OLIVER NAGY

Approaching equilibrium in a dynamic network

Approaching equilibrium in a dynamic network

Oliver Nagy

Approaching equilibrium in a dynamic network

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof. dr. ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op donderdag 16 januari 2025
klokke 13.00 uur

door

Oliver Nagy

geboren te Bratislava, Slovakia
in 1994

Promotores:

Prof. dr. W. Th. F. den Hollander
Prof. dr. R. W. van der Hofstad (Technische Universiteit Eindhoven)

Co-promotor:

Dr. L. Avena (Università degli Studi di Firenze)

Promotiecommissie:

Prof. dr. ir. G. L. A. Derks
Prof. dr. P. Stevenhagen

Dr. A. Ben-Hamou (Sorbonne Université)
Dr. D. V. Krioukov (Northeastern University)
Prof. dr. N. V. Litvak (Technische Universiteit Eindhoven)

This dissertation and the research presented therein were supported by the Dutch Research Council (NWO) through the Gravitation-grant NETWORKS-024.002.003.

Cover: Original sketch by the author, digitally post-processed using, among other tools, DALL-E by OpenAI.

Contents

CHAPTER 1

Introduction

§1.1 Overview

This dissertation presents four separate chapters, all striving towards the same goal: to understand dynamic phenomena occurring in dynamic networks and to explore the emergence of *equilibrium* in a broad sense.

More concretely:

I. Chapter 2 deals with mixing times of non-backtracking random walks on a *class* of graphs initially sampled according to the *configuration model* (with mild restrictions on the degree sequence) and later evolving via *edge rewiring*.

   We show that the *mixing time* of a non-backtracking random walk on such dynamic random graphs asymptotically differs from the mixing time of the same process on the initial static graph only by a multiplicative factor equal to the tail of the distribution of a particularly simple stopping time–the first time when the random walk steps over a previously rewired edge.

II. In Chapter 3, we focus on a specific kind of dynamic geometry generated by *dynamic permutations* started from the identity permutation and evolving via the application of *random transpositions*, possibly with additional constraints. The process that we focus on is an *infinite-speed* random walk, which arises as a distributional limit of a random walk whose stepping rate, relative to the stepping rate of the permutation dynamics, tends to infinity as the size of the permutation grows to infinity.

   We fully characterize the mixing profile of this process. We do so for two different kinds of dynamics, called *coagulative* and *coagulative-fragmentative*. In both cases, we observe an unusual *discontinuity* in the mixing profile, different from previously observed *cut-off* phenomena, occurring at a *random* time whose law we identify.

III. Chapter 4 is in the realm of engineering mathematics. We study the communication network within a swarm of satellites that work together as an *interferometer*. While in reality this is a highly dynamic problem, we show that under appropriate assumptions it can be modelled as a static *k-nearest neighbour graph*.

   We explore the effects of different choices of the parameter k on network connectivity, network resilience, and energy expenditure within the network. We derive a closed-form distribution for transmission costs in the swarm in a box of infinite size and numerically estimate the correction for finite-size effects. Furthermore, we numerically compute and heuristically explain the scaling properties of other quantities of interest, such as the maximal and average length of the baselines of the emergent interferometer.

IV. Chapter 5 is focused on the fragmented field of *centrality measures*. We propose a *quantitative* way to compare *arbitrary* pairs of centrality measures defined on the same graph. The main idea is to look at the ordering of vertices *induced* by the centrality measures and use this to define new quantities of interest, such as the *centrality comparison curve*.

We apply the method described in this chapter in several ways: we describe a natural approximation scheme for difficult-to-compute centrality measures, compare pairs of commonly used centrality measures, and propose new conjectures based on the observed behaviour of the centrality comparison curve.

Since these chapters were written with different audiences in mind, they also differ in their style of presentation. Chapters 2 and 3 are written to the standard of rigour expected in mathematical journals. Chapter 4 is aimed towards the engineering community, and hence it is written in a style similar to publications in relevant IEEE journals. The final scientific chapter, Chapter 5, is written for researchers in complex networks, and the style is reminiscent of *Physical Review* journals.

The remainder of this chapter consists of two parts: Section 1.2 serves as a brief introduction and overview of the main topics covered in Chapters 2-5, while Section 1.3 is a guide to the overall organisation of the rest of this dissertation.

§1.2 Brief introduction to the content

This section contains a narrative introduction to the main topics covered in Chapters 2-5. A more technical introduction is given in the introductory sections of the respective chapters.

§1.2.1 Random walks and their mixing properties

Random walks are one of the most fundamental stochastic processes, and have been studied in many contexts. We refer the reader to [97] for a modern introduction to this fascinating subject. In this section, we focus on *discrete-time* random walks on *finite graphs*. Unless explicitly stated otherwise, the word "graph" denotes an *undirected finite* graph.

The most elementary example is the *simple* random walk on a graph $G = (V_G, E_G)$, which is the vertex-valued Markov chain $(X(t))_{t\in\mathbb{N}_0}$, started from an initial vertex v_0, with transition probabilities

$$\mathbb{P}\left(X(t+1) = v \mid X(t) = u\right) = \begin{cases} \frac{1}{\deg(u)} & \text{if } \{u, v\} \in E_G, \\ 0 & \text{otherwise,} \end{cases}$$

where $\deg(u)$ denotes the degree of the vertex $u \in V_G$. Stationary distributions are easy to identify as long as the underlying graph is static. In particular, if G is finite, connected, and with finite degrees, then the unique stationary distribution of $(X(t))_{t\in\mathbb{N}_0}$ is the distribution μ^{stat} that places the following mass on a vertex $v \in V_G$:

$$\mu^{\text{stat}}(v) = \frac{1}{2|E_G|} \deg(v).$$

The above definition can be modified in many ways. One possible modification is the *non-backtracking* random walk $(Y(t))_{t\in\mathbb{N}_0}$, where we disallow the random walk to traverse the same edge twice right after each other. Even though $(Y(t))_{t\in\mathbb{N}_0}$ obviously depends on the last *two* states, it is possible to reformulate it over a different state space

in a way that Markov property is preserved (see, for example, [92] or Section 2.2.2). The stationary distribution of $(Y(t))_{t\in\mathbb{N}_0}$ is μ^{stat} introduced above. This modification is important in Chapter 2. In Chapter 3 we will work with an *infinite-speed* random walk, which is a distributional limit of a random walk whose relative stepping rate with respect to the stepping rate of the underlying dynamics tends to infinity as the size of the graph grows to infinity.

If the random walk in question admits a unique stationary distribution, then it is natural to ask *how fast* the process approaches the stationary distribution. A common phrase used to describe this approach to equilibrium is *mixing*.

Total variation distance. There are multiple ways to quantify mixing. One of them is by using the total variation distance between the distribution at time t and the target equilibrium distribution. More formally, take two probability measures π, ρ defined over a common probability space (A, Ω). The total variation distance between π and ρ is defined as

$$d_{\text{TV}}(\pi, \rho) = \sup_{\omega\in\Omega} |\pi(\omega) - \rho(\omega)|,$$

i.e., it is the maximal discrepancy of the probability of an event $\omega \in \Omega$ according to π and ρ. In the special setting of Markov chains defined on a finite state space S, the total variation distance between two probability measures μ and ν on the same finite state space S can be expressed as:

$$d_{\text{TV}}(\mu, \nu) = \frac{1}{2}\sum_{x\in S} |\mu(x) - \nu(x)| = \sum_{x\in S} [\mu(x) - \nu(x)]_+ = \max_{A\subseteq S} [\mu(A) - \nu(A)]. \tag{1.1}$$

If $\mu_X(t)$ is the distribution of the process $(X(t))_{t\in\mathbb{N}_0}$ at time t and μ^{stat} is its unique stationary distribution, then $d_{\text{TV}}(\mu_X(t), \mu^{\text{stat}})$, seen as a function of t, evidently quantifies the approach to equilibrium. For any $\varepsilon \in (0, 1)$, we define the ε-mixing time $t_{\text{mix}}(\varepsilon)$ as the first time when $d_{\text{TV}}(\mu_X(t), \mu^{\text{stat}}) \leq \varepsilon$. We refer the reader to [101] for a pedagogical introduction to Markov chain mixing and to the unfinished monograph [5] for an overview of advanced techniques and many results that are by now considered to be folklore in the field.

Cut-off. One fascinating phenomenon related to mixing is the *cut-off phenomenon*. First observed for a random walk on the symmetric group [49], the cut-off phenomenon represents a jump discontinuity in the limiting mixing profile. Subsequently, this phenomenon has been observed in various other settings (see the overview in [101, Chapter 18]).

Characterising general abstract conditions that imply the presence of a cut-off is one of the major open problems in the theory of Markov processes. The recent article [129] identifies a general sufficient condition that implies the presence of a cut-off for so-called *non-negatively curved* Markov chains. The follow-up article [130] shows a criterion that is equivalent to the presence of a cut-off under the assumptions of sparsity and fast mixing, which provides a unified framework that is applicable to many previously studied models. These results use a specific quantity from information theory called

varentropy. There is some hope that arguments based on varentropy may yield the so-far elusive general conditions for the presence of a cut-off.

Cut-off has also been observed in the special setting of random walks on random graphs (see also Section 1.2.2). We refer the reader to [71, Section 1.2.1] for a historical overview of the field. We specifically point out the work in [103, 26, 21], where cut-off on time scale $\log n$ was established for both simple and non-backtracking random walks on a general class of sparse *undirected* random graphs with *good expansion properties*. More recently, cut-off has been established for random graphs with community structure [75, 20], and for random graphs with an added random matching [76].

The setting of random walks on *dynamic* random graphs is even more demanding. This line of research started only recently. The first step in this direction was [124], which studies the mixing time of a random walk on subcritical dynamic percolation on $\mathbf{Z}^d$. Later, in [123], the case of supercritical dynamic percolation on $\mathbf{Z}^d$ was considered. Along the same lines, [138] established cut-off on dynamic supercritical Erdős-Rényi graphs, which can be seen as dynamic percolation on the complete graph. A different line of research was started in [9, 10], which studied the mixing properties of a non-backtracking random walk on a configuration model equipped with rewiring dynamics, under mild conditions on the degree sequence. The presence of a cut-off is controlled by the rewiring rate of the dynamics, which introduces an interesting *trichotomy* in the mixing profile. Chapter 2 can be seen as a direct generalisation of this research line, with the aim to push the coupling technique implicitly introduced in [10] to its limit and to show its applicability to a wide class of dynamic random graphs. Chapter 3 extends the analysis in a different direction, where the underlying dynamic geometry is disconnected, but nevertheless allows for a frequent merging and splitting of its constituent components.

§1.2.2 Random graphs

A random graph is a graph-valued random variable. The first random graph models were studied by sociologists in the early twentieth century [111]. Later, they were rediscovered by mathematical biologists [137], and soon thereafter they became an object of interest in their own right, starting with the seminal papers [53, 67]. Recently, the interest in the field was reignited after physicists started to study emergent networks in society and nature (see [18, 19, 115, 116, 113, 150]), a field nowadays known as *complex networks*. There are many good introductory books for random graphs, and we choose to highlight two of them: [62] for a combinatorial approach and the series [78, 79] for a probabilistic treatment.

In this section, we will focus on sparse random graphs, which can be informally defined as graphs where the number of edges is comparable to the number of vertices. More specifically, we will focus on two important random graph models in their sparse regimes, the *Erdős-Rényi random graph* and the *configuration model*.

Erdős-Rényi random graph. There are two models commonly referred to as the Erdős-Rényi random graph. The first, introduced in [67], can be seen as an edge percolation on the complete graph K_n, where edges are kept with probability p, inde-

pendently of other edges. We will call this the Gilbert-Erdős-Rényi graph $G^{GER}(n,p)$. The second, introduced in [53], is a graph chosen uniformly at random from all simple graphs with M edges and n vertices. We will call this the Erdős-Rényi random graph $G^{ER}(n,M)$.

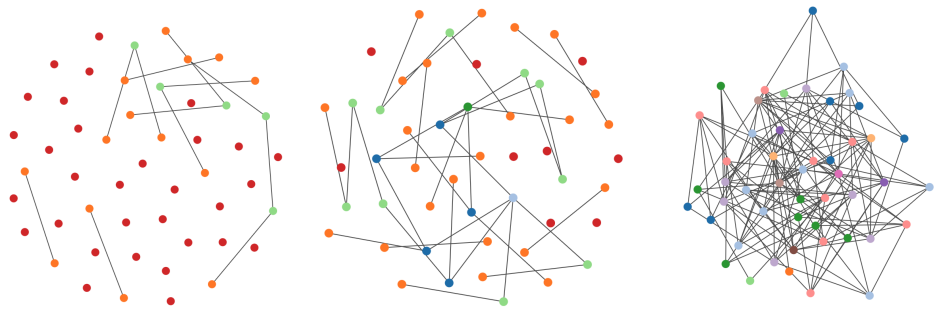


Figure 1.1: Gilbert-Erdős-Rényi graph $G^{GER}(n,p)$ for $n = 50$ and different values of p: 0.01 for the left graph, 0.03 for the middle one and 0.1 for the right one. Created using tools available on `https://www.networkpages.nl/`.

The two models are closely related. Before exploring this connection, we need to introduce the concept of a *monotone graph property*. A graph property $\mathcal{P}$ is monotone if every subgraph of a graph with that property also has this property. It turns out that the two versions of the Erdős-Rényi graph are, under mild conditions, asymptotically equivalent with respect to the laws of monotone graph properties in the limit as $n \to \infty$. A precise statement of this asymptotic equivalence can be found in [62, Theorem 1.4]. A classical example of a property where asymptotic equivalence fails is the (non-monotone) property "the graph has an even number of edges". While for $G^{ER}(n,M)$ the law of this property is degenerate and fully determined by M, for $G^{GER}(n,p)$ the law is non-degenerate. Thanks to this equivalence result, it is often enough to study one of the two versions of the Erdős-Rényi graph. For this reason, we can restrict ourselves to the perspective of the "combinatorial" Erdős-Rényi graph $G^{ER}(n,M)$ in the rest of this section.

Erdős-Rényi graphs can be partitioned into 3 parameter regimes, based on the behaviour of the largest component:

(a) **Sub-critical regime.** In this regime, the Erdős-Rényi graph consists of small components, which are mostly trees. It corresponds to the setting when $M = cn$, with $c < \frac{1}{2}$. The size of the largest connected component in this parameter range is $O_{\mathbb{P}}(\log n)$ and the graph is typically a forest.

(b) **Critical window.** Famously, the Erdős-Rényi graph exhibits a so-called *phase transition*. The critical window corresponds to the parameter range $M = \frac{n}{2} \pm O(n^{2/3})$. The largest connected component in the middle of this regime, i.e., when $M = \frac{n}{2}$, has size $O_{\mathbb{P}}(n^{2/3})$, and connected components are typically trees. The precise behaviour of the Erdős-Rényi graph in the critical regime is a delicate question, but the answer to it is not essential for this dissertation. We refer the interested reader to the seminal paper [86], which is one of the most detailed

studies to date of the critical Erdős-Rényi graph.

(c) **Super-critical regime.** This regime corresponds to the parameter range of $M = cn$ with $c > \frac{1}{2}$. On an event occurring with high probability, the graph contains a unique connected component of size $\Theta(n)$, and all other connected components are of size $O(\log n)$. More precisely, the following theorem applies:

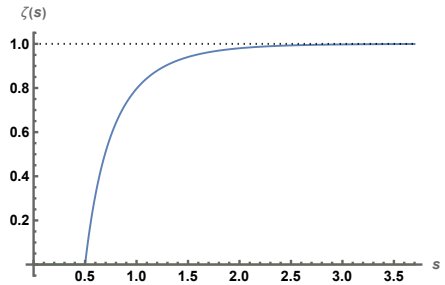


Figure 1.2: Graph of $\zeta(s)$. Note that the mapping $s \to \zeta(s)$ is monotone, $\zeta(s) = 0$ for $s \in [0, \frac{1}{2}]$, and $\zeta(s)$ approaches 1 as $s \to \infty$.

Theorem 1.2.1 (Adapted from [62], Theorem 2.14). *If $M = cn \leq \binom{n}{2}$, $c > \frac{1}{2}$, then $G^{ER}(n, M)$ with high probability consists of a* unique *giant component with $n[\zeta(c)+o(1)]$ edges, where $\zeta(c) \in (0, 1)$ is the positive solution of the equation $1 - \zeta(c) = e^{-2c\zeta(c)}$, and all the other connected components are of size at most $O(\log n)$.*

The super- and sub-critical regimes are connected via the so-called *discrete duality principle*. This says that the non-giant components of the supercritical Erdős-Rényi graph behave similarly to the connected components in a subcritical Erdős-Rényi model (see [79, Theorem 4.15] for a precise statement).

Erdős-Rényi graphs are among the most studied random graph models, and many of their fine structural properties have been identified. Instead of reiterating these results here, we point the reader to [36, 87, 79] for further details and references. While the Erdős-Rényi graph is of immense theoretical and historical importance, it can be too "homogeneous" for some practical applications. Other random graph models, such as the *configuration model*, can be used to work around these limitations.

Configuration model. Much of the following is inspired by [78, Chapter 7].

