = \frac{2}{3}e^{1/2}\varepsilon_k^{1/2}\left(1 + O\left(\varepsilon_k^{1/2}\right) + O\left(n^{-1/30}\right)\right).$$

Therefore, we get

$$\frac{Z_k}{\varepsilon_k^{3/2}n} = \frac{2\sqrt{e}}{\varepsilon_k n} + O\left(\frac{\varepsilon_k^{1/2}}{\varepsilon_k n}\right) + O\left(\frac{n^{-1/30}}{\varepsilon_k n}\right) + O\left(\frac{Z_k^2}{\varepsilon_k^{3/2}n}\right)$$

and (56) follows. \square

5 Good region and a stochastic differential equation

5.1 Rough behaviour of the fluctuations

The goal of this section is to control the fluctuations of the Markov chain (X, V, S) around its fluid limit by using the drift and variance estimates of Section 4. Since most of the proofs of this

section are straightforward adaptations from [6, Section 4.2], we focus primarily on emphasizing the differences. We keep the same notation as in Section 4. In particular, we recall that θ^n is the first time where X hits 0 and that $\tilde{\theta}^n = \theta^n \wedge (t^*n - n^{3/5-1/100})$, with $t^* = 1 - \frac{3}{2e}$. We also recall that we are working with the multigraph model $\mathbf{G}(n, m_n)$ with $m_n = \frac{e}{2}n + O(\sqrt{n})$.

We start with the control of the rescaled fluctuations \tilde{A} , \tilde{B} and \tilde{C} in the “bulk” of the interval $[0, t^*]$. For $\eta > 0$, we write $k_0(\eta) = \lfloor (t^* - \eta)n \rfloor$. For $\eta, K > 0$, we denote by $\mathcal{R}_{\eta, K}^n$ the set of triples (a, b, c) such that $|a|, |b|, |c| \leq K$ and

$$\mathbb{P} \left(\tilde{A}_{k_0(\eta)}^n = a, \tilde{B}_{k_0(\eta)}^n = b, \tilde{C}_{k_0(\eta)}^n = c \right) > 0.$$

We also define the event $\mathcal{E}_{\eta, K}^n$ by

$$\mathcal{E}_{\eta, K}^n = \left\{ \tilde{\theta}^n > k_0(\eta) \text{ and } \left(\tilde{A}_{k_0(\eta)}^n, \tilde{B}_{k_0(\eta)}^n, \tilde{C}_{k_0(\eta)}^n \right) \in \mathcal{R}_{\eta, K}^n \right\}.$$

We will make the dependence in n implicit when there is no ambiguity.

Lemma 11 (Fluctuations in the bulk). *For all $\eta, \delta > 0$, there exists $K_{\eta, \delta} > 0$ such that for n large enough, we have*

$$\mathbb{P} \left(\mathcal{E}_{\eta, K_{\eta, \delta}}^n \right) \geq 1 - \delta. \quad (74)$$

Proof. The proof of this result is similar to that of [6, Lemma 2], with the difference that here, the initial fluctuations at time $k = 0$ are random and of order \sqrt{n} instead of just being 0. More precisely, by Lemma 1 and the choice (28) of the initial condition $(\mathcal{X}(0), \mathcal{Y}(0), \mathcal{S}(0))$, the initial fluctuations $(\tilde{A}_0^n, \tilde{B}_0^n, \tilde{C}_0^n)$ are tight.

On the other hand, just like in [6], we can apply the results of Ethier and Kurtz [7, Theorem 2.3] *conditionally on the initial fluctuations* $(\tilde{A}_0^n, \tilde{B}_0^n, \tilde{C}_0^n)$. More precisely, if $z_n \in \mathbb{R}^3$ are such that $\mathbb{P} \left((\tilde{A}_0^n, \tilde{B}_0^n, \tilde{C}_0^n) = z_n \right) > 0$ and $z_n \rightarrow z_\infty$, then the law of $(\tilde{A}_{k_0(\eta)}^n, \tilde{B}_{k_0(\eta)}^n, \tilde{C}_{k_0(\eta)}^n)$ conditionally on $(\tilde{A}_0^n, \tilde{B}_0^n, \tilde{C}_0^n) = z_n$ converges to the law of a random Gaussian vector $G_\eta(z_\infty)$ whose mean and covariances depend on η and z_∞ . That is, for any bounded continuous function f , we have

$$\mathbb{E} \left[f \left(\tilde{A}_{k_0(\eta)}^n, \tilde{B}_{k_0(\eta)}^n, \tilde{C}_{k_0(\eta)}^n \right) \mid (\tilde{A}_0^n, \tilde{B}_0^n, \tilde{C}_0^n) = z_n \right] \xrightarrow{n \rightarrow +\infty} \mathbb{E} [f(G_\eta(z_\infty))]. \quad (75)$$

Therefore, let us fix a subsequence. Up to extracting a subsubsequence, we may assume that $(\tilde{A}_0^n, \tilde{B}_0^n, \tilde{C}_0^n)$ converges in distribution to a random variable Z_∞ . By the Skorokhod convergence theorem, let us assume that the convergence is almost sure. Then by (75), for any bounded continuous f , we have

$$\mathbb{E} \left[f \left(\tilde{A}_{k_0(\eta)}^n, \tilde{B}_{k_0(\eta)}^n, \tilde{C}_{k_0(\eta)}^n \right) \mid (\tilde{A}_0^n, \tilde{B}_0^n, \tilde{C}_0^n) \right] \xrightarrow{n \rightarrow +\infty, a.s.} \mathbb{E} [f(G_\eta(Z_\infty))].$$

By dominated convergence, this shows that $(\tilde{A}_{k_0(\eta)}^n, \tilde{B}_{k_0(\eta)}^n, \tilde{C}_{k_0(\eta)}^n)$ converges in distribution. We have proved that any subsequence admits a subsubsequence along which the triple $(\tilde{A}_{k_0(\eta)}^n, \tilde{B}_{k_0(\eta)}^n, \tilde{C}_{k_0(\eta)}^n)$ converges in distribution, which shows that it is tight.

Moreover, the same proof applies to show that the variable $\max_{0 \leq k \leq k_0(\eta)} |\tilde{A}_k^n|$ is tight, which implies $X_k > 0$ for $0 \leq k \leq k_0(\eta)$ with probability $1 - o(1)$, so $\tilde{\theta}^n > k_0(\eta)$. \square

We will now see the time $k_0(\eta)$ as the starting point of our process “near the end”. In order to come back from the multigraph to the simple graph model in Section 6, it will be convenient to

prove a version of Theorem 2 that holds conditionally on $(X, V, S)_{k_0(\eta)}$. Moreover, although such a conditional result was not written as such in [6], its proof for the model studied in [6] would be exactly the same. We recall that $\mathbb{D}_i(n)$ stands for the number of vertices of degree i in the Karp–Sipser core of $\mathbb{G}(n, m_n)$.

Theorem 3. *There is $\eta_0 > 0$ such that the following holds. Let $0 < \eta < \eta_0$ and $K > 0$, and let (a_n, b_n, c_n) be a sequence such that $(a_n, b_n, c_n) \in \mathcal{R}_{\eta, K}^n$ for all n . Then we have the following convergence for conditional distributions:*

$$\left(\begin{array}{c} n^{-3/5} \cdot \mathbb{D}_2(n) \\ n^{-2/5} \cdot \mathbb{D}_3(n) \\ n^{-1/5} \cdot \mathbb{D}_4(n) \\ \mathbb{D}_5(n) \\ \sum_{i \geq 6} \mathbb{D}_i(n) \end{array} \right) \left| \left\{ \left(\tilde{A}^n, \tilde{B}^n, \tilde{C}^n \right)_{k_0(\eta)} = (a_n, b_n, c_n) \right\} \right. \xrightarrow[n \rightarrow +\infty]{(d)} \left(\begin{array}{c} \frac{2^{9/5} 3^{4/5}}{e^{3/5}} \vartheta^{-2} \\ \frac{2^{16/5} 3^{1/5}}{e^{2/5}} \vartheta^{-3} \\ \frac{2^{13/5} 3^{3/5}}{e^{1/5}} \vartheta^{-4} \\ \text{Poi} \left(\frac{48}{5} \vartheta^{-5} \right) \\ 0 \end{array} \right), \quad (76)$$

where $\vartheta = \inf\{t \geq 0 : W_t = t^{-2}\}$ for a standard Brownian motion W started from 0.