In practice, we might want to model a graph of which we only know the *degrees* of the individual vertices, but we have no information about the realisation of the edges. In such cases, the *configuration model* is a suitable tool. Alternatively, this model is useful in cases where we wish to impose *restrictions* on the degree sequence, for instance, that all vertices have a degree larger than some constant. The original motivation for the introduction of the configuration model in [35] was to study the asymptotic behaviour of the number of regular graphs via a probabilistic argument.

For any degree sequence $\boldsymbol{d} = (d_1, \ldots, d_n)$ such that $S = \sum_{i=1}^{n} d_i$ is even, we can sample the configuration model $\mathsf{CM}(\boldsymbol{d})$ using the following procedure:

1. For every vertex $v \in V \equiv \{1, \ldots, n\}$, attach d_v half-edges to v and declare all half-edges unpaired.

2. Pick 2 unpaired half-edges uniformly at random, create an edge between the vertices to which the half-edges were connected, and declare these two half-edges paired.

3. Repeat step 2 as long as there are unpaired half-edges left.

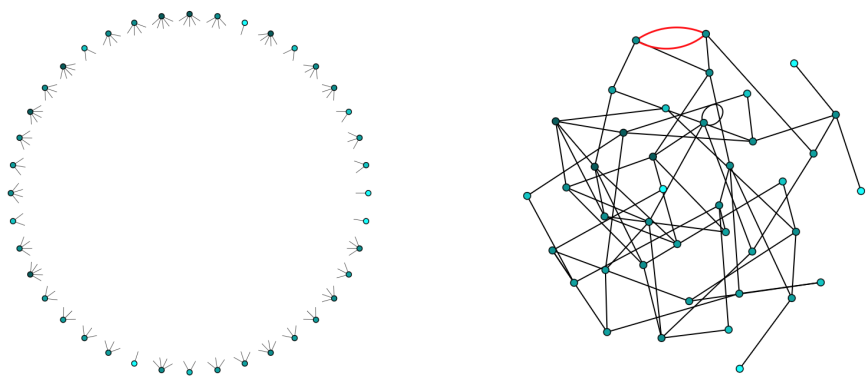


Figure 1.3: Possible initial and final states in the configuration model sampling algorithm. Created using tools available on `https://www.networkpages.nl/`.

Since S is even, the above algorithm is well-defined and terminates as long as S is finite. The procedure above also explains the origin of the name "half-edges": in the construction above, we can see these auxiliary objects as stubs of edges attached to the individual vertices that are waiting to be paired with another half-edge to create a proper edge.

The procedure outlined above need not result in a simple graph. It can happen that the resulting *multigraph* has self-loops and multiple edges, which might be problematic in some applications. The total number of configurations (i.e., different pairings of half-edges) with degree sequence $\boldsymbol{d}$ is given by $(S-1)!! = (S-1)(S-3) \times \ldots \times 1$, and the law of $\mathsf{CM}(\boldsymbol{d})$ is the uniform law over these $(S-1)!!$ possible configurations. Note that a resulting multigraph can be represented by multiple configurations, since half-edges attached to the same vertex are indistinguishable. Consequently, $\mathsf{CM}(\boldsymbol{d})$ is *not* distributed uniformly over all multigraphs with degree sequence $\boldsymbol{d}$. For a multigraph G with degree sequence $\boldsymbol{d}$, [78, Proposition 7.7] gives the explicit probability that a sample of $\mathsf{CM}(\boldsymbol{d})$ yields G. Furthermore, if we condition on $\mathsf{CM}(\boldsymbol{d})$ being simple, then the resulting conditional law is the uniform law over simple graphs with the degree sequence $\boldsymbol{d}$ [78, Proposition 7.15].

Like the Erdős-Rényi graph, the configuration model has been studied extensively and there are textbooks that cover detailed structural results. We point the interested reader to [62, Chapter 10], [78, Chapter 7] and [79, Chapters 4 and 7].

§1.2.3 Dynamic permutations

Permutations are fundamental mathematical objects that have a prominent place in many subfields of mathematics. A permutation is a bijection from a set A to itself. Here, we will take A to be the set of integers $A = \{1, \dots, n\}$. Denote by S_n the set of all permutations of $\{1, \dots, n\}$. If we equip S_n with the natural operation of map composition $\circ$, then we obtain the group $(S_n, \circ)$, commonly known as the *symmetric group* of degree n.

Introductory definitions. Before introducing the central model of this section, let us start with some definitions. A permutation is called a *cycle* of size k if there exists an element $x \in A$ such that $\pi^k(x) = x$ and all the other elements that cannot be obtained by applying π to x are mapped onto themselves. A cycle of size 1 is called a *fixed point* of a permutation. Any permutation $\pi \in S_n$ admits a decomposition into (element-wise) disjoint cycles, which is unique up to the order of composition of these cycles. This inspires a particular way of writing down transpositions, sometimes known as the *cyclic notation*, where a transposition is expressed as a composition of its constituent cycles. Since it is customary to omit the composition operator, the permutation $\pi \in S_4$ that maps $1 \to 2$, $2 \to 4$, $3 \to 3$, $4 \to 1$ can be written as $\pi = (1, 2, 4)(3)$. When the degree of the permutation (i.e., the size of the permuted set, here n) is obvious from the context, we can even omit the fixed points and write $\pi = (124)$. Finally, denote by $T_n \subset S_n$ the set all permutations from S_n that contain one cycle of size 2 while all the other cycles are of size 1. The elements of T_n are called transpositions, since they transpose 2 elements of A and map all the other elements to themselves.

Dynamic permutation. We will restrict our focus to a particular permutation-valued Markov chain. This Markov chain evolves as follows. Fix $n \in \mathbb{N}$ and start the Markov chain from the identity permutation, $X_0 = \mathrm{Id} \in S_n$. The transition rule is that at time $t \in \mathbb{N}$ we sample a uniform transposition $\pi_t \in T_n$ and compose it with the previous permutation. In symbols, for any $t \in \mathbb{N}$, $X_t = X_{t-1} \circ \pi_t$, where $\pi_t \sim \mathsf{Unif}(T_n)$. We will call this Markov chain a *dynamic permutation*.

From a different perspective, the dynamic permutation can be seen as a coagulation-fragmentation process on the level of permutation cycles. When we apply a transposition that swaps two elements in different permutation cycles, these two cycles merge, while when the two elements being swapped belong to the same cycle, this cycle splits into two.

The stationary distribution of the dynamic permutation is the uniform measure over S_n, and the mixing time is $\frac{1}{2} n \log n + O(n)$ [49]. Furthermore, this model (or more precisely, a model very closely related to it) was the first model shown to exhibit the *cut-off* phenomenon [49]. Both of these results were obtained by a clever use of the *representation theory* of the symmetric group. This situation was unsatisfactory to some probabilists, who were longing for a coupling-based argument for the mixing time (e.g., see open problems in [131]). A method explored in the textbook [5] yields the asymptotic scaling of $O(n^2)$ for the mixing time, which is too coarse compared to the representation-theoretic approach. A *non-Markovian* coupling that gives the estimate

$t_{\text{mix}} = Cn \log n + O(n)$, for some constant C, was devised only much later [41]. A different probabilistic approach that yields the estimate $t_{\text{mix}} = O(n \log n)$, based on *strong stationary times*, was developed earlier in [106].

Cycle structure of the dynamic permutation. An interesting question about the dynamic permutation is the distribution of its *cycle structure*, i.e., the normalized (divided by n) order statistics of sizes of cycles in the cycle decomposition of the dynamic permutation. Already [58] gives a coupling proof that the uniform random permutation has a cycle structure that, as $n \to \infty$, converges in distribution to $\mathsf{PoiDir}(1)$, the Poisson-Dirichlet measure with parameter 1. We refer the reader to [83] for an overview of its properties.

The $\mathsf{PoiDir}(1)$ measure is invariant with respect to the dynamics obtained as the $n \to \infty$ limit of the dynamics generated by the dynamic permutation. This limit process is sometimes called the *continuous coagulation-fragmentation process*. While $\mathsf{PoiDir}(1)$ is one of the invariant measures of the aforementioned process, it is not clear whether it is the only one. Vershik conjectured that the $\mathsf{PoiDir}(1)$ measure is indeed the unique invariant measure [143] and this conjecture was answered affirmatively in [48]. A major breakthrough occurred in the article [135], which gives a different proof that the $\mathsf{PoiDir}(1)$ measure is the unique invariant measure of the continuous coagulation-fragmentation process. This article also shows that the $\mathsf{PoiDir}(1)$ structure of cycles occurs in the dynamic permutation much earlier than the mixing time. More precisely, there is a subset of the dynamic permutation that, in the limit as $n \to \infty$ and for any $c > \frac{1}{2}$, attains this equilibrium cycle structure already after cn steps (see [135, Theorem 1.1] for a precise statement).

Many ideas that became standard in the study of dynamic permutations were introduced in [135], such as *Schramm's coupling* and the representation of the dynamic permutation as a graph growth process. These ideas are discussed further in Chapter 3 and Appendix B, where we make use of them.

Later developments. The techniques highlighted above were later expanded and refined in different directions. For example, Schramm's coupling was further refined in the follow-up article [27], where a modified version of this coupling was used to study the mixing of dynamic permutations endowed with a more general dynamics. We also mention [29], which contains a detailed account of Schramm's coupling. On the other hand, group-theoretic methods were further explored in [6, 25] and follow-up articles.

Close relatives of the dynamic permutation model have been studied extensively under different names: the *mean-field Tóth model* [142], the *interchange process* on the complete graph (see [6, 29, 74] and references therein), or the *multi-urn Bernoulli–Laplace diffusion model* [128], where our setting corresponds to a particular choice of parameters.

§1.2.4 Satellite interferometers and their communication network

An interferometer is an extremely sensitive measurement device that exploits interference phenomena in electromagnetic waves. We do not aim to explain the operational

principles of interferometry in this section, and we refer the interested reader to the monograph [141]. The features of an interferometer that are important for Chapter 4 are:

- An astronomical interferometer usually consists of *multiple* telescopes. These can be, for example, multiple satellites, each of them observing the same object. More conventionally, they are different ground-based detectors observing the same signal.
- The relative position of the individual detectors that form an interferometer must be known precisely. The line through the physical space that connects the two detectors is often called a *baseline*. Long baselines are important for higher resolution of the resulting image, while shorter baselines put limits on the maximal angular size of a detectable source.

There are many complications arising from an interferometric setup of an experiment, for example, complicated post-processing of the measured data and problems with observation of low-luminosity sources. For Chapter 4, it is essential to know that useful data can be obtained only when a pair of signals from different detectors is processed jointly. This joint processing of two signals is often called *cross-correlation*.

One of the major complications in post-processing of interferometric data obtained by ground-based telescopes is the influence of the atmosphere. This can vary from wind and other meteorological mechanisms, all the way to molecular effects such as scattering and absorption in the atmosphere. Furthermore, in some frequency bands there is significant interference due to human-made signals. One way to avoid these difficulties is to put the interferometer in space. A particularly intriguing idea is to abandon large satellites, such as the recently launched James Webb Space Telescope, in favour of a swarm of tens to hundreds of smaller satellites. One of the major drawbacks of this approach is that the data communication network that facilitates cross-correlation of signals is no longer trivial. The links within such a network need not be bidirectional and may change in time.

There are numerous further complications related to data transmission, such as data loss, various sources of noise, data routing problems, out-of-order arrival of data packets, and many others. These problems, and the techniques used to mitigate them, are discussed neither in this section nor in Chapter 4. If we choose not to consider these effects, we can restrict the analysis solely to the graph induced by the communication network.

If the swarm does not have a mechanism that enforces a particular position of the satellites, the positions of the satellites, and therefore also the communication network, are governed by *spatial randomness*. To model this type of network, we can use *spatial random graphs*, which are random graphs where the vertices are embedded into some space chosen to reflect the details of the network. In the case of a satellite swarm, a typical choice is a finite box $B \subset \mathbb{R}^3$. Furthermore, if the motion of the satellites is slow enough, then it suffices to consider *static* graphs, since changes in the communication network over relevant time scales will be negligible.

One of the natural *spatial random graph* models that can be used to describe this type of communication network is the k-nearest neighbour graph.

The k-nearest neighbour graph. A k-nearest neighbour graph (hereafter, k-NN graph) is a spatial random graph constructed over a set of vertices V such that $|V| > k$. Furthermore, all the vertices belong to the same metric space (B, d). For every vertex $v \in V$, draw k edges that connect v to its k nearest neighbours with respect to the metric d. This model has received substantial attention, both in mathematics and in engineering [1, 15, 16, 14, 52]. For example, [52] discusses probabilistic properties, like the expected number of connected components, of the k-NN graph over a random set of points.

The study of connected components in the k-NN graph has a twisted history. Originally, engineers believed that there exist so-called *magic numbers*, i.e., universal values of k that in typical situations guarantee connectivity of the k-NN graph [94, 117]. It has been suggested that these "magic numbers" can be any of the integers between 3 and 8. A breakthrough in this line of research was made in the article [153]. The authors were working in a setting where nodes were placed inside a square of finite size, and their positions were governed by a Poisson point process. The authors realized that to maintain connectivity as the area of the square (denoted by A) tends to infinity, while the intensity of the Poisson process remains constant, each node must be connected to a number of neighbours that grows at least as fast as $\log A$. This marks a departure away from the idea of a universal "magic number". Furthermore, the same article provides some bounds for the critical value $c^\star$, multiplying $\log A$, describing the zero-one law related to this notion of connectivity. These results have stimulated rigorous mathematical investigation. In a series of articles [15, 16], the authors formalize the idea that, in the setting of [153], connectivity is indeed obtained when k is of the order $\log A$, and first provide a much tighter bound on the critical multiplication constant ($0.3043 \leq c^\star \leq 0.5139$). Later, they invented a method to precisely compute the critical constant $c^\star$. Further results in these directions can be found in [14].

From an engineering perspective, a major difference between the k-NN graph and a graph that would result from a communication network between transmitters each transmitting to their k nearest neighbours is that the transmission network would be *directed*, unlike the k-NN graph. Nevertheless, k-NN graphs with undirected edges have been used in studies of optimal data routing [32, 31, 105], wireless sensor coverage [90, 84], wireless transmitter localisation [118, 152], or dynamical allocation problems in 6G networks [148].

§1.2.5 Network centrality measures

Real-world networks, both directed and undirected, are often inhomogeneous, with neighbourhoods of vertices varying substantially across the graph. This is not surprising, since many of the phenomena modelled by these networks exhibit the same level of inhomogeneity. A natural question arises: Since the network is inhomogeneous, are any of its vertices more important than others?

On the other hand, what does it even mean to say that a vertex is more important than some other vertex? Consider, for example, a friendship network among a group of people. In this network, individuals are represented by vertices and two individuals

are connected by an edge if they are friends with each other. One reasonable choice would be to call the most important vertex in the network the one with the highest degree, since it represents the person who has the most friends. But perhaps it is not the popular people who are relevant for a particular study, but the people who are members of multiple friend circles. Or we can consider it important that someone's friendship network is tightly-knit, and we would prefer a vertex with its immediate neighbours to form a clique.

The ambiguity demonstrated in the previous example prevents us from defining *the* network centrality measure. In general, a network centrality measure is a mapping from the vertex set V_G to the real numbers. The details of this mapping are chosen to reflect what being central can mean. Some popular examples of network centrality measures are:

- **Degree centrality.** This is simply the degree of a given vertex. For any $v \in V$:

$$R^{\mathrm{deg}}(v) = \deg(v).$$

  Similarly, it is possible to define in-/out-degree centrality for directed graphs.

- **Eigenvector centrality.** First defined in [96], eigenvector centrality is the corresponding element of the left eigenvector of the adjacency matrix A associated to the positive eigenvalue of maximal modulus, normalized so that all its elements sum up to 1. On the other hand, other normalisation are also possible. In symbols:

$$R^{\mathrm{ev}}(v) = \boldsymbol{a}_v,$$

  where $\boldsymbol{a}$ is a vector such that

$$\lambda_{\max}\boldsymbol{a}^T = A\boldsymbol{a}^T \qquad \text{and} \qquad \sum_{i=1}^{|V_G|} \boldsymbol{a}_i = 1.$$

  Katz centrality [91] and PageRank [122] can be seen as further refinements.

- **Betweenness centrality.** Introduced in [61], betweenness centrality is related to the number of shortest paths within the graph that run through a given vertex. It is defined as

$$R^{\mathrm{B}}(v) = \sum_{a,b \in V_G} \frac{\sigma_v(a,b)}{\sigma(a,b)},$$

  where $\sigma(a,b)$ is the number of distinct paths of minimal length between a and b, and $\sigma_v(a,b)$ is the same with the added condition that these paths contain v somewhere between a and b. By convention, for any $a, b \in V_G$, $\sigma(a,a) = 1$ and $\sigma_a(a,b) = \sigma_b(a,b) = 0$. Load centrality [68] is a slight variant of betweenness.

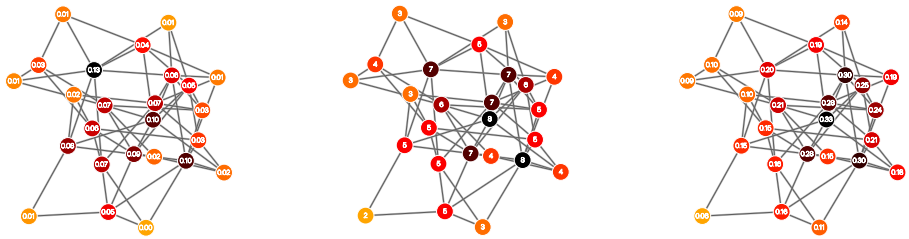


Figure 1.4: Betweenness, degree and eigenvector centrality for the same graph. Darker colours represent higher centrality of nodes. Created using tools available on the website `https://www.networkpages.nl/`.

Other centrality measures can be found in Appendix D or [114, Chapter 7]. To illustrate how fragmented the field is, note that [85] lists more than 400 different centrality measures. It does not help that results about centrality measures are scattered across different journals in different fields. In light of this, it is not surprising that recent comprehensive reviews were written by researchers in various disciplines, such as computer science [34], mathematics [134], electrical engineering [146], and social sciences [30].

There has been some effort to relate various centrality measures to each other. For example, [54] explains the relation between degree and closeness centrality. The fact that such a connection exists is not surprising, since many centrality measures are correlated with each other [102, 119]. On the other hand, there are only few results that explain these correlations. It is more common to see centrality measures compared with respect to their suitability for a particular applied goal, such as connectivity hub identification [108] or keyword extraction from a graph-based representation of text [39].

Naturally, network centrality measures have a prominent place in real-world applications. For example, [2] uses centrality measures to study attack tolerance and robustness of complex networks, [122, 40] use centrality measures in the context of Internet search engines, and [93, 107, 149, 126] use them to study infection spreading. Many further examples can be found in the review articles referenced above.

§1.3 Organisation of this dissertation

Here we outline the organisation of the remainder of this dissertation. The scientific part spans Chapters 2-5. Each of these chapters is based on a different research paper. Appendices are included right after the appropriate chapter, and are labelled by consecutive *letters*. For example, Appendix B contains material relevant for Chapter 3. After the scientific part, Chapter 6 lists some questions spurred by the research contained in Chapters 2-5. Finally, the dissertation concludes with a bibliography, Dutch and English summary, acknowledgements and a short curriculum vitæ.

CHAPTER 2

Linking the mixing times of random walks on static and dynamic random graphs

This chapter is based on the following article:

L. Avena, H. Güldaş, R. van der Hofstad, F. den Hollander, and O. Nagy. Linking the mixing times of random walks on static and dynamic random graphs. *Stoch. Proc. Appl.*, 153:145–182, 2022.

Abstract

In this chapter, which is a culmination of our previous research efforts, we provide a *general framework* for studying mixing profiles of non-backtracking random walks on dynamic random graphs generated according to the configuration model. The quantity of interest is the scaling of the mixing time of the random walk as the number of vertices of the random graph tends to infinity. Subject to mild general conditions, we link two mixing times: one for a *static* version of the random graph, the other for a class of *dynamic* versions of the random graph in which the edges are randomly *rewired*, but the degrees are preserved. With the help of coupling arguments, we show that the link is provided by a specific *randomised stopping time*, namely, the probability that the random walk has not yet stepped along a previously rewired edge.

To demonstrate the utility of our framework, we re-derive our earlier results on mixing profiles for *global* edge rewiring under weaker assumptions, and extend these results to an *entire class* of rewiring dynamics parametrised by the range of the rewiring relative to the position of the random walk. Along the way we establish that all the graph dynamics in this class exhibit the trichotomy we found earlier, namely, no cut-off, one-sided cut-off or two-sided cut-off.

For *interpolations* between *global* edge rewiring, the only Markovian graph dynamics considered here, and *local* edge rewiring (i.e., only those edges that are incident to the random walk can be rewired), we show that the trichotomy splits further into a hexachotomy, namely, three different mixing profiles with no cut-off, two with one-sided cut-off, and one with two-sided cut-off. Proofs are built on a new and flexible coupling scheme, in combination with sharp estimates on the degrees encountered by the random walk in the static and the dynamic version of the random graph. Some of these estimates require sharp control on possible short-cuts in the graph between the edges that are traversed by the random walk.

§2.1 Introduction

In this chapter, we generalise and extend the techniques developed in [9, 10] to cover a *class* of random graph dynamics that satisfy mild regularity conditions. Our core result is a generic link between the mixing times on the static and the dynamic random graph. Subject to mild conditions on the degrees of the vertices and the dynamics of the underlying graph, we show that, up to an error that vanishes as the number of vertices tends to infinity, the total variation distance to the stationary distribution on the *dynamic* random graph is given by the total variation distance on the *static* random graph multiplied by the probability that the random walk has not yet stepped along a previously rewired edge. Phrased in symbols, we show that

$$\mathcal{D}^{\mathrm{dyn}}_{x,\xi}(t) = \mathbb{P}_{x,\xi}(\tau > t)\, \mathcal{D}^{\mathrm{stat}}_{x,\xi}(t) + o_{\mathbb{P}}(1), \tag{2.1}$$

where x is the starting vertex of the random walk, ξ is the starting configuration of the random graph, $\mathcal{D}^{\mathrm{dyn}}_{x,\xi}(t)$ and $\mathcal{D}^{\mathrm{stat}}_{x,\xi}(t)$ are the total variation distance between the distribution of the random walk at time t and the stationary distribution for the dynamic, respectively, the static random graph, and τ is the first time the random walk crosses a rewired edge (see Theorem 2.1.4 below for a precise statement). The latter acts as a randomised stopping time and plays a central role in our analysis.

We emphasise that our approach is no longer tied to one specific random graph dynamics. In fact, we present a *general framework* that can be applied to a large class of random graph dynamics for which the degree structure is preserved, including dynamics that depend on the *position* of the random walk and dynamics that are *non-Markovian*. To do so we use a coupling that works well for non-backtracking random walks. We show that (2.1) holds under *general conditions* that appear to be the weakest possible, and that can be verified in specific examples.

To showcase our framework, we identify the scaling of the random walk mixing time for three choices of the dynamics, where the rewiring is done in a certain range around the current position of the random walk. Depending on the speed and the range of the rewiring, the mixing time may exhibit no cut-off, one-sided cut-off or two-sided cut-off, a trichotomy that was also found in earlier work (see Section 2.1.5 for an extensive literature overview). Interestingly, for a class of dynamics that interpolate between local and global (the "near-to-global" dynamics introduced in Section 2.1.3), we observe a *hexachotomy* – six subregimes (see Figure 2.3 below), two of which include *critical crossover times* where the mixing profile *changes shape*.

From a technical point of view, it is amusing to note that we are able to establish our previous results from [9, 10] under weaker assumptions, despite the fact that we are no longer using techniques that are tailor-made to global dynamics. Furthermore, we believe that near-to-global dynamics is interesting in its own right, not only thanks to its six different mixing subregimes, but also due to a non-trivial challenge of dealing with the influence of so-called "shortcuts" (see Section 2.4.3).

§2.1.1 Model and notation

It is convenient to describe our model in terms of *half-edges*. Write V to denote the vertex set of the graph, $|V| =: n$ the number of vertices, and $\deg(v)$ the degree of

vertex $v \in V$. To each vertex $v \in V$ we associate $\deg(v)$ half-edges, forming the set $H_v := \{h_i\}_{i=1}^{\deg(v)}$. The set of all half-edges is $H := \bigcup_{v \in V} H_v$. We denote the vertex v for which $h \in H_v$ by $v(h) \in V$. If $x, y \in H_v$, $x \neq y$, then we write $x \sim y$ and say that x and y are *siblings* of each other. Using $|X|$ to denote the cardinality of set X, we define the degree of a half-edge $h \in H$ as

$$\deg_H(h) := |\{h_i \in H_{v(h)} \colon\ h_i \sim h\}| = \deg(v(h)) - 1. \tag{2.2}$$

We identify an edge with a pair of half-edges. A *configuration* is a pairing ξ of half-edges with the property that $\xi(h) \neq h$ and $\xi(\xi(h)) = h$ for all $h \in H$. The set of all configurations on H is denoted by Conf_H, and the uniform distribution on Conf_H is denoted by U_{Conf_H}. With a slight abuse of notation, we will use the same symbol ξ to denote the set of pairs of half-edges forming ξ, so $\{x, y\} \in \xi$ means that $\xi(x) = y$ and $\xi(y) = x$. Note that ξ may represent a multi-graph, possibly with self-loops. A random graph corresponding to a configuration where the half-edges are paired uniformly at random is called the *configuration model* (see [35], [78, Chapter 7]). The quantities above depend on n, but this dependence will be mostly suppressed from the notation.

We study Markov chains $\{(X_t, C_t)\}_{t \in \mathbb{N}_0}$, where $X_t \in H$ denotes the *non-backtracking* random walk component and $C_t \in \mathit{Conf}_H$ corresponds to the *evolution of the underlying graph*. The evolution is chosen in such a way that it does not change the degree sequence of the graph (and consequently does not change the stationary distribution of the random walk on the graph), and can be visualised by breaking up pairs of half-edges and pairing them again, both according to prescribed rules. At each time $t \in \mathbb{N}$, we first update the configuration and then let the walk move.

Remark 2.1.1 (Notation). Note that (X_{t-1}, C_{t-1}) is the state just before the transition at time t, while (X_t, C_t) is the state just after the transition at time t. ♦

Our main result concerns the total variation distance between the distribution of the random walk component and the stationary uniform distribution on the set of half-edges U_H, defined as

$$\mathcal{D}^{\mathrm{dyn}}_{x,\xi}(t) := d_{\mathrm{TV}}\left(\mathbb{P}_{x,\xi}(X_t \in \cdot), U_H(\cdot)\right). \tag{2.3}$$

Here, the total variation distance between two probability measures μ and ν on the same finite state space S is defined by

$$d_{\mathrm{TV}}(\mu, \nu) := \frac{1}{2} \sum_{x \in S} |\mu(x) - \nu(x)| = \sum_{x \in S} [\mu(x) - \nu(x)]_+ = \sup_{A \subseteq S} [\mu(A) - \nu(A)], \tag{2.4}$$

We are concerned with the behaviour of $\mathcal{D}_{x,\xi}(t)$ for "typical" choices of x and ξ. We formalise the notion of typicality in the following definition:

Definition 2.1.2 (With high probability). Recall that $n = |V|$ and let $\mu := U_H \times U_{\mathit{Conf}_H}$. A statement that depends on the initial half-edge x and the initial configuration ξ is said to hold *with high probability*, abbreviated whp, if the μ-measure of the set of pairs (x, ξ) for which the statement holds tends to 1 as $n \to \infty$. ■

Another important object is the first time the random walk steps along a previously rewired edge:

Definition 2.1.3 (Randomized stopping time). Let R_t be the set of edges being rewired at time t, $R_{\leq t} := \bigcup_{s=1}^{t} R_s$, and let I_t denote the indicator of the event that the random walk steps along a previously rewired edge at time t, i.e., $I_t = 1$ when $X_{t-1} \in R_{\leq t}$ and $I_t = 0$ otherwise. We define the *randomized stopping time* τ as

$$\tau := \min\{t \in \mathbb{N} \colon\ I_t = 1\}. \tag{2.5}$$

■

Note that, since rewiring happens before the random walk steps, X_{t-1} is the position of the random walk just before it steps over an edge that is rewired at time t.

For $x \in H$ and $\xi \in \mathit{Conf}_H$, we denote by $\mathcal{D}^{\mathrm{stat}}_{x,\xi}(t)$ the total variation distance of the random walk on the static random graph to the stationary uniform distribution U_H at time t, and by $\mathbb{P}_{x,\xi}(\tau > t)$ the probability that $\tau > t$, both given the starting state (x, ξ).

§2.1.2 Mixing for general rewiring mechanisms

The main theorem of this chapter is the following statement linking the total variation distance to the stationary distribution for the static and the dynamic version of the random graph:

Theorem 2.1.4 (Link between static and dynamic mixing). *Suppose that $t = O(\log n)$. Subject to Conditions 2.3.1 and 2.3.4 (see Section 2.3.1), the following holds whp in x and ξ:*

$$\mathcal{D}^{\mathrm{dyn}}_{x,\xi}\big(t\,\big) = \mathbb{P}_{x,\xi}(\tau > t)\, \mathcal{D}^{\mathrm{stat}}_{x,\xi}(t) + o_{\mathbb{P}}(1). \tag{2.6}$$

Conditions 2.3.1 and 2.3.4 are regularity conditions on degree sequence and graph dynamics, respectively. The former is standard in the literature and ensures that the underlying graph is sparse and that the non-backtracking random walk is well-defined. The latter, representing one of the novelties of this article, ensures that the non-backtracking random walk is well-mixed when it steps along a previously rewired edge and the time at which this happens does not depend on the fine details of its past trajectory.

The proof of Theorem 2.1.4 is based on a *coupling argument* in which the random walk on the dynamically rewired random graph is coupled to a modified random walk on the static random graph that at certain random times makes uniform jumps. These jumps correspond to the times at which the random walk steps along a previously rewired edge. The coupling must be good enough to beat the errors in the comparison. A key ingredient of the coupling is that the non-backtracking random walk on the configuration model is whp *self-avoiding* on the scale of the mixing time.

Note that while $\{(X_t, C_t)\}_{t\in\mathbb{N}}$ is Markov, any of the marginals $\{X_t\}_{t\in\mathbb{N}}, \{C_t\}_{t\in\mathbb{N}}$ need not be. The framework we present in this chapter makes no assumptions about $\{C_t\}_{t\in\mathbb{N}}$ being Markov. This is demonstrated in Section 2.4, where "global-to-global"

rewiring is the only considered mechanism for which $\{C_t\}_{t\in\mathbb{N}}$ is Markov, while "near-to-global" and "local-to-global" rewiring correspond to a non-Markovian graph evolution.

§2.1.3 Application to specific rewiring mechanisms

We next consider three choices of random rewiring, referred to as *local-to-global*, *near-to-global* and *global-to-global*, controlled by two parameters: (1) r_n, representing the radius of the ball around the current location of the random walk in which edges are allowed to be rewired with an edge that is drawn uniformly at random from the set of all edges; (2) α_n, representing the probability that an edge in this ball is rewired per unit of time. By rewiring we mean breaking up two pairs of chosen edges into four half-edges and tying these up at random (for details, see Section 2.4).

At every unit of time a *subset* of the edges is rewired. The rewiring of each edge is always with an edge that is chosen *uniformly at random* from the set of *all* edges. For the subset of edges that is rewired we consider three choices:

- **Local-to-global ($r_n = 1$):**
  The edge that corresponds to the current position of the random walk has probability α_n to be rewired.
- **Near-to-global ($1 < r_n < r_{\max}$):**
  All the edges in the r_n-ball around the current position of the random walk have probability α_n to be rewired, independently of each other.
- **Global-to-global ($r_n = r_{\max}$):**
  All the edges have probability α_n to be rewired, independently of each other.

Here, $r_{\max}$ is the maximal radius (see (2.12) below), provided that the graph is connected (which happens whp under the conditions that will be stated below).

Global-to-global rewiring was considered in [9] and [10], while local-to-global rewiring was considered in an unpublished chapter of the doctorate thesis [71]. In the present chapter, however, we prove results under *weaker assumptions*. For an overview of previous work, see Section 2.1.5. Near-to-global rewiring is new and turns out to hold surprises:

Theorem 2.1.5 (Scaling of cross-rewired time). *Suppose that* $\lim_{n\to\infty}\alpha_n = 0$ *and* $t = O(\log n)$. *Subject to Conditions 2.3.1(R1) and (R3) below, the following hold* whp *in* x *and* ξ*:*

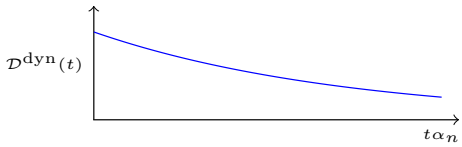
(A) *For local-to-global rewiring defined in Section 2.4.2:*

$$\mathbb{P}_{x,\xi}(\tau > t) = o_{\mathbb{P}}(1) + \mathrm{e}^{-c}, \quad c \in (0,\infty), \qquad t = \lfloor c/\alpha_n \rfloor. \tag{2.7}$$

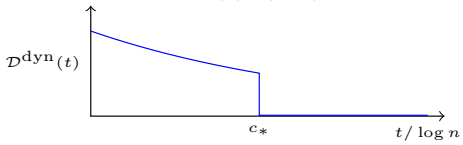
(B) *For near-to-global rewiring defined in Section 2.4.3, subject to Condition 2.3.5 below:*

(a) *If* $\lim_{n\to\infty}\alpha_n r_n^2 = \infty$, *then*

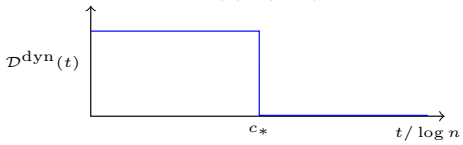
$$\mathbb{P}_{x,\xi}(\tau > t) = o_{\mathbb{P}}(1) + \mathrm{e}^{-c^2/2}, \quad c \in (0,\infty), \qquad t = \lfloor c/\sqrt{\alpha_n} \rfloor. \tag{2.8}$$



Regime (1): (2.15).

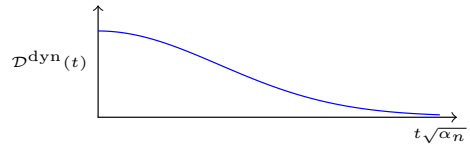


Regime (2): (2.16).

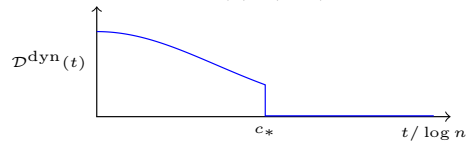


Regime (3): (2.17).