We first check that this result easily implies Theorem 2.

Proof of Theorem 2 given Theorem 3. The proof just consists of combining Lemma 11 with Theorem 3. More precisely, let us fix $\eta > 0$ small enough so that the conclusion of Theorem 3 is true. Let $g : \mathbb{R}^5 \rightarrow \mathbb{R}$ be continuous and bounded, and let $\delta > 0$. Finally, let $K = K_{\eta, \delta}$ be given by Lemma 11. For the sake of brevity, we denote by $\mathbf{D}(n)$ and $\mathbf{D}(\infty)$ the two 5-dimensional vectors appearing in (76). Then we have

$$\begin{aligned} & |\mathbb{E}[g(\mathbf{D}(n))] - \mathbb{E}[g(\mathbf{D}(\infty))]| \\ & \leq 2\|g\|_\infty \mathbb{P}(\mathcal{E}_{\eta, K}^c) + \mathbb{E} \left[\mathbb{1}_{\mathcal{E}_{\eta, K}} \left| \mathbb{E}[g(\mathbf{D}(n)) | \mathcal{F}_{k_0(\eta)}] - \mathbb{E}[g(\mathbf{D}(\infty))] \right| \right]. \quad (77) \end{aligned}$$

For n large enough, the first term is bounded by $2\delta\|g\|_\infty$ by Lemma 11. On the other hand, by the Markov property, we have

$$\mathbb{E}[g(\mathbf{D}(n)) | \mathcal{F}_{k_0(\eta)}] = \mathbb{E} \left[g(\mathbf{D}(n)) \mid \left(\tilde{A}, \tilde{B}, \tilde{C} \right)_{k_0(\eta)} \right],$$

and Theorem 3 shows that this converges to $\mathbb{E}[g(\mathbf{D}(\infty))]$, uniformly in the values $\left(\tilde{A}, \tilde{B}, \tilde{C} \right)_{k_0(\eta)}$, provided $\mathcal{E}_{\eta, K}$ occurs. In particular, the second term of (77) is bounded by δ for n large enough, which proves all the convergences of Theorem 2. Finally, the claim that the core is a configuration model is an immediate consequence of Lemmas 4 and 2. \square

In order to prove Theorem 3, we will apply the same strategy as in [6] and make sure for each intermediate result that the conditioning on $\mathcal{F}_{k_0(\eta)}$ is not a problem. The first step is to obtain rough upper bounds for the fluctuations \tilde{A}_k , \tilde{B}_k and \tilde{C}_k . We will rely on the drift and variance estimates from Section 4. This is the analog of [6, Proposition 6].

Proposition 5 (Rough upper bounds). *There is $\eta_0 > 0$ such that for all $0 < \eta < \eta_0$ and $\delta > 0$ and $K > 0$, there exists a constant K' such that for n large enough, under the event $\mathcal{E}_{\eta, K}$, we have*

$$\mathbb{P} \left(\max_{k_0(\eta) \leq k < \tilde{\theta}^n} \left\{ \frac{\tilde{A}_k}{|\log(\varepsilon_k)|^{3/4}}, \tilde{B}_k, \tilde{C}_k \right\} \leq K' \mid \mathcal{F}_{k_0(\eta)} \right) \geq 1 - \delta. \quad (78)$$

Sketch of Proof. The proof of this proposition is very similar to that of [6, Proposition 6]. For this reason, we will just recall the main steps. First, rather than \tilde{C} , we consider instead the process \tilde{E} defined by $\tilde{E}_k = \tilde{C}_k - \sqrt{\varepsilon} \tilde{B}_k$ in view of the form of the drift estimate (55). We thus say that $(\tilde{A}_k, \tilde{B}_k, \tilde{C}_k)$ is in the *good region* as long as $\max\left(\frac{\tilde{A}_k}{|\log(\varepsilon_k)|^{3/4}}, \tilde{B}_k, \tilde{E}_k\right) \leq K'$. We denote by L the first time at which the process exits this good region and evaluate separately the probability to exit this good region via one of the three processes $\tilde{A}, \tilde{B}, \tilde{E}$. In each case, we will decompose the process into its predictable and martingale parts. For all the proof, we fix $\eta > 0$, whose value will be precised later, and work under the event $\mathcal{E}_{\eta, K}$.

CONTROL OF \tilde{B} . We start from $k_0(\eta) = \lfloor (t^* - \eta)n \rfloor$. We can assume $K' > K$, so the event $\mathcal{E}_{\eta, K}$ implies that $L > k_0(\eta)$, so we can apply the estimates of Propositions 3 and 4 until the exit time L of the good region. We then decompose the fluctuations along scales of the form $k_j := \lfloor (t^* - 2^{-j}\eta)n \rfloor$. Note that the estimate (54) is even better than its equivalent [6, Equation (16)]. By the same computation as for [6, Proposition 6], we obtain that for all $k_0(\eta) \leq k \leq \tilde{\theta}^n$, we have

$$\mathbb{1}_{k \leq L} \left| \sum_{\ell=k_0(\eta)}^{k-1} \mathbb{E} \left[\Delta \tilde{B}_\ell | \mathcal{F}_\ell \right] \right| \leq K' \cdot (\text{Cst} \cdot \sqrt{\eta} |\log(\eta)|),$$

where $\text{Cst} > 0$ is an absolute constant. In particular, using Doob's maximal inequality, our estimate (58) in the good region and the assumption that $|\tilde{B}_{k_0(\eta)}| \leq K < K'/2$ if K' was chosen large enough, we obtain

$$\mathbb{P} \left(L < \tilde{\theta}^n \text{ and we exit the good region by } \tilde{B} | \mathcal{F}_{k_0(\eta)} \right) \leq \frac{4}{K'^2/4} \mathbb{E} \left[(M_L^B)^2 | \mathcal{F}_{k_0(\eta)} \right], \quad (79)$$

where M^B is the martingale part of \tilde{B} . Note that this estimate is slightly simpler to obtain than in [6] since we condition on $\mathcal{F}_{k_0(\eta)}$, so we do not need to handle the term $\mathbb{P} \left(|\tilde{B}_{k_0(\eta)}| > K' \right)$. By orthogonality of martingale increments in L^2 , Equation (79) becomes

$$\begin{aligned} \mathbb{P} \left(L < \tilde{\theta}^n \text{ and we exit the good region by } \tilde{B} | \mathcal{F}_{k_0(\eta)} \right) &\leq \frac{16}{K'^2} \sum_{k=k_0}^{\infty} \mathbb{E} \left[\mathbb{1}_{k \leq L} \text{Var}(\Delta \tilde{B}_k | \mathcal{F}_k) | \mathcal{F}_{k_0(\eta)} \right] \\ &\leq \frac{16K''\eta}{K'^2}, \end{aligned}$$

where K'' is an absolute constant coming from the estimate (58). This can be made smaller than δ by taking K' large enough.