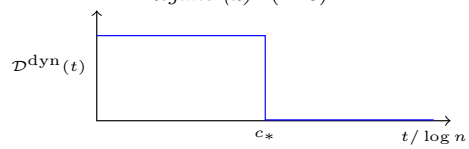
Figure 2.1: Trichotomy of $\mathcal{D}^{\mathrm{dyn}}(t)$ for local-to-global rewiring (Corollary 2.1.7).



Regime (1): (2.24).

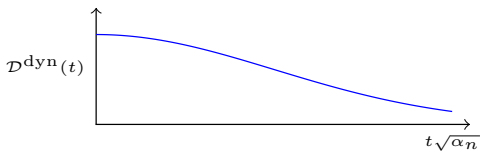


Regime (2): (2.25).

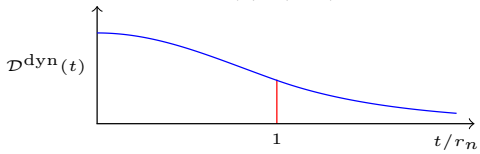


Regime (3): (2.26).

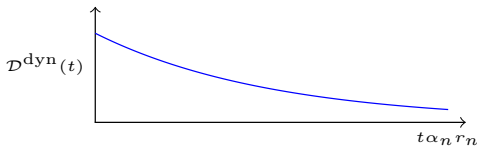
Figure 2.2: Trichotomy of $\mathcal{D}^{\mathrm{dyn}}(t)$ for global-to-global rewiring (Corollary 2.1.9).



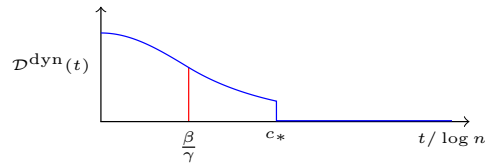
Regime 1(a): (2.18).



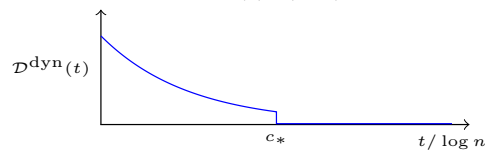
Regime 1(b): (2.19).



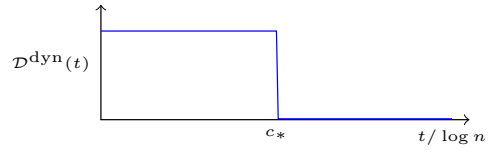
Regime 1(c): (2.20).



Regime 2(b): (2.21).



Regime 2(c): (2.22).



Regime 3(c): (2.23).

Figure 2.3: Hexachotomy of $\mathcal{D}^{\mathrm{dyn}}(t)$ for near-to-global rewiring (Corollary 2.1.8). The red lines indicate a crossover in the shape of the curve.

(b) *If* $\lim_{n\to\infty} \alpha_n r_n^2 = \beta \in (0,\infty)$*, then*

$$\mathbb{P}_{x,\xi}(\tau > t) = o_{\mathbb{P}}(1) + \begin{cases} \mathrm{e}^{-\beta c^2/2}, & c \in (0,1], \\ \mathrm{e}^{-\beta(2c-1)/2}, & c \in (1,\infty), \end{cases} \qquad t = \lfloor c\, r_n \rfloor, \tag{2.9}$$

(c) *If* $\lim_{n\to\infty} \alpha_n r_n^2 = 0$*, then*

$$\mathbb{P}_{x,\xi}(\tau > t) = o_{\mathbb{P}}(1) + \mathrm{e}^{-c}, \quad c \in (0,\infty), \qquad t = \lfloor c/\alpha_n r_n \rfloor. \tag{2.10}$$

(C) *For global-to-global rewiring defined in Section 2.4.4:*

$$\mathbb{P}_{x,\xi}(\tau > t) = o_{\mathbb{P}}(1) + \mathrm{e}^{-c^2/2}, \quad c \in (0,\infty), \qquad t = \lfloor c/\sqrt{\alpha_n} \rfloor. \tag{2.11}$$

Note that the tail probability $\mathbb{P}_{x,\xi}(\tau > t)$ exhibits a three-way split for near-to-global rewiring, with an additional crossover at time $t = r_n$ when $\lim_{n\to\infty} \alpha_n r_n^2 = \beta \in (0,\infty)$.

Condition 2.3.5 says that the empirical degree distribution converges to a limit as $n \to \infty$, and so do its first and second moments. It implies that whp the radius (i.e., the typical distance between vertices) of the random graph is

$$r_{\max} = [1 + o_{\mathbb{P}}(1)] \frac{\log n}{\log \nu}, \tag{2.12}$$

where ν is the size-biased mean of the limiting empirical degree distribution [79, Theorem 7.1], which is assumed to satisfy $\nu \in (1,\infty)$. Thus, for near-to-global rewiring we can only choose

$$r_n = [\rho + o(1)] \log n, \quad \rho \in [0, \rho_{\max}), \qquad \rho_{\max} = \frac{1}{\log \nu}. \tag{2.13}$$

Condition 2.3.5 is needed for Theorem 2.1.5(B) only. The fact that it is not needed for Theorem 2.1.5(C) weakens the conditions in [9, 10]. We expect Theorem 2.1.5(B) to fail without Condition 2.3.5. Namely, when the degree distribution has infinite variance, graph distances are of smaller order than $\log n$, and in fact are of order $\log\log n$ under an appropriate power-law assumption on the empirical degree distribution [43, 79, 81, 82]. In the latter setting, for $r_n = c \log n$ we expect near-to-global rewiring to behave similarly as global-to-global rewiring.

In order to apply Theorem 2.1.4, we need to also control $\mathcal{D}^{\mathrm{stat}}_{x,\xi}(t)$. For this we use the following result from [21], which requires *additional regularity conditions* stronger than Condition 2.3.1 (see Appendix A.2 for further details):

Theorem 2.1.6 (Scaling of static mixing time). *Subject to* (A.3) *and Condition A.2.1, the following holds whp in* x *and* ξ*:*

$$\mathcal{D}^{\mathrm{stat}}_{x,\xi}(t) = \begin{cases} 1 - o_{\mathbb{P}}(1), & \text{if} \quad t = \lfloor c \log n \rfloor, \quad c < c_*, \\ o_{\mathbb{P}}(1), & \text{if} \quad t = \lfloor c \log n \rfloor, \quad c > c_*, \end{cases} \tag{2.14}$$

where $c_* \in (0,\infty)$ *is the constant defined in* (A.3)*.*

Combining Theorems 2.1.4–2.1.6 we end up with the following results:

Corollary 2.1.7 (Scaling of dynamic mixing time for local-to-global rewiring). *Consider the local-to-global rewiring defined in Section 2.4.2. Suppose that* $\lim_{n\to\infty}\alpha_n = 0$ *and* $t = O(\log n)$*. Subject to Condition 2.3.1(R1), Condition A.2.1 and* (A.3)*, the following hold whp in* x *and* ξ*:*

(1) *If* $\lim_{n\to\infty}\alpha_n \log n = \infty$*, then*

$$\mathcal{D}^{\mathrm{dyn}}_{x,\xi}\big(\lfloor c/\alpha_n\rfloor\big) = o_{\mathbb{P}}(1) + \mathrm{e}^{-c}, \quad c \in [0,\infty). \tag{2.15}$$

(2) *If* $\lim_{n\to\infty}\alpha_n \log n = \gamma \in (0,\infty)$*, then*

$$\mathcal{D}^{\mathrm{dyn}}_{x,\xi}\big(\lfloor c\log n\rfloor\big) = o_{\mathbb{P}}(1) + \begin{cases} \mathrm{e}^{-\gamma c}, & c \in [0,c_*),\\ 0, & c \in (c_*,\infty). \end{cases} \tag{2.16}$$

(3) *If* $\lim_{n\to\infty}\alpha_n \log n = 0$*, then*

$$\mathcal{D}^{\mathrm{dyn}}_{x,\xi}\big(\lfloor c\log n\rfloor\big) = o_{\mathbb{P}}(1) + \begin{cases} 1, & c \in [0,c_*),\\ 0, & c \in (c_*,\infty). \end{cases} \tag{2.17}$$

Corollary 2.1.8 (Scaling of dynamic mixing time for near-to-global rewiring). *Consider the near-to-global rewiring defined in Section 2.4.3. Suppose that* $\lim_{n\to\infty}\alpha_n = 0$ *and* $t = O(\log n)$*. Subject to Condition 2.3.1(R1), Condition 2.3.5, Condition A.2.1 and* (A.3)*, the following hold whp in* x *and* ξ*:*

(1) *If* $\lim_{n\to\infty}\alpha_n r_n \log n = \infty$ *and*

(a) $\lim_{n\to\infty}\alpha_n r_n^2 = \infty$*, then*

$$\mathcal{D}^{\mathrm{dyn}}_{x,\xi}\big(\lfloor c/\sqrt{\alpha_n}\rfloor\big) = o_{\mathbb{P}}(1) + \mathrm{e}^{-c^2/2}, \quad c \in [0,\infty). \tag{2.18}$$

(b) $\lim_{n\to\infty}\alpha_n r_n^2 = \beta \in (0,\infty)$*, then*

$$\mathcal{D}^{\mathrm{dyn}}_{x,\xi}\big(\lfloor c\, r_n\rfloor\big) = o_{\mathbb{P}}(1) + \begin{cases} \mathrm{e}^{-\beta c^2/2}, & c \in (0,1],\\ \mathrm{e}^{-\beta(2c-1)/2}, & c \in (1,\infty). \end{cases} \tag{2.19}$$

(c) $\lim_{n\to\infty}\alpha_n r_n^2 = 0$*, then*

$$\mathcal{D}^{\mathrm{dyn}}_{x,\xi}\big(\lfloor c/\alpha_n r_n\rfloor\big) = o_{\mathbb{P}}(1) + \mathrm{e}^{-c}, \quad c \in [0,\infty). \tag{2.20}$$

(2) *If* $\lim_{n\to\infty}\alpha_n r_n \log n = \gamma \in (0,\infty)$ *and*

(b) $\lim_{n\to\infty}\alpha_n r_n^2 = \beta \in (0,\infty)$*, then*

$$\mathcal{D}^{\mathrm{dyn}}_{x,\xi}\big(\lfloor c\log n\rfloor\big) = o_{\mathbb{P}}(1) + \begin{cases} \mathrm{e}^{-(\gamma c)^2/2\beta}, & c \in [0,\beta/\gamma],\\ \mathrm{e}^{-(2\gamma c-\beta)/2}, & c \in (\beta/\gamma, c_*),\\ 0, & c \in (c_*,\infty). \end{cases} \tag{2.21}$$

(c) $\lim_{n\to\infty}\alpha_n r_n^2 = 0$*, then*

$$\mathcal{D}^{\mathrm{dyn}}_{x,\xi}\big(\lfloor c\log n\rfloor\big) = o_{\mathbb{P}}(1) + \begin{cases} \mathrm{e}^{-\gamma c}, & c \in (0,c_*),\\ 0, & c \in (c_*,\infty). \end{cases} \tag{2.22}$$

(3) *If* $\lim_{n\to\infty} \alpha_n r_n \log n = 0$ *and*

(c) $\lim_{n\to\infty} \alpha_n r_n^2 = 0$, *then*

$$\mathcal{D}^{\mathrm{dyn}}_{x,\xi}\big(\lfloor c \log n \rfloor\big) = o_{\mathbb{P}}(1) + \begin{cases} 1, & c \in [0, c_*), \\ 0, & c \in (c_*, \infty). \end{cases} \tag{2.23}$$

Corollary 2.1.9 (Scaling of dynamic mixing time for global-to-global rewiring). *Consider the global-to-global rewiring defined in Section 2.4.4. Suppose that* $\lim_{n\to\infty} \alpha_n = 0$ *and* $t = O(\log n)$. *Subject to Condition 2.3.1(R1), Condition A.2.1 and* (A.3), *the following hold whp in* x *and* ξ*:*

(1) *If* $\lim_{n\to\infty} \alpha_n (\log n)^2 = \infty$, *then*

$$\mathcal{D}^{\mathrm{dyn}}_{x,\xi}\big(\lfloor c/\sqrt{\alpha_n} \rfloor\big) = o_{\mathbb{P}}(1) + \mathrm{e}^{-c^2/2}, \quad c \in [0, \infty). \tag{2.24}$$

(2) *If* $\lim_{n\to\infty} \alpha_n (\log n)^2 = \gamma \in (0, \infty)$, *then*

$$\mathcal{D}^{\mathrm{dyn}}_{x,\xi}\big(\lfloor c \log n \rfloor\big) = o_{\mathbb{P}}(1) + \begin{cases} \mathrm{e}^{-\gamma c^2/2}, & c \in [0, c_*), \\ 0, & c \in (c_*, \infty). \end{cases} \tag{2.25}$$

(3) *If* $\lim_{n\to\infty} \alpha_n (\log n)^2 = 0$, *then*

$$\mathcal{D}^{\mathrm{dyn}}_{x,\xi}\big(\lfloor c \log n \rfloor\big) = o_{\mathbb{P}}(1) + \begin{cases} 1, & c \in [0, c_*), \\ 0, & c \in (c_*, \infty). \end{cases} \tag{2.26}$$

Note that the dynamic mixing time exhibits a *trichotomy* that distinguishes between *fast* dynamics (regime (1)), *moderate* dynamics (regime (2)) and *slow* dynamics (regime (3)). There is *no* cut-off for fast dynamics, *one-sided* cut-off (at $c = c_*$) for moderate dynamics, and *two-sided* cut-off (at $c = c_*$) for slow dynamics.

Note that near-to-global rewiring exhibits an even richer structure with a *hexa-chotomy*. See Figs. 2.1–2.3 for the various scaling shapes (where the indices x and ξ are suppressed).

Remark 2.1.10 (Role of Condition 2.3.4). Corollaries 2.1.7–2.1.9 do not mention Condition 2.3.4 explicitly, even though this is needed for Theorem 2.1.4. The reason is that the three rewiring mechanisms under consideration satisfy Condition 2.3.4, as shown in Section 2.4. ♦

§2.1.4 Discussion

1. Each of the three choices of rewiring shows a *trichotomy* between fast dynamics ($\alpha_n r_n \gg 1/\log n$), moderate dynamics ($\alpha_n r_n \asymp 1/\log n$) and slow dynamics ($\alpha_n r_n \ll 1/\log n$), with $r_n = 1$ for local-to-global rewiring, $1 < r_n < r_{\max}$ for near-to-global rewiring and $r_n = r_{\max}$ for global-to-global rewiring. For fast dynamics the mixing time is of smaller order than $\log n$, which is the mixing time on the static random graph, and so *speed-up* occurs. For moderate and slow dynamics the mixing time is of order $\log n$, and so *no speed-up* occurs. The one-sided cut-off for moderate dynamics

shows that there is a *competition* between static and dynamic. For fast dynamics only Conditions 2.3.1 and 2.3.4 are needed, while for moderate and slow dynamics (A.3) and Condition A.2.1 are needed as well. For fast dynamics the scaling does *not* depend on the choice of degrees, subject to the mild regularity imposed by Condition 2.3.1. On other hand, for moderate and slow dynamics it does, because the constant c_* equals the limit as $n \to \infty$ of the empirical average of the logarithm of the degrees of the half-edges.

2. Whereas for local-to-global and global-to-global rewiring the trichotomy controls the scaling, for near-to-global rewiring several *subregimes* show up. In particular, crossovers in the mixing time occur at critical values of the scaling parameter c (see (2.19) and (2.21)). These arise from a crossover in the cross-rewired time that appears as soon as $t \asymp r_n$ (see (2.9)). What happens is that all edges on the r_n-future of the path can be rewired before the random walk reaches them, but only until time $t - r_n$: for any time $s \in (t - r_n, t]$ only $t - s$ edges are left on the future path until time t. The extra condition in Condition 2.3.5 ensures that whp the r_n-balls carried around by the random walk do not overlap significantly, i.e., short-cuts of length $\leq r_n$ are negligible until time $t = O(\log n)$.

3. Regime (2b) for near-to-global rewiring corresponds to $\rho = \beta/\gamma$. Subject to Condition 2.3.5 we have

$$\frac{1}{c_*} = \sum_{m \in \mathbb{N}} p^\star(m) \log m, \tag{2.27}$$

with

$$p^\star(m) = \frac{1}{N}(m+1)p(m+1), \quad m \in \mathbb{N}_0, \qquad N = \sum_{m \in \mathbb{N}} mp(m), \tag{2.28}$$

where $p(m) = \lim_{n\to\infty} p_n(m)$ with $p_n = \frac{1}{n}\sum_{v \in V} \delta_{\deg(v)}$ the empirical degree distribution (see (2.2), Theorem A.2.2 and (A.3); Condition A.2.1, which is needed for Theorem A.2.2, implies that $p(1) = p(2) = 0$). By Jensen's inequality,

$$\sum_{m \in \mathbb{N}} p^\star(m) \log m \leq \log \left(\sum_{m \in \mathbb{N}} p^\star(m)\, m \right) = \log \nu. \tag{2.29}$$

Consequently, $c_* \geq \rho_{\max}$ (see (2.13), (2.40) and (A.3)), with equality if and only if p is a point mass. Thus, the cut-off threshold c_* exceeds the maximal value of the radius, as shown in Figure 2.3.

4. The coupling of the random walk on the dynamically rewired random graph to the modified random walk is implicit in the proof of the main theorem in [10]. There the main idea was that the path probabilities for the two random walks coincide for self-avoiding paths, and it was shown that the two random walks are with high probability self-avoiding. The crucial observation was that, on a typical configuration drawn according to the configuration model, the random walks are self-avoiding with high probability. The particular form of Condition 2.3.4 was motivated by this observation, and suggests that the same results may hold when the initial graph is drawn according to some other distribution, on which non-backtracking random walks are typically self-avoiding.

5. The graph regularity conditions in Condition 2.3.4 are mild, but can be violated. Consider for example a modification of the local-to-global rewiring in which the probability α_n of the half-edge X_{t-1} being rewired at time t depends on a specific choice of X_{t-1}, e.g. $\alpha_n(X_{t-1}) = 1/\deg_H(X_{t-1})$. This would lead to a violation of Condition 2.3.4(D1). Condition 2.3.4(D2) can be violated by a graph rewiring mechanism that at each time gives preferential treatment to some half-edges. For example, fix a set of half-edges F with $|F| \ll n$, and define a "local-to-F" graph dynamics where the edge that might get rewired at time t with $\{X_{t-1}, C_{t-1}(X_{t-1})\}$ is chosen from the set of edges generated by the configuration C_{t-1} such that each edge contains at least half-edge from F. This obviously results in a violation of the Condition 2.3.4(D2).

6. The scaling regimes considered in Theorem 2.1.5, Corollaries 2.1.7–2.1.9 and Figures 2.1–2.3 are chosen so as to end up with *non-trivial scaling profiles*. Apart from conditions on t in terms of α_n and r_n, there is also the implicit condition that $t = O(\log n)$. However, since the probability that $\tau > t$ is monotone decreasing in both α_n and r_n, we get trivial scaling profiles outside these regimes.

§2.1.5 Previous work

The past decade has witnessed much activity towards understanding processes – both random and deterministic – on dynamic networks [121, 47, 59, 46, 23, 66, 7, 95, 132, 147]. Research is motivated not only by mathematical interest, but also by numerous applications in computer science and data science. One of the emerging efforts is concerned with the study of mixing times of random walks on dynamic networks, and how they compare with those of random walks on static networks. The present chapter fits within this line of research.

The article [9] introduced a version of a *dynamic configuration model* in which a fraction of the edges gets rewired at each step of the random walk according to a global-to-global rewiring mechanism. An important result in [9] was an expression for the mixing time of a *non-backtracking* random walk under conditions that guarantee a locally tree-like structure of the graph and fast dynamics. In a follow-up article [10], the same authors extended their results to moderate and slow dynamics. In particular, they obtained a *trichotomy* for the mixing time of non-backtracking random walks, of the type as stated in Corollary 2.1.9. In the current chapter, however, we achieve this trichotomy under *weaker assumptions*.

Trichotomies were also found in subsequent work. The closest to our setting is [42], where the authors consider a dynamic *directed* version of the configuration model. Contrary to our setting, for the directed graph the rewiring no longer preserves the stationary measure, and the analysis in [42] is restricted to a rewiring mechanism in which all the edges are freshly resampled at each step of the random walk. Two trichotomies are derived for the worst-case total variation distance, respectively, for the joint Markov process given by the graph and the random walk and for the non-Markov process given by the random walk marginal. Trichotomies can also emerge in the presence of other random mechanisms that do not directly change the graph. This is well illustrated in [42, 144, 147], where crossovers were established for random walks

on random graphs with various PageRank-like transitions. The results are analogous to Theorem 2.1.4, with the role of the randomized stopping time τ replaced by the first time the walk gets "teleported" by a PageRank-like transition.

Mixing studies for random walks on dynamic random graphs started with [124], which considered random walks on *dynamic percolation* clusters on a d-dimensional *discrete torus*, i.e., a stochastic version of percolation where edges appear and disappear independently at a given rate. In [124] and subsequent works [123, 77], mixing times were identified for several parameter regimes controlling the rates of the random walk and the random graph dynamics. Similar results were obtained for dynamic percolation on the *complete graph* [138, 133]. One of the main difficulties with the dynamic percolation setting is that the stationary distribution of the random walk changes over time, which explains why results tend to be restricted to specific parameter regimes.

Some further advances were achieved in [133, 13], where general bounds on mixing times, and other quantities such as hitting, cover and return times, were derived for certain classes of evolving graphs under proper *expansion assumptions*. Typically, random walk mixing on a dynamic graph is *faster* than on a static graph, although [13] contains some (artificial) examples where the dynamics makes the mixing slower. Speed-up of mixing times for general Markov chains was recently analysed in [45], which also contains an overview of related results.

Unlike for dynamic graphs, mixing times of random walks on *static* random graphs form a well-established subject. For the present chapter, it is important to note the work in [103, 26, 21], where (two-sided) cut-offs on timescale $\log n$ were established for both simple and non-backtracking random walks on a fairly general class of sparse *undirected* random graphs with *good expansion properties*. More recently, similar results were obtained for static random graphs with *directed* edges [37, 38] or with a *community structure* [20].

§2.1.6 Outline

The remainder of this chapter is organised as follows. In Section 2.2 we define the random walk and the random graph dynamics. In Section 2.3 we prove Theorem 2.1.4. In Section 2.4 we prove Theorem 2.1.5. In Appendix A.1 we show that the joint Markov chain of random walk and dynamically rewired random graph is irreducible, aperiodic and doubly-stochastic. In Appendix A.2 we recall the precise form of Theorem A.2.2. In Appendix A.3 we identify the general form of the transition matrix for rewirings and prove that the stationary distribution for the class of "anything-to-global" rewirings is the uniform distribution on H.

§2.2 Random graph dynamics and random walk

In this section we set up the model. In Section 2.2.1 we give a general description of the rewiring mechanism for the random graph (specific choices will be considered in Section 2.4). In Section 2.2.2 we define the non-backtracking random walk. In Section 2.2.3 we define the joint process of random graph and random walk.

§2.2.1 Random graph dynamics

We consider a general class of graph dynamics in which some edges are randomly rewired at each unit of time according to a prescribed rule. First a subset of edges to be rewired is chosen randomly, then these edges are broken into half-edges, and afterwards the resulting half-edges are paired randomly according to a prescribed distribution. The set of half-edges involved in the rewiring at time $t \in \mathbb{N}$ is denoted by R_t.

Suppose that $X_{t-1} = x$ and $C_{t-1} = \xi$. Then, at time t, the above dynamics gives rise to a distribution $Q_x(\xi, \cdot)$ on $Conf_H$. In [9], [10] a specific choice of dynamics was considered in which $Q_x(\xi, \cdot)$ did not actually depend on x. In such a situation, the configuration component forms a Markov chain itself.

§2.2.2 Random walk

We consider a *non-backtracking random walk* on a dynamic random graph in which some edges are rewired at each step. By non-backtracking we mean that the random walk cannot traverse the same edge *twice in a row*. Since in our model the underlying graph is dynamic and the edges change over time, the random walk is more conveniently defined as a random walk on the set of half-edges H. Recall that at time $t \in \mathbb{N}$ we update the configuration to $C_t = \xi$ and only then let the random walk make a move. Then the random walk moves according to the transition probabilities

$$P_\xi(x, y) := \begin{cases} \frac{1}{\deg_H(y)} & \text{if } \xi(x) \sim y \text{ and } \xi(x) \neq y, \\ 0 & \text{otherwise.} \end{cases} \tag{2.30}$$

More descriptively, when the random walk is on a half-edge x and the graph is in configuration ξ, the random walk moves to one of the siblings of the half-edge that the current half-edge x is paired with, chosen uniformly at random (see Figure 2.4). The transition probabilities are symmetric with respect to the pairing given by ξ, i.e., $P_\xi(x, y) = P_\xi(\xi(y), \xi(x))$. In particular, the transition matrix is doubly stochastic, and so the uniform distribution on H, denoted by U_H, is the stationary distribution for the random walk process:

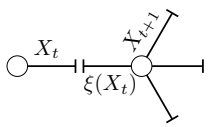


Figure 2.4: The random walk moves from half-edge X_t to half-edge X_{t+1}, one of the siblings of the half-edge $\xi(X_t)$ that X_t is paired to.

§2.2.3 Joint process

The law of the joint Markov chain $(X_t, C_t)_{t \in \mathbb{N}}$, starting from initial half-edge $x \equiv X_0$ and initial configuration $\xi \equiv C_0$, is given by the conditional probabilities

$$\mathbb{P}_{x,\xi}(X_t = z, C_t = \zeta \mid X_{t-1} = y, C_{t-1} = \eta) = Q_y(\eta, \zeta) P_\zeta(y, z), \quad t \in \mathbb{N}, \tag{2.31}$$

with

$$\mathbb{P}_{x,\xi}(X_0 = x, C_0 = \xi) = 1, \tag{2.32}$$

where the transition probabilities $Q_y(\cdot,\cdot)$ remain to be chosen. While the joint process is Markov, the marginal processes $X = (X_t)_{t\in\mathbb{N}}$ and $C = (C_t)_{t\in\mathbb{N}}$ need *not* be Markov. Consequently, the total variation distance $d_{\mathrm{TV}}(\mathbb{P}_{x,\xi}(X_t \in \cdot), U_H(\cdot))$ is not guaranteed to be decreasing in t, even when it converges to 0.

We emphasise that at each time step the graph evolution happens first and only then the random walk makes a move.

Furthermore, note that when the graph dynamics does not depend on the random walk, i.e., $Q_x(\cdot,\cdot) = Q_y(\cdot,\cdot)$ for all $x, y \in H$, the uniform distribution U_H is the stationary distribution for the random walk, i.e., for all $\xi \in \mathit{Conf}_H$ and $t \in \mathbb{N}$,

$$\sum_{x\in H} \frac{1}{|H|} \mathbb{P}_{x,\xi}(X_t \in \cdot) = U_H(\cdot). \tag{2.33}$$

This can easily be seen by noting that the random walk conditioned on a realisation of the graph dynamics is a time-inhomogeneous Markov chain for which U_H is the stationary distribution.

§2.3 Proof of the main theorem

In this section we build up the apparatus that is required to prove Theorem 2.1.4. In Section 2.3.1 we formulate the regularity conditions for the graph and its evolution. In Section 2.3.2 we introduce the modified random walk, which lives on the static random graph. In Section 2.3.3 we propose a coupling of the modified random walk and the dynamically rewired random walk. In Section 2.3.4 we analyse the errors in the coupling. In Section 2.3.5 we use the coupling to prove Theorem 2.1.4.

§2.3.1 Regularity conditions

In the formulation of Theorem 2.1.4 we refer to certain regularity conditions, which we lay out next. The first set of conditions concerns the degrees of the graph:

Condition 2.3.1 (Regularity of degrees).

(R1) $|H| = \Theta(n)$ *as* $n \to \infty$.

(R2) $\max_{v\in V} \deg(v) =: d_{\max} = o(n/(\log n)^2)$ *as* $n \to \infty$.

(R3) $\deg(v) \geq 2$ *for all* $v \in V$.

Condition 2.3.1(R1) ensures that the graph is sparse, and together with Condition 2.3.1(R2) guarantees that the paths of the random walk are with high probability self-avoiding on relevant time scales (see Lemma 2.3.9 below). Condition 2.3.1(R3) is a consistency condition ensuring that the non-backtracking random walk is well-defined.

As stated in the introduction, Condition 2.3.1 is standard in the literature, unlike the forthcoming Condition 2.3.4. To state this new condition, we require further notation.

Definition 2.3.2 (Dynamic self-avoidance).

(1) For $s, t \in \mathbb{N}$ with $s \le t$, define $[s,t] \coloneqq \{s, \dots, t\}$ and $[t] \coloneqq [1,t] = \{1, \dots, t\}$. For $r \in \mathbb{N}$, $t_1, \dots, t_r \in \mathbb{N}$ with $t_1 < \cdots < t_r \le t-1$, introduce a set of times

$$T \coloneqq \{t_1, \dots, t_r\}, \tag{2.34}$$

and sequences of half-edges

$$\begin{aligned} x_{[0,t-1]} &\coloneqq (x_0, \dots, x_{t-1}), \quad \bar{x}_{[0,t-1]} \coloneqq (\bar{x}_0, \dots, \bar{x}_{t-1}), \\ \hat{x}_{[r]} &\coloneqq (\hat{x}_1, \dots, \hat{x}_r), \quad \tilde{x}_{[r]} \coloneqq (\tilde{x}_1, \dots, \tilde{x}_r). \end{aligned} \tag{2.35}$$

The sequences $x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \tilde{x}_{[r]}$ are called *dynamically self-avoiding with respect to* T if the sequences of vertices

$$(v(x_0), \dots, v(x_{t-1})), \quad (v(\bar{x}_{t_1-1}), \dots, v(\bar{x}_{t_r-1})), \quad (v(\tilde{x}_1), \dots, v(\tilde{x}_r)), \tag{2.36}$$

are all distinct.

(2) Recall from Definition 2.1.3 that I_t is the indicator of the event that the random walk steps along a previously rewired edge at time t. Let $\mathsf{DSA}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]})$ be the event that (see Figure 2.5):

- $x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \tilde{x}_{[r]}$ are dynamically self-avoiding with respect to T.
- $I_s = 1$ for $s \in T$ and $I_s = 0$ for $s \in [t-1] \setminus T$.
- $C_0(x_s) = \bar{x}_s$ for $s = 0, \dots, t-1$, where C_0 is the configuration at time $t = 0$.
- $C_{t_i}(x_{t_i-1}) = \hat{x}_i$ for $i = 1, \dots, r$.
- $C_0(\hat{x}_i) = \tilde{x}_i$ for $i = 1, \dots, r$.
- $X_s = x_s$ for $s = 0, \dots, t-1$.

When the event $\mathsf{DSA}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]})$ occurs, we say that the random walk on the dynamic random graph has a *dynamically self-avoiding history up to time* t. We call $x_{[0,t-1]}$ *good* when $\deg_H(x_i) \le (\log n)^{2+\varepsilon}$ for $0 \le i \le t-1$ and some $\epsilon > 0$ fixed, i.e., for all half-edges in the sequence $x_{[0,t-1]}$ the degree is $O((\log n)^{2+\varepsilon})$. A sequence $x_{[0,t-1]}$ that is not good is called bad.

■

Remark 2.3.3 (Interpretation of sequences). On the event $\mathsf{DSA}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]})$, the sequences in Definition 2.3.2 have the following interpretation:

- $T \coloneqq \{t_1, \dots, t_r\} \subset [t-1]$:
  times when the random walk steps along a previously rewired edge,
- $x_{[0,t-1]} \coloneqq (x_0, \dots, x_{t-1})$:
  half-edges the random walk visits up to time $t-1$.

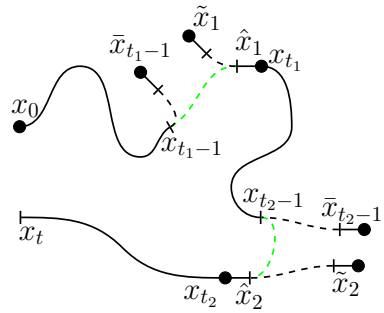


Figure 2.5: Illustration of the event $\mathsf{DSA}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]})$ *and the role of the sequences in Definition 2.3.2. Dashed black lines show the pairing in the initial configuration, while dashed green lines show the pairings that are seen by the random walk.*

- $\bar{x}_{[0,t-1]} \coloneqq (\bar{x}_0, \dots, \bar{x}_{t-1})$:
  half-edges $\bar{x}_i$ that are paired with half-edges $x_i \in x_{[0,t-1]}$ in the initial configuration.
- $\hat{x}_{[r]} \coloneqq (\hat{x}_1, \dots, \hat{x}_r)$:
  half-edges that are paired with x_{i-1} at times $i \in T$.
- $\tilde{x}_{[r]} \coloneqq (\tilde{x}_1, \dots, \tilde{x}_r)$:
  half-edges that are paired with $\hat{x}_1, \dots, \hat{x}_r$ in the initial configuration.