CONTROL OF \tilde{E} . To prove that the process does not exit the good region via \tilde{E} , we use more carefully the “restoring force” effect of its conditional drift. In particular, we introduce L_E^- the last time between $k_0(\eta)$ and L at which \tilde{E} is smaller than $K'/2$. For all $L_E^- \leq k \leq L$, using the form of the expected drift given by (54) and (55), we can prove that, if η was chosen small enough, we have for all $L_E^- < k < L$,

$$\mathbb{E} \left[\Delta \tilde{E}_k | \mathcal{F}_k \right] \leq -\frac{K'}{4\varepsilon_k n} < 0,$$

which “pulls back” the fluctuation \tilde{E} towards 0. Thus, on the event $\{k_0(\eta) < L_E^- < L < \tilde{\theta}^n \text{ and } \tilde{E}_L > K'\}$, the variation of the martingale part of \tilde{E} must be larger than $K'/2$ over $[L_E^-, L]$ to compensate the predictable part. Using Doob's maximal inequality and our variance estimates (58) and (59), we can show that the probability that the martingale part compensates the predictable part over

$[L_E^-, L]$ is smaller than $\text{Cst}\sqrt{\eta}/K^2$. If K' was chosen large enough, we thus obtain

$$\mathbb{P}\left(k_0(\eta) < L_E^- < L < T \text{ and } \tilde{E}_L > K' \mid \mathcal{F}_{k_0(\eta)}\right) \leq \delta.$$

Combining this with the symmetric case $\tilde{E}_L < -K'$, this finishes the control of \tilde{E} .

CONTROL OF \tilde{A} . The control of \tilde{A} is the most subtle since we use the strength of the restoring force of the drift (and not only its sign). The drift estimate (53) has the same form as [6, Equation (15)] in the good region. Moreover, the variance estimate (57) only differs by a (multiplicative) constant from [6, Equation (19)], which is not important for this rough control. Thus the same proof shows that the probability to exit the good region via \tilde{A} is small and concludes the proof.

We highlight that in all the proof, the reason why we needed to take η small is that the drift estimate of (53) (resp. (55)) remains of order $-\frac{\tilde{A}_k}{\varepsilon_k n}$ (resp. $-\frac{\tilde{E}_k}{\varepsilon_k n}$). Since the constants in Proposition 3 are absolute, the bound η_0 on how small η needs to be depends neither on δ nor on K . This will be important later to avoid circular dependencies (in Section 6, we will choose η depending on δ). \square

5.2 Reaching the end via a stochastic differential equation

We recall that the number of leaves after k steps is $X_k = n\mathcal{X}\left(\frac{k}{n}\right) + \varepsilon_k^{3/4}\sqrt{n}\tilde{A}_k$. By Proposition 5 and (38), we deduce that the process X_k stays positive as long as

$$t^*n - k \gg n^{3/5}(\log n)^{3/5}.$$

We will now refine the control on \tilde{A} in order to remove the $\log^{3/5}n$ factor. To look more precisely at times of order $n^{3/5}$ before the end, we introduce, for all $k \geq 0$, the notation

$$t_k := n^{-3/5}(t^*n - k) \quad \text{so that} \quad k = t^*n - t_k n^{3/5} \text{ and } \varepsilon_k = t_k n^{-2/5}.$$

Proposition 6 (Control of \tilde{A} near the end). *There is $\eta_0 > 0$ such that for all $0 < \eta < \eta_0$ and $\delta > 0$ and $K > 0$, there exists K' such that for n large enough, under the event $\mathcal{E}_{\eta, K}$, we have*

$$\mathbb{P}\left(\forall k \geq k_0(\eta) \text{ such that } t_k \geq K', \text{ we have } |\tilde{A}_k| < K' t_k^{1/8} \mid \mathcal{F}_{k_0(\eta)}\right) \geq 1 - \delta.$$

The proof is essentially the same as that of [6, Proposition 7], the only differences being the conditioning on $\mathcal{F}_{k_0(\eta)}$ (which does not change anything to the argument), and the value of the constant in front of the variance (which does not play an important role). For this reason, we do not reproduce it here. Equipped with this control on the fluctuations, we now have all the tools to study the convergence of the stopping time θ^n . We note that $t_{\theta^n} \in \mathbb{R}$ is given by

$$\theta^n = t^*n - t_{\theta^n} n^{3/5}.$$

We now show the convergence in distribution of t_{θ^n} to a certain random variable. We will need a statement in the same form as for Theorem 3, i.e. a uniform result conditionally on $\mathcal{F}_{k_0(\eta)}$.

Proposition 7 (Convergence of the stopping time). *There is $\eta_0 > 0$ such that the following holds. Let $0 < \eta < \eta_0$ and $K > 0$, and let $(a_n, b_n, c_n) \in \mathcal{R}_{\eta, K}^n$ for all n . Then we have the convergence of conditional distributions*

$$t_{\theta^n} \mid \left\{ \left(\tilde{A}^n, \tilde{B}^n, \tilde{C}^n \right)_{k_0(\eta)} = (a_n, b_n, c_n) \right\} \xrightarrow[n \rightarrow +\infty]{(d)} 6^{4/5} e^{-3/5} \vartheta^{-2}, \quad (80)$$

where $\vartheta := \inf\{t \geq 0 : W_t = -t^{-2}\}$ is the hitting time of the curve $t \mapsto -t^{-2}$ by a standard Brownian motion W started from 0 at time 0.

Proof. Let η_0 be given by Propositions 5 and 6, and let us fix $0 < \eta < \eta_0$, $K > 0$ and $(a_n, b_n, c_n) \in \mathcal{R}_{\eta, K}^n$ for all n . In all the proof, we work under the conditioning of (80). The form of the drift and variance estimates in Equations (53) and (56) suggests that near the end, the fluctuation process $(\tilde{A}_{t^*n - tn^{3/5}} : -\infty < t < t_{\theta^n})$ converges weakly (for the $\|\cdot\|_{\infty}$ -norm) toward a process $(\mathcal{A}_t)_{-\infty < t \leq \tau}$ satisfying

$$d\mathcal{A}_t = -\frac{\mathcal{A}_t}{4|t|}dt + \frac{2\sqrt{e}}{t}dW_t,$$

where W is a standard Brownian motion. To make this heuristic precise, the proof is similar to that of [6, Proposition 8]. In particular, we need to consider the renormalized process

$$\tilde{F}_k = \frac{\tilde{A}_k}{t_k^{1/4}}, \quad 0 \leq k \leq \theta^n.$$

We fix $\delta > 0$ and we take $K' > 0$ such that on an event of probability at least $1 - 2\delta$, the conclusions of Proposition 5 and 6 hold. We also fix $\zeta \in (0, K'^{-1})$ such that $K'\zeta^{1/8} \leq \delta$. To avoid stopping time issues, we extend \tilde{F} after time θ^n by a process \hat{F} with increments $\pm \frac{2\sqrt{e}}{t_k^{3/2}n^{3/5}}$ with probability $1/2$, as in the proof of [6, Proposition 8]. Then with probability at least $1 - 2\delta$, for all k such that $\zeta < t_k < \zeta^{-1}$, we have

$$\left\{ \begin{array}{l} |\hat{F}_{nt^* - \zeta^{-1}n^{3/5}}| < \delta, \quad (\text{by Prop. 6 and the assumption } K'\zeta^{1/8} \leq \delta), \\ \mathbb{E}[\Delta\hat{F}_k | \mathcal{F}_k] = o(n^{-3/5}) \cdot |\hat{F}_k| + o(n^{-3/5}), \\ \text{Var}(\Delta\hat{F}_k | \mathcal{F}_k) = \frac{2\sqrt{e}}{t_k^{3/2}n^{3/5}} + o(n^{-3/5}), \\ \|\Delta\hat{F}_k\|_{\infty} = o(1). \end{array} \right.$$

The proof of these estimates relying on Propositions 3 and 4 is exactly the same as in [6, Proposition 8], except for the last one. Indeed, in [6] the vertex degrees were bounded, so it was immediate that the increments of X are $O(1)$, so the increments of \hat{F} are $O(n^{-3/5})$. In our setting the vertex degrees are not bounded. However, an immediate consequence of (7) applied to $k = k_0(\eta)$ is that with high conditional probability (under the conditioning of (80)), all the vertex degrees at time $k_0(\eta)$ are smaller than $\log n$. This implies that the increments of X after time $k_0(\eta)$ are at most $\log^2 n$, so $\|\Delta\hat{F}_k\|_{\infty} = O\left(\frac{\log^2 n}{n^{3/5}}\right)$ on a very high probability event.