♦

With these definitions in hand, we can now state the conditions on the random graph dynamics:

Condition 2.3.4 (Regularity of graph dynamics). *Recall that I_t is the indicator of the event that a random walk steps over a rewired edge at time t (see Definition 2.1.3). For all $t = O(\log n)$ and all $T = \{t_1, \dots, t_r\} \subset [t-1]$ the following conditions hold (note that I_t is random given $(I_s)_{0<s<t}$):*

(D1) *For all $x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]}$ and $y_{[0,t-1]}, \bar{y}_{[0,t-1]}, \hat{y}_{[r]}, \tilde{y}_{[r]}$ that describe dynamically self-avoiding histories with respect to T,*

$$\begin{aligned}\big|\mathbb{P}(I_t = 1 \mid \mathsf{DSA}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]})) \\ - \mathbb{P}(I_t = 1 \mid \mathsf{DSA}(T, y_{[0,t-1]}, \bar{y}_{[0,t-1]}, \hat{y}_{[r]}, \tilde{y}_{[r]}))\big| = o(\tfrac{1}{\log n}),\end{aligned} \tag{2.37}$$

where the bound is uniform *in the histories.*

(D2) *For all $x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]}$ that describe dynamically self-avoiding histories with respect to T,*

$$\begin{aligned}d_{\mathrm{TV}}(\mathbb{P}(C_t(x_{t-1}) \in \cdot \mid \mathsf{DSA}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]}) \cap \{I_t = 1\}), U_H(\cdot)) \\ = o(\tfrac{1}{\log n}),\end{aligned} \tag{2.38}$$

where the bound is uniform *in the histories.*

In case (D1) and (D2) cannot be verified for all *sets of sequences that describe a dynamically self-avoiding history with respect to T, the following suffices:*

(D3) *If (D1) and (D2) hold for $x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]}$ and $y_{[0,t-1]}, \bar{y}_{[0,t-1]}, \hat{y}_{[r]}, \tilde{y}_{[r]}$ that describe dynamically self-avoiding histories with respect to T for which $x_{[0,t-1]}$ and $y_{[0,t-1]}$ are* good*, but not necessarily hold for those histories for which $x_{[0,t-1]}$ and $y_{[0,t-1]}$ are bad, then the path $X_{[0,t]}$ of the random walk on the dynamic random graph traced until time t satisfies:*

$$\mathbb{P}\left(X_{[0,t-1]} \text{ is bad}\right) = o(\tfrac{1}{\log n}). \tag{2.39}$$

Part (D1) states that the times at which the random walk steps over a rewired edge are almost independent of the fine details of the random walk, provided it has a good dynamically self-avoiding history. Part (D2) states that a random walk with a good dynamically self-avoiding history is close to being mixed right after it steps over a rewired edge. The error terms of order $o(1/\log n)$ are chosen such that we can carry out the estimates in Lemma 2.3.9 below. Part (D3) ensures that good dynamically self-avoiding histories are typical.

To identify the scaling of the mixing time for near-to-global rewiring in Corollary 2.1.8, we need an extra regularity condition:

Condition 2.3.5 (Regularity of degree distribution). *Let $p_n := \frac{1}{n}\sum_{v\in V}\delta_{\deg(v)}$ denote the empirical degree distribution. We require that:*

(R1*) $\lim_{n\to\infty} p_n = p$, *pointwise for some probability distribution p on $\mathbb{N}$.*

(R2*) $\lim_{n\to\infty}\sum_{m\in\mathbb{N}} m p_n(m) = \sum_{m\in\mathbb{N}} m p(m) < \infty$.

(R3*) $\lim_{n\to\infty}\sum_{m\in\mathbb{N}} m^2 p_n(m) = \sum_{m\in\mathbb{N}} m^2 p(m) < \infty$.

The size-biased mean minus one of p is

$$\nu = \frac{\sum_{m\in\mathbb{N}} m(m-1)p(m)}{\sum_{m\in\mathbb{N}} m p(m)} \tag{2.40}$$

and is assumed to satisfy $\nu > 1$.

We may interpret ν as the *average forward degree* of a uniformly chosen half-edge, which plays the role of the mean offspring in the branching-process approximation of the local limit of the configuration model. In view of Condition 2.3.1(R3), the condition $\nu > 1$ amounts to the requirement $p \neq \delta_2$.

§2.3.2 Modified random walk

We define a *modified random walk*, denoted by $(Y_t)_{t\in\mathbb{N}}$, as a random walk on a static random graph that at certain random times makes uniform jumps. Formally, we have a sequence $(J_t)_{t\in\mathbb{N}}$ of random variables adapted to a filtration $(\mathcal{F}_t)_{t\in\mathbb{N}}$, taking values in $\{0,1\}$ according to a pre-specified distribution on $\{0,1\}^{\mathbb{N}}$. For fixed $t \in \mathbb{N}$, J_t is seen as the indicator of the event that the modified random walk makes a uniform jump at

time t. The law of the modified random walk $(Y_t)_{t\in\mathbb{N}}$ on ξ that starts from the initial half-edge $x \equiv Y_0$, which is adapted to $(\mathcal{F}_t)_{t\in\mathbb{N}}$, is given by the conditional probabilities

$$\begin{aligned}&\mathbb{P}^{\mathrm{mod}}_{x,\xi}(Y_t = z \mid Y_{t-1} = y, J_1 = j_1, \dots, J_t = j_t)\\ &= \mathbb{P}^{\mathrm{mod}}_{x,\xi}(Y_t = z \mid Y_{t-1} = y, J_t = j_t) = \begin{cases} P_\xi(y,z) & \text{if } j_t = 0,\\ \frac{1}{|H|} & \text{if } j_t = 1, \end{cases} \quad t \in \mathbb{N},\end{aligned} \tag{2.41}$$

with

$$\mathbb{P}^{\mathrm{mod}}_{x,\xi}(Y_0 = x) = 1. \tag{2.42}$$

Note that, according to the definition, neither $(J_t)_{t\in\mathbb{N}}$ nor the pair $(Y_t, J_t)_{t\in\mathbb{N}}$ needs to be Markov, but $(Y_t)_{t\in\mathbb{N}}$ is Markov conditionally on a realisation of $(J_t)_{t\in\mathbb{N}}$.

Uniform jumps of the modified random walk can be rephrased in the following form. Let Y'_t be a uniformly chosen half-edge, independent of the random walk path and the jump times. If $J_t = 1$, then we choose a uniform sibling of Y'_t, say y, and set $Y_t = y$. Since Y'_t is uniform and one of its siblings is chosen uniformly at random, the resulting half-edge is distributed uniformly on H. Even though Y'_t is already a half-edge chosen uniformly at random, working with its sibling (which is also a half-edge chosen uniformly at random) will come in handy in the coupling argument in Section 2.3.3.

As an analogue of τ, we define σ to be the first time that the modified random walk makes a uniform jump, i.e.,

$$\sigma \coloneqq \inf\{t \in \mathbb{N} : J_t = 1\}. \tag{2.43}$$

§2.3.3 Coupling of modified and dynamically rewired random walk

We couple the law $\mathbb{P}_{x,\xi}(X_t \in \cdot)$ of the random walk on the dynamic random graph, with initial half-edge x and initial configuration ξ, to the law $\mathbb{P}^{\mathrm{mod}}_{x,\xi}(Y_t \in \cdot)$ of the modified random walk. We want the coupled random walks to *stick together as much as possible*. When the two random walks make different steps, we say that the coupling of the two random walks has *failed*. Until the coupling fails, the times at which the random walk on the dynamically rewired graph makes a step over a previously rewired edge correspond to the times at which the modified random walk makes a uniform jump.

Definition 2.3.6 (Coupling to a modified random walk). Let X_t be a non-backtracking random walk starting in the initial state (x, ξ), where $x \in H$, $\xi \in \mathit{Conf}_H$, and Y_t be a modified random walk on ξ starting in x. First, define a sequence of auxiliary random sets $(A_t)_{t\in\mathbb{N}_0}$. Call A_t the set of *active* half-edges at time t. Let A_0 be the set consisting of the initial half-edge of the random walk and its siblings, i.e., $A_0 \coloneqq H_{v(x)}$.

Define the coupling of the non-backtracking random walk X_t and the modified random walk Y_t at any time $t \in \mathbb{N}$ by the following rules:

(a) The coupling is successful, unless it has been declared failed.

(b) If $\xi(X_{t-1})$ or any of its siblings belong to A_{t-1}, then declare the coupling as **failed**.

(c) If $\deg_H(X_{t-1}) > (\log n)^{2+\varepsilon}$ (recall that $n \coloneqq |V|$), then declare the coupling as **failed**. If Condition 2.3.4(D3) is not needed, then this rule is suspended (see Remark 2.3.7 below for further details).

(d) If the coupling has not yet failed, then maximally couple the distribution of I_t, conditionally on the history of the random walk and the rewired edges seen by the random walk, to the distribution of J_t, conditionally on the values of the indicators $J_1, \dots, J_{t-1}$. The following three outcomes are possible:

  (a) If the coupling of the conditional distributions of I_t and J_t is successful and $I_t = J_t = 0$, then create A_t as a union of $\xi(X_{t-1})$ and all its siblings with A_{t-1}. Let the random walk on a dynamic graph make a move and set $Y_t \coloneqq X_t$.

  (b) If the coupling of the conditional distributions of I_t and J_t is successful and $I_t = J_t = 1$, then maximally couple the distribution of $C_t(X_{t-1})$, i.e., the half-edge paired with X_{t-1} in configuration C_t, conditionally on the history of the random walk and $I_t = 1$, to the distribution of Y_t':

    (a) If the coupling of $C_t(X_{t-1})$ and Y_t' is successful, and neither $C_t(X_{t-1})$ nor any of its siblings is already contained in A_{t-1}, then add $\xi(X_{t-1})$ and all its siblings, along with $C_t(X_{t-1})$ and all its siblings, to A_{t-1} in order to obtain A_t. Phrased in symbols:

$$\begin{aligned} A_t \coloneqq A_{t-1} &\cup \xi(X_{t-1}) \cup \{h \in H\colon\ h \sim \xi(X_{t-1})\} \\ &\cup C_t(X_{t-1}) \cup \{h \in H\colon\ h \sim C_t(X_{t-1})\}. \end{aligned} \tag{2.44}$$

    Let the random walk on the dynamic graph make a move, and set $Y_t \coloneqq X_t$.

    (b) Otherwise, declare the coupling as **failed**.

  (c) If the coupling of the conditional distributions of I_t and J_t is not successful, namely if $I_t \neq J_t$, then declare the coupling of the two random walks as **failed**.

(e) If the coupling has failed let X_t and Y_t evolve independently.

■

Remark 2.3.7 (Failure of the coupling after a high-degree half-edge is encountered). In Lemma 2.3.9 we will see that failure of the coupling as described in rule (c) of Definition 2.3.6 is needed only when Condition 2.3.4(D3) comes into play. This will only happen for one of the three examples in Section 2.4, namely, near-to-global. ♦

§2.3.4 Failures in the coupling

Remark 2.3.8 (Possible failures). At each time $t \in \mathbb{N}$, the random walk and the coupled modified random walk try to avoid stepping on the active half-edges A_{t-1}. The coupling of these two random walks fails in four cases described in Definition 2.3.6:

(a) In step (b):

(a) if the coupling of $C_t(X_{t-1})$ and Y'_t is not successful,
(b) if the two random walks step over a half-edge in A_{t-1}.

(b) In step (c), if the coupling of I_t and J_t is not successful.

(c) In step (b), if the pair of X_{t-1} in the starting configuration is already in A_{t-1}.

(d) In step (c), if the random walk encounters a half-edge X_{t-1} with a high degree.

Failure cases (b) and (c) correspond to the situation in which the random walks do not have dynamically self-avoiding histories. Consequently, *the random walks have dynamically self-avoiding histories before the coupling of the two random walks fails.* Failure case (a) corresponds to the situation in which the conditional distribution of $C_t(X_{t-1})$ is too far from the uniform distribution in total variation distance. Failure case (b) corresponds to the situation in which the conditional distribution of the times at which the random walk on the dynamically rewired graph and the conditional distribution of the times at which the modified random walk makes uniform jumps are far from each other in total variation distance. Finally, failure case (d) corresponds to the situation when during the graph exploration the random walk encounters a half-edge with an anomalously high degree. ♦

The next lemma states that these failure events are unlikely up to logarithmic times when Conditions 2.3.1 and 2.3.4 hold for the random walk on the dynamically rewired random graph:

Lemma 2.3.9 (Coupling estimates). *Suppose that $t = O(\log n)$, and that Conditions 2.3.1 and 2.3.4 hold for the random walk on the dynamically rewired graph. For all $1 \leq s \leq t$ and all $T_s = \{s_1, \ldots, s_r\} \subset [s-1]$, fix a sequence of half-edges*

$$x^{T_s}_{[0,s-1]}, \bar{x}^{T_s}_{[0,s-1]}, \hat{x}^{T_s}_{[r]}, \tilde{x}^{T_s}_{[r]} \tag{2.45}$$

that describes a good dynamically self-avoiding history with respect to T_s (see Definition 2.3.2). Consider the modified random walk for which the jump distribution has conditional distribution

$$\begin{aligned}\mathbb{P}^{\mathrm{mod}}_{x,\xi}\big(J_s = 1 \mid J_{s'} = 0 \text{ for } s' \in [s-1] \setminus T_s, J_{s''} = 1 \text{ for } s'' \in T_s\big)\\ := \mathbb{P}_{x,\xi}\Big(I_s = 1 \mid \mathsf{DSA}\big(T_s, x^{T_s}_{[0,s-1]}, \bar{x}^{T_s}_{[0,s-1]}, \hat{x}^{T_s}_{[r]}, \tilde{x}^{T_s}_{[r]}\big)\Big).\end{aligned} \tag{2.46}$$

Then, whp in x and ξ,

$$d_{\mathrm{TV}}\left(\mathbb{P}_{x,\xi}(X_t \in \cdot), \mathbb{P}^{\mathrm{mod}}_{x,\xi}(Y_t \in \cdot)\right) = o_{\mathbb{P}}(1), \tag{2.47}$$

and, with σ as defined in (2.43),

$$\mathbb{P}_{x,\xi}(\tau > t) = \mathbb{P}^{\mathrm{mod}}_{x,\xi}(\sigma > t) + o_{\mathbb{P}}(1). \tag{2.48}$$

Remark 2.3.10 (Jump distribution of the modified random walk). Observe that (2.46) describes the jump distribution of the modified random walk at any time $1 \leq s \leq t$ for any set of previous jump times T_s in a *non-anticipating* manner. If the sequence of half-edges in the event $\mathsf{DSA}(T_s, x^{T_s}_{[0,s-1]}, \bar{x}^{T_s}_{[0,s-1]}, \hat{x}^{T_s}_{[r]}, \tilde{x}^{T_s}_{[r]})$ in the right-hand

side of (2.46) is not compatible with the initial state (x, ξ) (i.e., when the conditioning is on an event of probability zero), then we set the right-hand side of (2.46) equal to zero. The proof below uses an annealing argument in which the "mismatched" events play no role. ♦

Proof of Lemma 2.3.9. Let $\mathbb{P}^{\text{couple}}_{x,\xi}$ denote the law of the coupling of the two non-backtracking random walks described in Section 2.3.2 with $X_0 = x$ and $C_0 = \xi$. Also, use $F \in \mathbb{N}$ to denote the time at which this coupling fails. Due to Condition 2.3.1(R3), these random walks are always well-defined. Since the two random walks agree up to the time F, that is until the coupling fails, we have

$$d_{\text{TV}}\left(\mathbb{P}_{x,\xi}(X_t \in \cdot), \mathbb{P}^{\text{mod}}_{x,\xi}(Y_t \in \cdot)\right) \leq \mathbb{P}^{\text{couple}}_{x,\xi}(F \leq t). \tag{2.49}$$

So, in order to prove the claim it suffices to show that, whp in x and ξ,

$$\mathbb{P}^{\text{couple}}_{x,\xi}(F \leq t) = o_{\mathbb{P}}(1). \tag{2.50}$$

To achieve this, we use an annealing argument on the initial graph and the initial location. Recall that $\mu = U_H \times U_{Conf_H}$, and let

$$\mathbb{P}^{\text{couple}} = \sum_{x,\xi} \mu(x, \xi)\, \mathbb{P}^{\text{couple}}_{x,\xi}. \tag{2.51}$$

We will show that

$$\mathbb{P}^{\text{couple}}(F \leq t) = o(1) \tag{2.52}$$

by exploring the initial configuration using the paths of the random walk and its coupled modified random walk until the coupling fails at time F.

(a) At time $s = 0$, choose a half-edge $x \in H$ uniformly at random. Set $X_0 = Y_0 = x$ and $A_0 = H_{v(x)}$, the subset of H consisting of x and its siblings.

(b) At time $s \in \mathbb{N}$, first explore the half-edge to which $X_{s-1} = Y_{s-1}$ is paired in the initial configuration ξ, then let the coupled random walks evolve in accordance with Definition 2.3.6, and update A_s accordingly.

This exploration process covers the part of the graph seen by the random walks, along with the parts affected by the rewiring at the positions of the random walks, and stops as soon as the coupling of the two random walks fails.

We will carry out the proof in a setting where Conditions 2.3.4(D1) and (D2) hold. At the end of the proof we will briefly comment on the changes required when Condition 2.3.4(D3) comes into play.

Suppose that the coupling of the two random walks has not failed before time s. Failure at time s can occur in the following three cases (see also Remark 2.3.8):

(a) The coupling of I_s and J_s fails in step (c) of Definition 2.3.6.

(b) The coupling of $C_s(X_{s-1})$ and Y'_s fails in step (b) of Definition 2.3.6.

(c) The random walks jointly step over a half-edge that lies in A_{s-1} in either step (b) or step (b) of Definition 2.3.6.

For case (a), we note that, since the distribution of J_t for the modified random walk is given by (2.46), Condition 2.3.4(D1) implies that the probability of coupling failure is $o(1/\log n)$.

For case (b) we note that, by Remark 2.3.8, before the coupling of the two random walks fails, the random walk has a dynamically self-avoiding history. By Condition 2.3.4(D2), the total variation distance between the conditional distribution of $C_s(X_{s-1})$ and the uniform distribution U_H is $o(1/\log n)$. Since Y'_s is also distributed uniformly on H, the probability of the event in case 2 is $o(1/\log n)$.

For case (c), we first need an upper bound on the size of A_{s-1}. Each time we explore the initial configuration, we add at most $d_{\max}$ half-edges to the set of active half-edges. In case a rewiring occurs, then we add at most $2d_{\max}$ half-edges to the set of active half-edges. This gives us the following crude bound:

$$|A_{s-1}| \leq 3sd_{\max}. \tag{2.53}$$

For a fail event in step (b), we see that the probability that $C_s(X_{s-1}) \in A_{s-1}$ is smaller than

$$\frac{|A_{s-1}|}{|H|} + o(1/\log n) \leq \frac{3sd_{\max}}{|H|} + o(1/\log n), \tag{2.54}$$

since the random walk has a dynamically self-avoiding history before the coupling of the two random walks fails (see Remark 2.3.8), so the total variation distance between the conditional distribution of $C_s(X_{s-1})$ and the uniform distribution U_H is $o(1/\log n)$, by Condition 2.3.4(D2).

For a fail event in step (b), we see that the probability that $C_0(X_{s-1}) \in A_{s-1}$ is smaller than

$$\frac{|A_{s-1}|}{|H| - 4s + 4} \leq \frac{3sd_{\max}}{|H| - 4s + 4}, \tag{2.55}$$

since up to time s we form at most $2s-2$ pairs in C_0, of which $s-1$ on the random walk path and an additional $s-1$ if rewiring occurs at each step up to time s.

The above estimates give us

$$\mathbb{P}^{\text{couple}}(F = s \mid F > s-1) \leq \frac{6sd_{\max}}{|H| - 4s + 4} + o\Big(\frac{1}{\log n}\Big). \tag{2.56}$$

Taking a union bound up to time t, and using that by assumption $t = O(\log n)$, $d_{\max} = o(n/(\log n)^2)$ (Condition 2.3.1(R2)) and $|H| = \Theta(n)$ (Condition 2.3.1(R1)), we get

$$\mathbb{P}^{\text{couple}}(F \leq t) \leq \frac{3t(t+1)d_{\max}}{|H| - 4t} + o(1) = o(1). \tag{2.57}$$

Since this bound holds for the annealed quantity $\mathbb{P}^{\text{couple}}(F \leq t)$, using Markov inequality we get for the quenched probabilities the desired claim that

$$\mathbb{P}^{\text{couple}}_{x,\xi}(F \leq t) = o_{\mathbb{P}}(1). \tag{2.58}$$

In case we rely on Condition 2.3.4(D3), a fourth possible failure of the coupling shows up, namely, if the random walk encounters a half-edge of degree larger than $(\log n)^{2+\varepsilon}$. The probability of this failure is $o(1/\log n)$ by Condition 2.3.4(D3). The

estimates for the other possible failures carry over, because if the coupling did not fail at some time s due to a meeting with a high-degree half-edge, then the random walk path traced up to time s is good and we can apply the same arguments as above.

Finally, note that events $\{\tau > t\}$ and $\{\sigma > t\}$ can be expressed as finite unions of events of the type $\{X_t \in \cdot\}$ or $\{Y_t \in \cdot\}$, respectively. Since (2.47) bounds the total variation distance between $\mathbb{P}_{x,\xi}(X_t \in \cdot)$ and $\mathbb{P}^{\mathrm{mod}}_{x,\xi}(Y_t \in \cdot)$, the claim in (2.48) follows. □

§2.3.5 Link between dynamic and static

In this section we prove Theorem 2.1.4. Recall from Section 2.3.2 that for any fixed $T = \{t_1, \dots, t_r\} \subset [t]$, the modified random walk conditionally on the event $J(T) \coloneqq \{J_s = 0 \text{ for } s \in [t] \setminus T, J_s = 1 \text{ for } s \in T\}$ is a time-inhomogeneous Markov chain that makes random-walk steps at times $s \in [t] \setminus T$ and jumps to half-edges chosen uniformly at random at times $s \in T$.

We start by two observations about $\mathbb{P}^{\mathrm{mod}}_{x,\xi}$. Conditionally on $T \subset [t]$ being non-empty, we know that $\sigma < t$ by definition (recall eq. (2.43)), and so the random walk at time t on a graph satisfying Condition 2.3.1 is well-mixed for any starting $x \in H, \xi \in \mathit{Conf}_H$. Expressed in symbols, this means that

$$\mathbb{P}^{\mathrm{mod}}_{x,\xi}(Y_t \in \cdot \mid \sigma \le t) = U_H(\cdot). \tag{2.59}$$

On the other hand, since the modified random walk up to time t conditionally on the event $\{\sigma > t\}$ is the same as the random walk on the static graph, for any $x \in H$ and $\xi \in \mathit{Conf}_H$, we have

$$d_{\mathrm{TV}}\left(\mathbb{P}^{\mathrm{mod}}_{x,\xi}(Y_t \in \cdot \mid \sigma > t), U_H(\cdot)\right) = \mathcal{D}^{\mathrm{stat}}_{x,\xi}(t). \tag{2.60}$$

Now we just combine our previous results. Using the triangle inequality twice, we obtain

$$\begin{aligned} d_{\mathrm{TV}}\left(\mathbb{P}^{\mathrm{mod}}_{x,\xi}(Y_t \in \cdot), U_H(\cdot)\right) \le{}& \mathbb{P}^{\mathrm{mod}}_{x,\xi}(\sigma > t)\; d_{\mathrm{TV}}\left(\mathbb{P}^{\mathrm{mod}}_{x,\xi}(Y_t \in \cdot \mid \sigma > t), U_H(\cdot)\right) \\ &+ \mathbb{P}^{\mathrm{mod}}_{x,\xi}(\sigma \le t)\; d_{\mathrm{TV}}\left(\mathbb{P}^{\mathrm{mod}}_{x,\xi}(Y_t \in \cdot \mid \sigma \le t), U_H(\cdot)\right), \end{aligned} \tag{2.61}$$

and

$$\begin{aligned} d_{\mathrm{TV}}\left(\mathbb{P}^{\mathrm{mod}}_{x,\xi}(Y_t \in \cdot), U_H(\cdot)\right) \ge{}& \mathbb{P}^{\mathrm{mod}}_{x,\xi}(\sigma > t)\; d_{\mathrm{TV}}\left(\mathbb{P}^{\mathrm{mod}}_{x,\xi}(Y_t \in \cdot \mid \sigma > t), U_H(\cdot)\right) \\ &- \mathbb{P}^{\mathrm{mod}}_{x,\xi}(\sigma \le t)\; d_{\mathrm{TV}}\left(\mathbb{P}^{\mathrm{mod}}_{x,\xi}(Y_t \in \cdot \mid \sigma \le t), U_H(\cdot)\right). \end{aligned} \tag{2.62}$$

Inserting (2.59) and (2.60) into the previous two inequalities, we get the equality

$$d_{\mathrm{TV}}\left(\mathbb{P}^{\mathrm{mod}}_{x,\xi}(Y_t \in \cdot), U_H(\cdot)\right) = \mathbb{P}^{\mathrm{mod}}_{x,\xi}(\sigma > t)\, \mathcal{D}^{\mathrm{stat}}_{x,\xi}(t). \tag{2.63}$$

Using Lemma 2.3.9, we see that, whp in x and ξ,

$$\mathcal{D}^{\mathrm{dyn}}_{x,\xi}(t) = \mathbb{P}_{x,\xi}(\tau > t)\, \mathcal{D}^{\mathrm{stat}}_{x,\xi}(t) + o_{\mathbb{P}}(1), \tag{2.64}$$

which concludes the proof of Theorem 2.1.4. □

§2.4 Applications of the general framework

In Section 2.4.1 we introduce three choices of rewiring. In Sections 2.4.2–2.4.4 we identify, for each of these choices, the scaling of the probability that the random walk does not step along a previously rewired edge, which settles Theorem 2.1.5.

In Appendix A.1 we show that each of the three choices of rewiring leads to an irreducible and aperiodic joint Markov chain for the random walk and the random graph.

§2.4.1 Three choices of rewiring

We explore rewirings that fit into a larger scheme of random graph dynamics, namely, where the decision of which edges to rewire depends on their distance to the current position of the random walk.

Definition 2.4.1 (Sets of edges to be rewired). Recall that the configuration ξ is a pairing of all the half-edges (which induces a set of edges) and H is the set of all half-edges. By abuse of notation, in Section 2.1.1 we introduced the expression $\{a, b\} \in \xi$, $a, b \in H$, to mean that the half-edges a, b form an edge in the configuration ξ. For any $\xi \in \mathit{Conf}_H$, $h \in H$ and $r_n \in \mathbb{N}$, define the following sets of edges:

$$\begin{aligned} \mathrm{Local}_\xi(h) &:= \{\{h, g\} \in \xi\}, \\ \mathrm{Near}_{\xi, r_n}(h) &:= \left\{ \{k, l\} \in \xi \colon \begin{array}{ll} k \in H \colon \mathbb{P}(X_{t+\rho} = k \mid X_{t-1} = h, \, \xi \text{ fixed}) > 0 & \\ & \text{for } 0 \le \rho < r_n, \\ l \in H \colon \, l = \xi(k) & \end{array} \right\}. \end{aligned} \tag{2.65}$$

■

In words, $\mathrm{Local}_\xi(h)$ is the edge to which the half-edge h belongs and $\mathrm{Near}_{\xi, r_n}(h)$ are the edges that can be reached after $r_n - 1$ steps by the non-backtracking random walk when the graph is in configuration ξ (and is not evolving). Obviously, $\mathrm{Near}_{\xi,1}(h)$ is equivalent to $\mathrm{Local}_\xi(h)$.

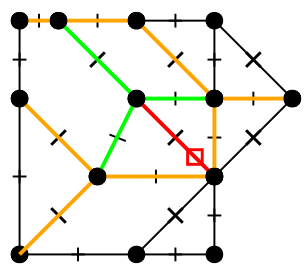


Figure 2.6: Illustration of the sets in Definition 2.4.1. The red box denotes the current position of the random walk. The red edge forms the local set, which is also the near set with $r_n = 1$. The red and green edges form the near set with $r_n = 2$. The red, green and orange edges form the near set with $r_n = 3$.

With the above notation we can define the dynamics:

Definition 2.4.2 (Random walk with (K_t)-to-(L_t) rewiring). Recall that X_{t-1} is the position of the random walk before the transition at time t and C_{t-1} is the configuration of the random graph before an update at time t. Let $(K_t)_{t\in\mathbb{N}}, (L_t)_{t\in\mathbb{N}}$ be sequences of sets of edges, which can be different at each time t. Define the random walk with (K_t)-to-(L_t) rewiring as the following process:

(a) At each time $t \in \mathbb{N}$, for each edge $e \in K_t$ draw a Bernoulli random variable Z_t^e with parameter α_n, independently of everything else.

(b) (a) If $Z_t^e = 1$, then select edge e for rewiring.
(b) If $Z_t^e = 0$, then edge e will not be rewired.

Write R_t to denote the set of edges that get rewired at time t.

(c) (a) If $|R_t| \geq |L_t \setminus R_t|$, then break-up all the edges in $R_t \cup L_t$ into half-edges and re-pair them at random. More formally, pick $\frac{1}{2}|R_t \cup L_t|$ different half-edges (the half-edges forming $R_t \cup L_t$) and order them randomly. Also order randomly the half-edges not chosen in the previous step. The new pairing is generated by pairing the successive elements from the first and the second ordered sets described above.

(b) Otherwise, for every $e \in R_t$, choose $e' \in L_t \setminus R_t$ uniformly at random without replacement. Denote the set of all edges e' chosen in the previous step by R'_t. Break up R_t into half-edges and order them randomly. Do the same with R'_t. Just as in (a), the new pairing is given by the successive elements of the first and the second ordered set.

The new pairing of half-edges obtained in either (a) or (b) above is the new graph configuration C_t.

(d) The random walk moves from X_{t-1} to X_t on the evolved graph C_t.

■

Remark 2.4.3 (Sets of edges generated from a configuration). When in the sequel we write $L_t \equiv \xi \in \mathit{Conf}_H$, we mean that the set of edges L_t is generated by the configuration ξ, which is a pairing of the entire set of half-edges H. ♦

§2.4.2 Local-to-global rewiring

In this section we focus on a rewiring mechanism that is called local-to-global. Using the language of Definition 2.4.2, this would be a rewiring with $K_t = \mathrm{Local}_{C_{t-1}}(X_{t-1})$ (see Definition 2.4.1) and $L_t \equiv C_{t-1}$ (see Remark 2.4.3). Observe that the set K_t is explicitly dependent on the position of the random walk X_{t-1} before the transition at time t occurs. For $\xi, \eta \in \mathit{Conf}_H$ and $x \in H$, define

$$Q_x^R(\xi,\eta) := \begin{cases} \frac{1}{|H|-2} & \text{if } \xi(\eta(x)) = \eta(\xi(x)) \text{ and } |\xi \setminus \eta| = 2, \\ 0 & \text{otherwise.} \end{cases} \tag{2.66}$$

Then the transition matrix for the random graph from configuration ξ to configuration η when the random walk is at position x equals

$$Q_x(\xi,\eta) = (1-\alpha_n)I(\xi,\eta) + \alpha_n Q_x^R(\xi,\eta), \tag{2.67}$$

where $I(\xi,\eta) = 1$ if $\eta = \xi$, and $I(\xi,\eta) = 0$ otherwise, i.e., I is the identity matrix. The first term of (2.67) captures the situation when rewiring does not happen and the graph remains the same. On the other hand, the off-diagonal symmetric matrix $Q_x^R(\xi,\eta)$ in the second term represents the possible evolution of the graph by local-to-global rewiring. Note that the only possible transitions between graph states are those where the two configurations ξ and η differ in exactly two pairs of half-edges. The condition $\xi(\eta(x)) = \eta(\xi(x))$ in (2.66) says that rewiring always happens at the position of the random walk. The value $\frac{1}{|H|-2}$ comes from the fact that at time t the rewiring mechanism can choose to pair the half-edge X_t to any half-edge chosen randomly from $H \setminus \{X_{t-1}, C_{t-1}(X_{t-1})\}$, which is a set of size $|H| - 2$.

Since Q_x^R is symmetric for all $x \in H$, we see that the measure U_{Conf_H}, defined by

$$U_{Conf_H}(\zeta) := \frac{1}{|Conf_H|} \qquad \forall \zeta \in Conf_H \tag{2.68}$$

is the stationary distribution for Q_x^R for any $x \in H$. This implies that U_{Conf_H} is also the stationary distribution for Q_x for all $x \in H$.

Remark 2.4.4 (Symmetry of transition matrix for graph dynamics). Local-to-global rewiring is one of the examples where the transition matrix is symmetric. Symmetry does not hold generally, even within the restricted class of "something-to-global" rewirings. Still, for such rewirings the transition matrices are always doubly stochastic. For more details see Appendix A.3. ♦

Using this fact, we have the following result for the joint Markov chain:

Proposition 2.4.5 (Stationary distribution). *For any $\alpha_n \in [0,1]$, $U_H \times U_{Conf_H}$ is the stationary distribution for the random walk with local-to-global rewiring with parameter α_n.*

Proof. Recall from Section 2.2.2 that P_η is the transition matrix for the non-backtracking random walk on the graph η. Since U_H is stationary for P_η for any $\eta \in Conf_H$, and U_{Conf_H} is stationary for Q_x for any $x \in H$, it follows that for any $y \in H$ and $\eta \in Conf_H$,

$$\begin{aligned}
&\sum_{x\in H}\sum_{\xi\in Conf_H} U_H(x)U_{Conf_H}(\xi)\,\mathbb{P}_{x,\xi}(X_1 = y, C_1 = \eta) \\
&\quad= \sum_{x\in H}\sum_{\xi\in Conf_H} U_H(x)U_{Conf_H}(\xi)\,Q_x(\xi,\eta)P_\eta(x,y) \\
&\quad= \sum_{x\in H} U_H(x)\,P_\eta(x,y) \sum_{\xi\in Conf_H} U_{Conf_H}(\xi)\,Q_x(\xi,\eta) \\
&\quad= U_{Conf_H}(\eta)\sum_{x\in H} U_H(x)P_\eta(x,y) = U_{Conf_H}(\eta)U_H(y).
\end{aligned} \tag{2.69}$$

□

It is not obvious that the joint Markov chain is irreducible and aperiodic. In Appendix A.1 we show that this is nonetheless the case when $\alpha_n \in (0,1)$, and so the distribution of the joint Markov chain at time t converges to $U_H \times U_{Conf_H}$ as $t \to \infty$.

An important implication is that the distribution of the random walk alone at time t converges to U_H as $t \to \infty$. Indeed, for any $x \in H$, $\xi \in \mathit{Conf}_H$ and $t \in \mathbb{N}$, we have

$$\mathcal{D}^{\mathrm{dyn}}_{x,\xi}(t) \le d_{\mathrm{TV}}\left(\mathbb{P}_{x,\xi}((X_t, C_t) \in \cdot), U_H \times U_{\mathit{Conf}_H}(\cdot)\right), \tag{2.70}$$

and since the right-hand side tends to 0 as $t \to \infty$, $\mathcal{D}^{\mathrm{dyn}}_{x,\xi}(t)$ also tends to 0 as $t \to \infty$. On the other hand, this argument does not automatically imply that $\mathcal{D}^{\mathrm{dyn}}_{x,\xi}(t)$ is non-increasing in t.