Using standard diffusion approximation results (see e.g. [13]) and the Dubins–Schwarz theorem, and letting $(\delta, \zeta) \rightarrow (0, 0)$, we deduce the following weak convergence over all compact subsets of $(0, \infty)$:

$$\left(\hat{F}_{t^*n - tn^{3/5}}\right)_{0 < t < \infty} \xrightarrow[n \rightarrow \infty]{} 2 \cdot e^{1/4} \left(W_{\frac{1}{\sqrt{t}}}\right)_{0 < t < \infty}, \quad (81)$$

still under the conditioning (80), and where W is a standard Brownian motion with $W_0 = 0$. Then we obtain

$$t_{\theta^n} \xrightarrow[n \rightarrow \infty]{(d)} \tau := \sup\{t \geq 0 \mid 2 \cdot e^{1/4} W_{\frac{1}{\sqrt{t}}} = -\frac{e}{3}t\},$$

still under the same conditioning. By scaling, we finally find $\tau = \frac{6^{4/5}}{e^{3/5}}\vartheta^{-2}$, where ϑ is given by

$$\vartheta = \inf\{t \geq 0 \mid W_t = t^{-2}\}. \quad (82)$$

□

5.3 Size and composition of the Karp–Sipser core

We have now all the tools to finish the proof of Theorem 3. We first recall that by Lemma 4, conditionally on $(V_{\theta^n}, S_{\theta^n})$, the core follows the distribution $\mathbf{G}(0, V_{\theta^n}, S_{\theta^n})$. Therefore, we first need to estimate $(V_{\theta^n}, S_{\theta^n})$.

For this, let us work under the same conditioning as in (80). By Proposition 7, the renormalized stopping time t_{θ^n} converges in distribution towards $6^{4/5}e^{-3/5}\vartheta^{-2}$. Hence, by definition of the fluctuations B , we can write

$$V_{\theta^n} = \underbrace{B_{\theta^n}}_{\stackrel{\text{Prop.5}}{=} O(\sqrt{n} \log(n)^{3/4})} + \underbrace{n^{\mathcal{V}} \left(\frac{\theta^n}{n} \right)}_{\stackrel{(40)}{\sim} 2t_{\theta^n} n^{3/5}} = 2t_{\theta^n} n^{3/5} + o\left(n^{3/5}\right)$$

in probability. Similarly, at time θ_n , the surplus is given by

$$\begin{aligned} S_{\theta^n} &= \underbrace{n \mathcal{S} \left(\frac{\theta^n}{n} \right)}_{\stackrel{(42)}{\sim} \frac{4\sqrt{e}}{3} t_{\theta^n}^{3/2} n^{2/5}} + \underbrace{C_{\theta^n}}_{\stackrel{\text{Prop.5}}{=} O(n^{3/10} \log(n))}, \\ &= \frac{4\sqrt{e}}{3} t_{\theta^n}^{3/2} n^{2/5} + o\left(n^{2/5}\right) \end{aligned}$$

in probability. In other words, still under the conditioning of (80), we have the convergence in distribution

$$\begin{aligned} \left(\frac{V_{\theta^n}}{n^{3/5}}, \frac{S_{\theta^n}}{(V_{\theta^n})^{2/3}} \right) &\stackrel{(d)}{n \rightarrow \infty} \left(2 \cdot 6^{4/5} e^{-3/5} \vartheta^{-2}, \frac{4\sqrt{e}(6^{4/5} e^{-3/5} \vartheta^{-2})^{3/2}}{3(2 \cdot 6^{4/5} e^{-3/5} \vartheta^{-2})^{2/3}} \right) \\ &= \left(2^{9/5} 3^{4/5} e^{-3/5} \vartheta^{-2}, 2 \cdot 3^{-1/3} \vartheta^{-5/3} \right). \end{aligned} \quad (83)$$

To pass from the two parameters $V_{\theta^n}, S_{\theta^n}$ to a full understanding of the vertex degrees, we will use the following lemma.

Lemma 12. *Let $x > 0$ and let $(v_n), (s_n)$ be sequences of nonnegative integers such that $v_n \rightarrow +\infty$ and $s_n = xv_n^{2/3} + o(v_n^{2/3})$ as $n \rightarrow +\infty$. We also assume that $\log^4 n \leq v_n \leq n$ and that $s_n \leq n$. Then we have the following convergences in distribution:*

$$\begin{aligned} \frac{\#\{\text{vertices of degree 2 in } \mathbf{G}(0, v_n, s_n)\}}{v_n} &\stackrel{(d)}{n \rightarrow \infty} 1, \\ \frac{\#\{\text{vertices of degree 3 in } \mathbf{G}(0, v_n, s_n)\}}{xv_n^{2/3}} &\stackrel{(d)}{n \rightarrow \infty} 1, \\ \frac{\#\{\text{vertices of degree 4 in } \mathbf{G}(0, v_n, s_n)\}}{(3/4) \cdot xv_n^{1/3}} &\stackrel{(d)}{n \rightarrow \infty} 1, \\ \#\{\text{vertices of degree 5 in } \mathbf{G}(0, v_n, s_n)\} &\stackrel{(d)}{n \rightarrow \infty} \text{Poisson} \left(\frac{9x^3}{20} \right), \\ \#\{\text{vertices of degree at least 6 in } \mathbf{G}(0, v_n, s_n)\} &\stackrel{(d)}{n \rightarrow \infty} 0. \end{aligned}$$

From here, we can easily deduce Theorem 3.

Proof of Theorem 3. Let η_0 be given by Proposition 7. As before, we fix $0 < \eta < \eta_0$, $K > 0$ and $(a_n, b_n, c_n) \in \mathcal{R}_{\eta, K}^n$ for all n . As it is the only case with two levels of randomness involved, we

only write down the proof completely for vertices of degree 5. We recall that $\mathbb{D}_5(n)$ stands for the number of vertices of degree 5 in the Karp–Sipser core of $\mathbb{G}(n, m_n)$. We will also write $\mathbb{E}_{a,b,c}[\cdot]$ for conditional expectations on the event $(\tilde{A}, \tilde{B}, \tilde{C})_{k_0(\eta)} = (a_n, b_n, c_n)$.

Let $g, h : \mathbb{R} \mapsto \mathbb{R}$ be two bounded continuous functions. By conditioning on $(V_{\theta^n}, S_{\theta^n})$ and using Lemma 4, we can write

$$\mathbb{E}_{a,b,c} \left[g \left(\frac{V_{\theta^n}}{n^{3/5}} \right) h(\mathbb{D}_5(n)) \right] = \mathbb{E}_{a,b,c} \left[g \left(\frac{V_{\theta^n}}{n^{3/5}} \right) H \left(V_{\theta^n}, \frac{S_{\theta^n}}{\sqrt{\theta^n}} \right) \right], \quad (84)$$

where

$$H(v, x) = \mathbb{E} \left[h \left(\#\{\text{vertices of degree 5 in } \mathbb{G}(0, v, xv^{2/3})\} \right) \right].$$

By the Skorokhod embedding theorem, we may assume that the convergence in distribution (83) is almost sure. Moreover, Lemma 12 implies that when $v_n \rightarrow +\infty$ and $x_n \rightarrow x$, the quantity $H(v_n, x_n)$ converges to $\mathbb{E} [h(\text{Poisson}(9x^3/20))]$. Therefore, we have almost sure convergence of the quantity inside the expectation in the right-hand side of (84). By dominated convergence, we finally obtain

$$\mathbb{E}_{a,b,c} \left[g \left(\frac{V_{\theta^n}}{n^{3/5}} \right) h(\mathbb{D}_5(n)) \right] \xrightarrow{n \rightarrow +\infty} \mathbb{E}_{a,b,c} \left[g \left(\frac{2^{9/5} 3^{4/5}}{e^{3/5} \theta^2} \right) \times \mathbb{E} \left[h \left(\text{Poisson} \left(\frac{9}{20} \left(\frac{2}{3^{1/3} \theta^{5/3}} \right)^3 \right) \right) \mid \vartheta \right] \right].$$

In other words, the pair $\left(\frac{V_{\theta^n}}{n^{3/5}}, \mathbb{D}_5(n) \right)$ under $\mathbb{P}_{a,b,c}$ converges to a pair of random variables where the first coordinate has the law of $2^{9/5} 3^{4/5} e^{-3/5} \theta^{-2}$ and, conditionally on the first, the second follows a Poisson distribution. This proves the convergence of $\mathbb{D}_5(n)$.

Finally, the argument for vertices of other degrees is the same, except that we need to normalize $\mathbb{D}_i(n)$ by the right power of V_{θ^n} before applying h . Note that it is immediate that all the convergences in distribution of Theorem 3 are joint, as all convergences except one are actually convergences in probability to a constant. \square

Finally, we can now prove Lemma 12, which is a consequence of the Poisson approximation for the vertex degrees provided by Lemma 6.

Proof of Lemma 12. The proof consists of a first and second moment computation which makes use of Lemma 6. For the sake of brevity, in this proof, we will write $\mathbb{G}(n)$ for $\mathbb{G}(0, v_n, s_n)$. We also write $z_n = z(v_n, s_n)$. By Lemma 5, our assumptions imply $z_n = 3xv_n^{-1/3} + o(v_n^{-1/3})$ as $n \rightarrow +\infty$. In particular, this implies that v_n, s_n and z_n satisfy the assumptions of Lemma 6. We also recall that the v_n vertices of $\mathbb{G}(n)$ are labelled from 1 to v_n . Using (5), for $2 \leq d \leq 5$, we have

$$\begin{aligned} \mathbb{E} [\#\{\text{vertices of degree } d \text{ in } \mathbb{G}(n)\}] &= v_n \cdot \mathbb{P}(\text{deg}(1) = d) \\ &= v_n \frac{z_n^d}{d! f(z_n)} \left(1 + O \left(\frac{1}{v_n \cdot v_n^{-1/3}} \right) \right) \\ &= (1 + o(1)) \frac{2(3x)^{d-2}}{d!} \cdot v_n^{\frac{5-d}{3}}. \end{aligned}$$

Similarly, using (5) and (7), we have

$$\begin{aligned} \mathbb{E} [\#\{\text{vertices of degree } \geq 6 \text{ in } \mathbb{G}(n)\}] &= O \left(v_n \sum_{d=6}^{\log n} \frac{z_n^d}{d! f(z_n)} \right) + O(n^{-9}) \\ &= O(v_n z_n^4) + O(n^{-9}). \end{aligned}$$

In particular this is $o(1)$, which is sufficient to prove the part of the Lemma on vertices of degree at least 6.

We now move on to the second moment computation. By Equation (8), for all $2 \leq d \leq 4$, we have

$$\begin{aligned} \mathbb{E} \left[\#\{\text{vertices of degree } d \text{ in } G(n)\}^2 \right] &= v_n \cdot \mathbb{P}(\text{deg}(1) = d) + v_n(v_n - 1) \mathbb{P}(\text{deg}(1) = \text{deg}(2) = d) \\ &= O(v_n^{\frac{5-d}{3}}) + \left(v_n^2 + O(v_n) \right) \frac{z_n^{2d}}{d!^2 f(z_n)^2} \left(1 + O\left(\frac{\log^2 n}{v_n^{2/3}} \right) \right) \\ &= (1 + o(1)) \frac{4(3x)^{2d-4}}{d!^2} v_n^{\frac{10-2d}{3}}. \\ &= (1 + o(1)) \mathbb{E} \left[\#\{\text{vertices of degree } \ell \text{ in } G(n)\}^2 \right], \end{aligned}$$

where we have used the assumption $v_n \geq \log^4 n$ to get rid of the error term. This is sufficient to prove the convergence of the number of vertices of degree d for $d \in \{2, 3, 4\}$. However, for $d = 5$, we cannot expect concentration around the expectation, so we will need to compute all the moments to prove convergence in distribution.

More precisely, we fix $m \geq 1$. For $x \in \mathbb{R}$, we denote by $(x)_m := x(x-1)\dots(x-m+1)$ the falling factorial of x . Using Equation (10) from Lemma 6, we have

$$\begin{aligned} \mathbb{E} \left[(\#\{\text{vertices of degree 5 in } G(n)\})_m \right] &= \sum_{j_1=1}^{v_n} \sum_{j_2 \neq j_1} \cdots \sum_{j_m \notin \{j_1, \dots, j_{m-1}\}} \mathbb{P}(\text{deg}(j_1) = \cdots = \text{deg}(j_m) = 5) \\ &= (v_n)_m \cdot (1 + o(1)) \left(\frac{2 \cdot 3^3 x^3 v_n^{-1}}{5!} \right)^m \\ &= (1 + o(1)) \left(\frac{9}{20} x^3 \right)^m \end{aligned}$$

as $n \rightarrow +\infty$. This shows that the factorial moments of the number of vertices of degree 5 converge to those of a Poisson variable, which implies convergence in distribution. \square

6 From the configuration model back to Erdős–Rényi

We now extend our result to simple graphs in order to prove Theorem 1. We will rely on the fact that the graph $G(n, m)$ has the law of the multigraph $\mathbf{G}(n, m)$, conditioned to be simple. Note that passing from one model to the other is not obvious: since the probability that $\mathbf{G}(n, m)$ is simple does not go to one in the regime we are interested in, the conditioning might “twist” the limiting distribution. Another natural approach, which is followed in [8], would be to use the fact that $G(n, m)$ and $\mathbf{G}(n, m)$ can be coupled so that the edit distance between them (i.e. the number of edges to add or remove to pass from one to the other) is $O(1)$. Unfortunately, this does not adapt to our problem, since contrary to the matching number, the size of the Karp–Sipser core is not a Lipschitz function of the graph for the edit distance.

To pass from one model to the other, it will be useful to know the probability that a configuration model is a simple graph, so we recall the following classical result (see e.g. [14, Corollary 5.3]). If $\mathbf{d} = (d_i)_{i \geq 1}$ is a sequence of integers which is 0 eventually, we write $|\mathbf{d}| = \frac{1}{2} \sum_{i \geq 1} i d_i$. If this quantity is an integer, we denote by $\text{CM}(\mathbf{d})$ a configuration model with d_i vertices of degree i for all i .

Lemma 13. Let $\mathbf{d}^{(n)}$ be degree sequences such that $|\mathbf{d}^{(n)}| \rightarrow +\infty$ and such that $\max\{i \geq 1 | d_i^{(n)} > 0\} = o\left(\sqrt{|\mathbf{d}^{(n)}|}\right)$. Then

$$\mathbb{P}\left(\text{CM}(\mathbf{d}^{(n)}) \text{ is simple}\right) = \exp\left(-\frac{1}{4}\left(\frac{\sum_{i \geq 1} i^2 d_i^{(n)}}{|\mathbf{d}^{(n)}|}\right)^2 + \frac{1}{4}\right) + o(1)$$

as $n \rightarrow +\infty$.

Again, for $\eta > 0$, we recall that $k_0(\eta) = \lfloor (t^* - \eta)n \rfloor$. We will decompose the event that the graph $\mathbf{G}(n, m_n)$ is simple as the intersection of three events:

- the event $\text{Simp}_{\rightarrow k_0(\eta)}$ that the execution of the Karp–Sipser algorithm does not encounter any multiple edge or loop until step $k_0(\eta)$,
- the event $\text{Simp}_{k_0(\eta) \rightarrow \theta^n}$ that the execution of the Karp–Sipser algorithm does not encounter any multiple edge or loop between step $k_0(\eta)$ and its end θ^n ,
- the event that the Karp–Sipser core is simple.

Note that if a graph contains two multiple edges, then these two edges will be discovered at the same step of the Karp–Sipser algorithm, so the decomposition above is not ambiguous. We want to show that the simplicity of $\mathbf{G}(n, m_n)$ is asymptotically independent from the degrees in its Karp–Sipser core. Roughly speaking, the argument will be to handle the three events of this decomposition as follows:

- the event $\text{Simp}_{\rightarrow k_0(\eta)}$ is almost independent from the Karp–Sipser core because Theorem 3 holds conditionally of $\mathcal{F}_{k_0(\eta)}$ (this is why Theorem 2 is not sufficient and we needed Theorem 3),
- the event $\text{Simp}_{k_0(\eta) \rightarrow \theta^n}$ has probability close to 1,
- by Lemma 13, the probability that the Karp–Sipser core is simple conditionally on its vertex degrees is almost deterministic.

To make this sketch precise, we start with the second item.

Lemma 14. Let $\delta > 0$. Then there is $\eta > 0$ such that for n large enough, we have

$$\mathbb{P}\left(\text{Simp}_{k_0(\eta) \rightarrow \theta^n}\right) \geq 1 - \delta.$$

Proof. We recall that G_k is the multigraph obtained after k steps of the Karp–Sipser algorithm applied to the random multigraph $\mathbf{G}(n, m_n)$. We will show that if η is small enough, then with high probability, all the loops and multiple edges in $G_{k_0(\eta)}$ are either a loop on a vertex of degree 2 or two edges joining the same pair of vertices of degree 2. In particular, loops and multiple edges do not have leaves in their connected component, so they will never be removed by the Karp–Sipser algorithm.

Let $\eta > 0$ (we will fix its value later). For the sake of brevity, we will just write k_0 for $k_0(\eta)$. We denote by \mathcal{E}_η the event that $|\tilde{A}_{k_0}|, |\tilde{B}_{k_0}|, |\tilde{C}_{k_0}| \leq \log n$. We note that by Lemma 11, the probability of \mathcal{E}_η tends to 1 as $n \rightarrow +\infty$. On the event \mathcal{E}_η , Lemma 10 ensures that the assumptions of Lemma 6

are satisfied, so we can write

$$\begin{aligned}
& \mathbb{E} [\#\{\text{loops of } G_{k_0} \text{ attached to a vertex of degree at least 3}\} | \mathcal{F}_{k_0}] \\
&= \sum_{j=X_{k_0}+1}^{X_{k_0}+V_{k_0}} \sum_{i \geq 3} \mathbb{P} \left(\deg_{G_{k_0}}(j) = i \mid \mathcal{F}_{k_0} \right) \frac{i(i-1)}{X_{k_0} + 2V_{k_0} + S_{k_0} - 1} \\
&\stackrel{\text{Lem. 6}}{=} \frac{V_{k_0}}{X_{k_0} + 2V_{k_0} + S_{k_0} - 1} \left(\sum_{i=3}^{\log(n)} i(i-1) \times O \left(\frac{Z_{k_0}^i}{i! f(Z_{k_0})} \right) + O(n^{-8}) \right),
\end{aligned}$$

where the implied constants are absolute. By writing $\frac{Z_{k_0}^i}{f(Z_{k_0})} = O(Z_{k_0}^{i-2}) = O(Z_{k_0} C^{i-3})$ where C is an upper bound on Z , we find that the last display is $O(Z_k) + O(n^{-8})$. Finally, if \mathcal{E}_η occurs, then by Lemma 10 we have $Z_{k_0} = O(\sqrt{\eta})$. Therefore, there is an absolute constant c such that for n large enough, we have

$$\mathbb{P}(G_{k_0} \text{ has a loop attached to a vertex with degree at least 3}) \leq c\sqrt{\eta}.$$

We now prove the same with multiple edges. On the event \mathcal{E}_η , we can write

$$\begin{aligned}
& \mathbb{E} [\#\{\text{pairs of multiple edges of } G_{k_0} \text{ adjacent to a vertex of degree at least 3}\} | \mathcal{F}_{k_0}] \\
&= \sum_{\substack{j_1, j_2 = X_{k_0}+1 \\ j_1 \neq j_2}}^{X_{k_0}+V_{k_0}} \sum_{\substack{i_1 \geq 3 \\ i_2 \geq 2}} \mathbb{P} \left(\deg_{G_{k_0}}(j_1) = i_1, \deg_{G_{k_0}}(j_2) = i_2 \mid \mathcal{F}_{k_0} \right) \frac{i_1(i_1-1) \cdot i_2(i_2-1)}{(X_{k_0} + 2V_{k_0} + S_{k_0} - 1)(X_{k_0} + 2V_{k_0} + S_{k_0} - 3)} \\
&\stackrel{\text{Lem. 6}}{=} \frac{V_{k_0}(V_{k_0}-1)}{(X_{k_0} + 2V_{k_0} + S_{k_0} - 1)(X_{k_0} + 2V_{k_0} + S_{k_0} - 3)} \left(O(n^{-4}) + \sum_{\substack{3 \leq i_1 \leq \log(n) \\ 2 \leq i_2 \leq \log(n)}} \frac{i_1(i_1-1)i_2(i_2-1)}{i_1!i_2!} O \left(\frac{Z_{k_0}^{i_1} Z_{k_0}^{i_2}}{f(Z_{k_0})^2} \right) \right) \\
&= O(n^{-4}) + O(Z_k).
\end{aligned}$$

Therefore, we have proved that the probability that G_{k_0} has a loop or multiple edges incident to a vertex of degree at least 3 is bounded by $c\sqrt{\eta}$ for some absolute constant c . This proves the Lemma by taking $\eta = \left(\frac{\delta}{c}\right)^2$. \square

Proof of Theorem 1. We first prove the result for the model $G(n, m_n)$ with a fixed number m_n of edges, where $m_n = \frac{c}{2}n + O(\sqrt{n})$. Passing from this model to $G\left[n, \frac{c}{n}\right]$ will follow easily by integrating over the possible values of the number of edges of $G\left[n, \frac{c}{n}\right]$.

Most of the proof will consist in proving the convergence of vertex degrees (1). To keep the notation light, we will write down the proof completely only for vertices of degree 5.

Let g be a bounded, continuous function from \mathbb{R} to \mathbb{R} . For $x > 0$, we also write

$$\widehat{g}(x) = \mathbb{E} \left[g \left(\text{Poi} \left(\frac{48}{5} x^{-5} \right) \right) \right],$$

where $\text{Poi}(\lambda)$ denotes a Poisson variable with parameter λ . We note that \widehat{g} is bounded and is a continuous function of θ . We also recall that for $i \geq 2$, we denote by $\mathbb{D}_i(n)$ the number of vertices of degree i in the Karp–Sipser core of $G(n, m_n)$. We want to prove that

$$\frac{\mathbb{E} \left[g(\mathbb{D}_5(n)) \mathbb{1}_{G(n, m_n) \text{ is simple}} \right]}{\mathbb{P}(G(n, m_n) \text{ is simple})} \xrightarrow{n \rightarrow +\infty} \mathbb{E}[\widehat{g}(\theta)], \quad (85)$$

where we recall that ϑ is defined by (82). We first notice that by the Poisson approximation of Lemma 6 (for $k = 0$) and by Lemma 13, the denominator of (85) is bounded away from 0 (it actually goes to $e^{-\frac{\varepsilon}{2} - \frac{\varepsilon^2}{4}}$ but we will not need the exact computation). We now study the numerator. We fix $\delta > 0$. Let $\eta > 0$ small be given by Lemma 14, and let $K = K_{\eta, \delta}$ be the constant provided by Lemma 11. We recall that $\mathcal{E}_{\eta, K}$ is the event that $\theta^n > k_0(\eta)$ and $|\tilde{A}_{k_0(\eta)}|, |\tilde{B}_{k_0(\eta)}|, |\tilde{C}_{k_0(\eta)}| \leq K$, and that $\mathbb{P}(\mathcal{E}_{\eta, K}) \geq 1 - \delta$ for n large enough by Lemma 11. We also write $\mathcal{E}'_\eta = \text{Simp}_{k_0(\eta) \rightarrow \theta^n}$. Finally, we write

$$\mathcal{E}'' = \left\{ \left| \frac{\sum_{i \geq 1} i^2 \mathbb{D}_i(n)}{4 \sum_{i \geq 1} i \mathbb{D}_i(n)} - \frac{1}{2} \right| \leq n^{-1/10} \right\}.$$

Note that as a consequence of Theorem 2, we have $\sum_{i \geq 2} i \mathbb{D}_i(n) = 2\mathbb{D}_2(n) \left(1 + O(n^{-1/5})\right)$ and $\sum_{i \geq 2} i^2 \mathbb{D}_i(n) = 4\mathbb{D}_2(n) \left(1 + O(n^{-1/5})\right)$ in probability, so $\mathbb{P}(\mathcal{E}'') \rightarrow 1$ as $n \rightarrow +\infty$. Combining these observations, for n large enough, we have

$$\mathbb{P}\left(\left(\mathcal{E}_{\eta, K} \cap \mathcal{E}'_\eta \cap \mathcal{E}''\right)^c\right) \leq 3\delta.$$

Therefore, we can write

$$\begin{aligned} & \mathbb{E} \left[g(\mathbb{D}_5(n)) \mathbb{1}_{\mathbf{G}(n, m_n) \text{ is simple}} \right] \\ &= O(\delta) + \mathbb{E} \left[\mathbb{1}_{\text{Simp} \rightarrow k_0(\eta)} \mathbb{1}_{\mathcal{E}_{\eta, K}} \mathbb{1}_{\mathcal{E}'_\eta} g(\mathbb{D}_5(n)) \mathbb{1}_{\mathcal{E}''} \mathbb{1}_{\text{KSCore}(\mathbf{G}(n, m_n)) \text{ is simple}} \right], \end{aligned}$$

where the constant implied by the O notation only depends on the function g . Since $\mathbb{P}(\mathcal{E}'_\eta) \geq 1 - \delta$ and everything inside of the expectation is bounded, we can remove the indicator $\mathbb{1}_{\mathcal{E}'_\eta}$ and absorb the difference in the $O(\delta)$ term. By conditioning on \mathcal{F}_{θ^n} and on the degrees in the Karp–Sipser core, we obtain

$$\begin{aligned} & \mathbb{E} \left[g(\mathbb{D}_5(n)) \mathbb{1}_{\mathbf{G}(n, m_n) \text{ is simple}} \right] \\ &= O(\delta) + \mathbb{E} \left[\mathbb{1}_{\text{Simp} \rightarrow k_0(\eta)} g(\mathbb{D}_5(n)) \mathbb{1}_{\mathcal{E}_{\eta, K}} \mathbb{1}_{\mathcal{E}''} \mathbb{P}(\text{KSCore}(\mathbf{G}(n, m_n)) \text{ is simple} | \mathcal{F}_{\theta^n}, (\mathbb{D}_i(n))_{i \geq 2}) \right]. \end{aligned}$$

By Lemma 13, under the event \mathcal{E}'' , the conditional probability for $\text{KSCore}(\mathbf{G}(n, m_n))$ to be simple is $e^{-3/4} + o(1)$, where $o(1)$ is uniform. Again, the error term can be absorbed in the $O(\delta)$ term, and we can then remove the indicator $\mathbb{1}_{\mathcal{E}''}$ as we did before for $\mathbb{1}_{\mathcal{E}'_\eta}$. By conditioning on $\mathcal{F}_{k_0(\eta)}$, the last display rewrites

$$O(\delta) + e^{-3/4} \mathbb{E} \left[\mathbb{1}_{\text{Simp} \rightarrow k_0(\eta)} \mathbb{1}_{\mathcal{E}_{\eta, K}} \mathbb{E} \left[g(\mathbb{D}_5(n)) | \mathcal{F}_{k_0(\eta)} \right] \right].$$

By the Markov property, on the event $\theta^n > k_0(\eta)$ (which is implied by $\mathcal{E}_{\eta, K}$), the conditional expectation $\mathbb{E} \left[g(\mathbb{D}_5(n)) | \mathcal{F}_{k_0(\eta)} \right]$ is a measurable function of $(\tilde{A}, \tilde{B}, \tilde{C})_{k_0(\eta)}$. Moreover, Theorem 3 guarantees that on $\mathcal{E}_{\eta, K}$, this quantity converges towards $\mathbb{E} \left[g(\hat{\vartheta}) \right]$, uniformly in the values $(\tilde{A}, \tilde{B}, \tilde{C})_{k_0(\eta)}$. Therefore, for n large enough, we have

$$\mathbb{E} \left[g(\mathbb{D}_5(n)) \mathbb{1}_{\mathbf{G}(n, m_n) \text{ is simple}} \right] = O(\delta) + e^{-3/4} \mathbb{E} \left[\hat{g}(\hat{\vartheta}) \right] \mathbb{P} \left(\text{Simp} \rightarrow k_0(\eta) \text{ and } \mathcal{E}_{\eta, K} \right).$$

As before, using this time Lemma 11 and the choice of the constant K , the event $\mathcal{E}_{\eta, K}$ can be absorbed by the $O(\delta)$ term. Coming back to (85) and using that $\mathbb{P}(\mathbf{G}(n, m_n) \text{ is simple})$ is bounded

away from 0, we have shown that for any $\delta > 0$, there is $\eta > 0$ such that for n large enough, we have

$$\mathbb{E} [g(\mathbb{D}_5(n)) | \mathbb{G}(n, m_n) \text{ is simple}] = O(\delta) + \frac{e^{-3/4} \mathbb{P}(\text{Simp}_{\rightarrow k_0(\eta)})}{\mathbb{P}(\mathbb{G}(n, m_n) \text{ is simple})} \mathbb{E} [\widehat{g}(\vartheta)], \quad (86)$$

where the constant implied by the O notation only depends on g . In particular, if g is constant and equal to 1, then so is \widehat{g} and the last formula becomes

$$\mathbb{P}(\text{Simp}_{\rightarrow k_0(\eta)}) = (1 + O(\delta)) e^{3/4} \mathbb{P}(\mathbb{G}(n, m_n) \text{ is simple}).$$

Plugging this back into (86), we get

$$\mathbb{E} [g(\mathbb{D}_5(n)) | \mathbb{G}(n, m_n) \text{ is simple}] = \mathbb{E} [\widehat{g}(\vartheta)] + O(\delta).$$

Since we can take δ arbitrarily small, this concludes the proof of convergence of $\mathbb{D}_5(n)$ for the random simple graph $\mathbb{G}(n, m_n)$.

To extend the result to $(D_i(n))_{i \geq 2}$ instead of just $D_5(n)$, the proof is exactly the same. We just need to replace $g(\mathbb{D}_5(n))$ by

$$g \left(\frac{\mathbb{D}_2(n)}{n^{3/5}}, \frac{\mathbb{D}_3(n)}{n^{2/5}}, \frac{\mathbb{D}_4(n)}{n^{1/5}}, \mathbb{D}_5(n), \sum_{i \geq 6} \mathbb{D}_i(n) \right),$$

where g is a bounded, continuous function from \mathbb{R}^5 to \mathbb{R} .

We now move on to the claim that conditionally on its vertex degrees, the Karp–Sipser core of $\mathbb{G}(n, m_n)$ is a configuration model conditioned to be simple. This will be true without assuming anything about m_n . For this, let \mathfrak{g} be a simple graph with no leaf and no isolated vertex. We need to show that $\mathbb{P}(\text{KSCore}(\mathbb{G}(n, m_n)) = \mathfrak{g})$ only depends on the vertex degrees of \mathfrak{g} . But we have

$$\mathbb{P}(\text{KSCore}(\mathbb{G}(n, m_n)) = \mathfrak{g}) = \frac{\mathbb{P}(\mathbb{G}(n, m_n) \text{ is simple and } \text{KSCore}(\mathbb{G}(n, m_n)) = \mathfrak{g})}{\mathbb{P}(\mathbb{G}(n, m_n) \text{ is simple})}.$$

The denominator does not depend on \mathfrak{g} , so we only need to consider the numerator. Since \mathfrak{g} is simple, it can be rewritten as

$$\begin{aligned} & \mathbb{P}(\text{Simp}_{\rightarrow k_0(\eta)} \text{ and } \text{Simp}_{k_0(\eta) \rightarrow \theta^n} \text{ and } \text{KSCore}(\mathbb{G}(n, m_n)) = \mathfrak{g}) \\ &= \mathbb{E} \left[\mathbb{1}_{\text{Simp}_{\rightarrow k_0(\eta)}} \mathbb{1}_{\text{Simp}_{k_0(\eta) \rightarrow \theta^n}} \mathbb{P}(\text{KSCore}(\mathbb{G}(n, m_n)) = \mathfrak{g} | \mathcal{F}_{\theta^n}) \right]. \end{aligned}$$

But by Lemmas 4 and 2, the core $\text{KSCore}(\mathbb{G}(n, m_n))$ is a configuration model conditionally on \mathcal{F}_{θ^n} and on its vertex degrees, so this only depends on the vertex degrees of \mathfrak{g} .

Finally, to pass from $\mathbb{G}(n, m_n)$ to the “true” Erdős–Rényi model $\mathbb{G}[n, \frac{e}{n}]$, let M_0^n be the total number of edges of $\mathbb{G}[n, \frac{e}{n}]$. By the Skorokhod embedding theorem, we may assume that $\frac{M_0^n - en/2}{\sqrt{n}}$ converges a.s. to a Gaussian variable, so in particular $M_0^n = \frac{e}{2}n + O(\sqrt{n})$. Therefore, if $g : \mathbb{R}^5 \rightarrow \mathbb{R}$ is bounded and continuous, we have the almost sure convergence

$$\mathbb{E}_{\mathbb{G}[n, \frac{e}{n}]} \left[g \left(\frac{\mathbb{D}_2(n)}{n^{3/5}}, \dots, \sum_{i \geq 6} \mathbb{D}_i(n) \right) \middle| M_0^n \right] \xrightarrow[n \rightarrow +\infty]{a.s.} \mathbb{E} [\widehat{g}(\vartheta)],$$

and (1) for $\mathbb{G}[n, \frac{e}{n}]$ follows by taking the expectation. Finally, for any simple graph \mathfrak{g} , we can write the decomposition

$$\mathbb{P}(\text{KSCore}(\mathbb{G}[n, p]) = \mathfrak{g}) = \sum_{m=0}^{\binom{n}{2}} \mathbb{P}(M_0^n = m) \mathbb{P}(\text{KSCore}(\mathbb{G}(n, m)) = \mathfrak{g}),$$

which only depends on \mathbf{g} through its vertex degrees. It follows that conditionally on its vertex degrees, the graph $\text{KScore}(G[n, p])$ is a configuration model conditioned to be simple. \square

References

- [1] J. ARONSON, A. FRIEZE, AND B. G. PITTEL, *Maximum matchings in sparse random graphs: Karp–Sipser revisited*, *Random Structures & Algorithms*, 12 (1998), pp. 111–177.
- [2] M. BAUER AND O. GOLINELLI, *Core percolation in random graphs: a critical phenomena analysis*, *The European Physical Journal B-Condensed Matter and Complex Systems*, 24 (2001), pp. 339–352.
- [3] ———, *Exactly solvable model with two conductor-insulator transitions driven by impurities*, *Physical Review Letters*, 86 (2001), pp. 2621–2624.
- [4] T. BOHMAN AND A. FRIEZE, *Karp–Sipser on random graphs with a fixed degree sequence*, *Combinatorics, Probability and Computing*, 20 (2011), pp. 721–741.
- [5] C. BORDENAVE AND M. LELARGE, *The rank of diluted random graphs*, in *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete algorithms*, SIAM, 2010, pp. 1389–1402.
- [6] T. BUDZINSKI, A. CONTAT, AND N. CURIEN, *The critical Karp–Sipser core of random graphs*, *Annals of Applied Probability*, (to appear).
- [7] S. N. ETHIER AND T. G. KURTZ, *Markov processes: characterization and convergence*, vol. 282, John Wiley & Sons, 2009.
- [8] M. GLASGOW, M. KWAN, A. SAH, AND M. SAWHNEY, *A central limit theorem for the matching number of a sparse random graph*, arXiv preprint arXiv:2402.05851, (2024).
- [9] S. JANSON, *Monotonicity, asymptotic normality and vertex degrees in random graphs*, *Bernoulli*, (2007), pp. 952–965.
- [10] M. JONCKHEERE AND M. SÁENZ, *Asymptotic optimality of degree-greedy discovering of independent sets in configuration model graphs*, *Stochastic Processes and their Applications*, 131 (2021), pp. 122–150.
- [11] R. M. KARP AND M. SIPSER, *Maximum matching in sparse random graphs*, in *22nd Annual Symposium on Foundations of Computer Science (sfcs 1981)*, IEEE, 1981, pp. 364–375.
- [12] E. KREAČIĆ, *Some problems related to the Karp–Sipser algorithm on random graphs*, PhD thesis, University of Oxford, 2017.
- [13] H. J. KUSHNER, *On the weak convergence of interpolated Markov chains to a diffusion*, *The Annals of Probability*, (1974), pp. 40–50.
- [14] B. D. MCKAY AND N. C. WORMALD, *Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$* , *Combinatorica*, 11 (1991), pp. 369–382.
- [15] N. C. WORMALD, *Differential equations for random processes and random graphs*, *The Annals of Applied Probability*, (1995), pp. 1217–1235.