We are now ready to prove the scaling results stated in Theorem 2.1.5(A) and Corollary 2.1.7:

Proof of results in Theorem 2.1.5(A) and Corollary 2.1.7: For fixed $t = O(\log n)$, fix some $T = \{t_1, \dots, t_r\} \subset [t-1]$ and some $x_{[0,t-1]}$, $\bar{x}_{[0,t-1]}$, $\hat{x}_{[r]}$ and $\tilde{x}_{[r]}$ that describe a dynamically self-avoiding history with respect to T. Conditionally on the event $\mathsf{DSA}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]})$, x_{t-1} cannot have been rewired before time t. Indeed, by construction the half-edges that are rewired before time t are $x_{t_1-1}, \dots, x_{t_r-1}$, $\bar{x}_{t_1-1}, \dots, \bar{x}_{t_r-1}$, $\hat{x}_1, \dots, \hat{x}_r$ and $\tilde{x}_1, \dots, \tilde{x}_r$, while x_{t-1} is not equal to any of these. So we have

$$\begin{aligned}&\mathbb{P}\big(I_t = 1 \mid \mathsf{DSA}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]})\big)\\ &= \mathbb{P}\big(Z_t^{\mathrm{Local}_{C_{t-1}}(X_{t-1})} = 1 \mid \mathsf{DSA}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]})\big) = \alpha_n.\end{aligned} \tag{2.71}$$

Since this holds for any choice of $x_{[0,t-1]}$, $\bar{x}_{[0,t-1]}$, $\hat{x}_{[r]}$ and $\tilde{x}_{[r]}$, Condition 2.3.4(D1) holds with zero error. As a consequence, Condition 2.3.4(D1) is trivially satisfied. Moreover,

$$\mathbb{P}(C_t(x_{t-1}) \in \cdot \mid \mathsf{DSA}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]}) \cap \{I_t = 1\}) = U_{H \setminus \{x_{t-1}, C_{t-1}(x_{t-1})\}}, \tag{2.72}$$

because after rewiring, the half-edge x_{t-1} cannot end up being paired with itself or the half-edge it was paired with before. This gives

$$d_{\mathrm{TV}}\left(\mathbb{P}\big(C_t(x_{t-1}) \in \cdot \mid \mathsf{DSA}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]}) \cap \{I_t = 1\}\big), U_H(\cdot)\right) = \frac{2}{|H|}. \tag{2.73}$$

Since this holds for any choice of $x_{[0,t-1]}$, $\bar{x}_{[0,t-1]}$, $\hat{x}_{[r]}$ and $\tilde{x}_{[r]}$, Condition 2.3.4(D2) holds with error $O(1/n)$.

On the other hand, the event $\{\tau = t\}$ is the same as the event $\{\min\{s \in \mathbb{N} : R_s = 1\} = t\}$, since when a rewiring occurs the random walk steps over a rewired edge with probability 1. This implies that, for any x and ξ,

$$\mathbb{P}_{x,\xi}(\tau > t \mid \mathsf{SA}(t)) = (1 - \alpha_n)^t = \mathrm{e}^{-[1+o(1)]\,\alpha_n t}, \tag{2.74}$$

where $\mathsf{SA}(t)$ is the event that the random walk is self-avoiding until time t. The first equality comes from the requirement that none of the edges the random walk steps over until time t gets rewired, the second equality uses that $\lim_{n\to\infty} \alpha_n = 0$. Since

$$\lim_{n\to\infty} \mathbb{P}_{x,\xi}(\mathsf{SA}(t)) = 1 \quad \text{whp uniformly in } t = O(\log n), \tag{2.75}$$

we obtain the scaling in Theorem 2.1.5(A). (The proof of (2.75) was given in [10, Lemma 3.1] for global-to-global rewiring, but easily carries over to local-to-global and near-to-global rewiring.) Given Condition 2.3.1(R1), we can use Corollary A.2.4, which combined with (A.3) yields Corollary 2.1.7. □

§2.4.3 Near-to-global rewiring

In this section we focus on near-to-global rewiring. In view of Definition 2.4.2, this is a rewiring with $K_t = \mathrm{Near}_{C_{t-1},r_n}(X_{t-1})$ (recall Definition 2.4.1) and $L_t \equiv C_{t-1}$ (see Remark 2.4.3) at any time t. Just like in the previous example, this is also a rewiring mechanism where the sets K_t are dependent on the current position of the random walk.

The layout is the same as in the previous section, the main difference being the presence of the additional parameter r_n that controls the size of the set of edges that are being considered for rewiring at each unit of time. We will see that this parameter controls the trichotomy. We only consider $r_n = O(\log n)$, since the expected diameter of the configuration model is of order $\log n$ (see (2.12) and [79, 80]). For $r_n = o(\log n)$ the behaviour is dominated by the local properties of the graph dynamics and is similar to that for the local-to-global rewiring studied in Section 2.4.2. On the other hand, once $r_n = \Theta(\log n)$ we get a significant contribution from a certain "boundary term" in the computation of the tail probability $\mathbb{P}(\tau > t \mid \mathsf{SA}(t))$, and we find a behaviour that is more similar to the global-to-global rewiring studied in Section 2.4.4.

First, we claim that the random walk is again irreducible and aperiodic:

Proposition 2.4.6 (Irreducibility and aperiodicity). *Non-backtracking random walk with near-to-global rewiring is aperiodic and irreducible.*

Proof. In Appendix A.1 we show that the joint Markov chain with local-to-global rewiring is irreducible and aperiodic. Since near-to-global rewiring admits all the transitions that are admitted for local-to-global rewiring, the proof carries over. □

Next, we claim that the stationary distribution is again uniform:

Proposition 2.4.7 (Stationary distribution). *For any $\alpha_n \in [0,1]$ and $r_n = O(\log n)$, $U_H \times U_{Conf_H}$ is the stationary distribution for the random walk with near-to-global rewiring with parameters α_n, r_n.*

Proof. Apply Proposition A.3.2 to establish that U_{Conf_H} is stationary for the chosen graph dynamics. After that the rest of the proof carries over from Proposition 2.4.5. □

We are now ready to prove Theorem 2.1.5(B) and Corollary 2.1.8. First we settle Condition 2.3.4(D2) for *good* histories. After that we identify the asymptotics of $\mathbb{P}(\tau > t \mid \mathsf{SA}(t))$ and settle Condition 2.3.4(D1) for *good* histories. Both are tricky because they force us to investigate the possible occurrence of *short-cuts* in the configuration (see Figure 2.7). The key ingredient in the proof is that short-cuts are unlikely when $t = O(\log n)$ and $r_n \leq (1-\varepsilon)\rho_{\max} \log n$ for some $\varepsilon > 0$, which requires the error term in Condition 2.3.4(D1). We finally settle Condition 2.3.4(D3). At the end we put the pieces together and wrap up the proof.

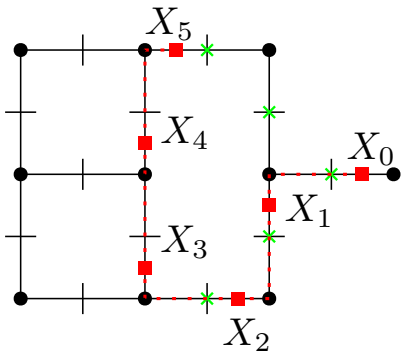


Figure 2.7: Illustration of shortcuts for the near-to-global rewiring with $r_n = 3$. A non-backtracking random walk starts on half-edge X_0 and moves along the dotted red path. Green crosses denote edges that might be rewired in the first step of the rewiring. Note that green crosses are not only on edges that contain the half-edges X_1 and X_2, but also on the edge that contain the half-edge X_5, which will be stepped over after more than $(r_n - 1)$ steps.

Proof of Condition 2.3.4(D2). Because the rewiring is done with the global set, we have

$$\mathbb{P}\big(C_t(x_{t-1}) \in \cdot \mid \mathsf{DSA}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]}) \cap \{I_t = 1\}\big) = U_{H\setminus\{x_{t-1}, C_{t-1}(x_{t-1})\}}, \tag{2.76}$$

and, just as in (2.73),

$$d_{\mathrm{TV}}\left(\mathbb{P}\big(C_t(x_{t-1}) \in \cdot \mid \mathsf{DSA}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]}) \cap \{I_t = 1\}\big), U_H(\cdot)\right) = \frac{2}{|H|}. \tag{2.77}$$

Thus, Condition 2.3.4(D2) is satisfied. □

Identification of $\mathbb{P}(\tau > t \mid \mathsf{SA}(t))$. On the event $\mathsf{SA}(t)$, for $1 \le k < l \le t$, let $S_{kl}^{r_n}$ be the indicator of the event that there is a *short-cut* of length $\le r_n$ between the half-edges visited by the random walk *at* times k and l, i.e., a connection *not* running along the path of the random walk itself. Abbreviate $\mathrm{SH}^{r_n}(t) = (S_{kl}^{r_n})_{1 \le k < l \le t}$. Then, for any x, ξ,

$$\begin{aligned}&\mathbb{P}_{x,\xi}(\tau > t \mid \mathsf{SA}(t),\ \mathrm{SH}^{r_n}(t)) \\ &\quad = \prod_{i=1}^{(t-r_n)_+} (1-\alpha_n)^{\sum_{l=i+1}^{i+r_n}(1+\sum_{k=1}^{i-1} S_{kl}^{r_n})} \prod_{i=(t-r_n)_+ +1}^{t} (1-\alpha_n)^{\sum_{l=i+1}^{t}(1+\sum_{k=1}^{i-1} S_{kl}^{r_n})}.\end{aligned} \tag{2.78}$$

This equality comes from the requirement that from time 1 until time $(t - r_n)_+$ none of the r_n half-edges on the *future path* must be rewired, while from time $(t - r_n)_+ + 1$ until time t none of the r_n half-edges on the *future path until time* t must be rewired.

Rewrite (2.78) as

$$\mathbb{P}_{x,\xi}(\tau > t \mid \mathsf{SA}(t),\ \mathrm{SH}^{r_n}(t)) = (1-\alpha_n)^{(t-r_n)_+ r_n + \frac{1}{2}[t-(t-r_n)_+][t-(t-r_n)_+-1]} \times (1-\alpha_n)^{\chi^{r_n}(t)}, \tag{2.79}$$

with

$$\begin{aligned}\chi^{r_n}(t) &= \sum_{i=1}^{(t-r_n)_+} \sum_{l=i+1}^{i+r_n} \sum_{k=1}^{i-1} S_{kl}^{r_n} + \sum_{i=(t-r_n)_++1}^{t} \sum_{l=i+1}^{t} \sum_{k=1}^{i-1} S_{kl}^{r_n} \\ &= \sum_{i=1}^{t} \Bigg(\sum_{\substack{k\in(0,\,i)\\ l\in(i,\,t\wedge(i+r_n)]}} S_{kl}^{r_n} \Bigg).\end{aligned} \tag{2.80}$$

The first factor in (2.79) equals

$$\begin{cases} \exp\big(-[1+o(1)]\,\frac{1}{2}\alpha_n t^2\big), & t \le r_n, \\ \exp\big(-[1+o(1)]\,\alpha_n[r_n(t-r_n)+\frac{1}{2}r_n^2]\big), & t \ge r_n, \end{cases} \tag{2.81}$$

and produces the scaling in Theorem 2.1.5(B) (recall (2.75)). We therefore need to show that the second factor in (2.79) is negligible. For this it suffices to show the following:

Lemma 2.4.8 (Bound on number of short-cuts). *Subject to Condition 2.3.5,* $\chi^{r_n}(t) = 0$ whp *uniformly in* $t = O(\log n)$ *and* $r_n \le (1-\varepsilon)\rho_{\max}\log n$ *for some* $\varepsilon > 0$.

Proof. Recall that $\mathsf{SA}(t)$ is the event that the random walk is self-avoiding until time t. Consider the ball $B_t(x)$ of radius t around the starting point x of the random walk. Recall that, conditionally on $\mathsf{SA}(t)$, (2.2) implies that the probability for the random walk to choose a t-step self-avoiding path consisting of half-edges $\vec{h} = (h_0, \dots, h_{t-1})$ in $B_t(x)$ equals

$$\prod_{i=0}^{t-1} \frac{1}{\deg_H(h_i)}. \tag{2.82}$$

Condition on $\vec{h}$. Note that $\mathsf{SA}(t)$ is equivalent to the event that all half-edges in $\vec{h}$ are distinct, which we assume from now on.

It is helpful to distinguish between *disjoint* short-cuts and *non-disjoint* short-cuts. A disjoint short-cut between two half-edges h_i and h_j is a short-cut that does not use any of the other half-edges in $\vec{h}$. Not all short-cuts are disjoint. Indeed, a disjoint short-cut gives rise to other short-cuts that are counted in $\sum_{1\le k\le t} S_{kt}^{r_n}$, which we call non-disjoint. For example, for $r_n \ge 2$, if there is a disjoint short-cut of one edge between h_i and h_{i+4}, then there necessarily is a short-cut between h_i and h_{i+5} also. The point is that $\chi^{r_n}(t) = 0$ precisely when there are no disjoint short-cuts. We must also bring the *graph dynamics* into the picture.

We call a disjoint short-cut a disjoint (s,i,j,k)-short-cut when the r_n-neighbourhood of the random walk at time s creates a disjoint short-cut consisting of k edges between h_i and h_j. This is only possible when $s \le i \le s + r_n$ and $k \le r_n$, since otherwise

h_i would not be in the r_n-neighbourhood of the random walk at time s, and when $j > i + r_n$, since otherwise the path of k-edges would not be a short-cut.

We aim to show that, for $r_n \leq (1-\varepsilon)\rho_{\max}\log n$ and $\varepsilon > 0$, the probability that there exists a disjoint (s,i,j,k)-short-cut vanishes as $n \to \infty$. To do so, we rely on the first-moment method. We make crucial use of the fact that the configuration model is the stationary distribution under our graph dynamics. This implies that, conditionally on $\vec{h}$, all other half-edges at time s are paired uniformly at random, so that we can use configuration model estimates. Given $\vec{h}$, the expected number of disjoint (s,i,j,k)-short-cuts is bounded by (see [79, Proposition 7.4])

$$[1+o(1)]\,\frac{\deg_H(h_i)\deg_H(h_j)}{\widehat{\ell_n}}\,\widehat{\nu}_n^{k-1}, \tag{2.83}$$

with

$$\widehat{\ell_n} = \ell_n - O(\log n), \qquad \widehat{\nu}_n = \nu_n\,\frac{\ell_n}{\widehat{\ell_n}}, \tag{2.84}$$

where ν_n is the size-biased mean of the empirical degree distribution p_n (recall (2.40)), ℓ_n is the sum of the degrees (= number of half edges), and the error term $o(1)$ is uniform in $k \leq C\log n$. The quantities in (2.84) introduce corrections that come from the fact that, conditionally on $\vec{h}$, only a subset of size $\widehat{\ell_n}$ of the half-edges is randomly paired at time s. Due to Condition 2.3.5, the sum over $1 \leq k \leq r_n \leq (1-\varepsilon)\rho_{\max}\log n$ of this expression is bounded by $(\max_{1\leq i\leq t}\deg_H(h_i))^2 n^{-\varepsilon/2}$ for n large enough. Thus, for $r_n \leq (1-\varepsilon)\rho_{\max}\log n$, by a union bound over $1 \leq i,j \leq t$, the probability that there exists a disjoint short-cut before time t is bounded by

$$r_n t^2 \Big(\max_{1\leq i\leq t}\deg_H(h_i)\Big)^2 n^{-\varepsilon/2}. \tag{2.85}$$

Since $t = O(\log n)$, we can use an annealing argument to show that, subject to Condition 2.3.5, $\max_{1\leq i\leq t}\deg_H(h_i) \leq t^2$ whp. Indeed, let $\tilde{h}_i$ denote the half-edge to which h_i is paired, so that $\deg_H(h_{i+1}) = \deg_H(\tilde{h}_i)$. Then, the distribution of $\deg_H(\tilde{h}_i)$ is the size-biased degree distribution minus 1. By Condition 2.3.5, the mean of this size-biased distribution is uniformly bounded, so that by the Markov inequality the probability that $\deg_H(h_{i+1}) \geq B$ is at most C/B for any $B > 0$ and some $C < \infty$. Hence the probability that $\max_{1\leq i\leq t}\deg_H(h_i) > t^2$ is at most $Ct/t^2 = o(1)$.

Since $r_n, t = O(\log n)$, we conclude that the probability that $\chi^{r_n}(t) > 0$ is whp at most

$$r_n t^6 n^{-\varepsilon/2} = o(1), \tag{2.86}$$

as required. □

We can now complete the identification of $\mathbb{P}(\tau > t \mid \mathsf{SA}(t))$. By Lemma 2.4.8, $\mathbb{P}_{x,\xi}(\tau > t \mid \mathsf{SA}(t),\ \mathrm{SH}^{r_n}(t))$ is asymptotically equal to the expression in (2.81) whp, uniformly in $r_n \leq (1-\varepsilon)\rho_{\max}\log n$ and $t = O(\log n)$. Taking the expectation w.r.t. $\mathrm{SH}^{r_n}(t)$, we get that the same is true for $\mathbb{P}_{x,\xi}(\tau > t \mid \mathsf{SA}(t))$. Taking the expectation w.r.t. x, ξ as well, we conclude that the same is true for $\mathbb{P}(\tau > t \mid \mathsf{SA}(t))$, as required. □

Proof of Condition 2.3.4(D1). For all paths that describe a dynamically self-avoiding history with respect to $T \subset [t-1]$, the probability that at time t the random walk steps along a rewired edge is

$$\mathbb{P}\big(I_t = 1 \mid \mathsf{DSA}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]})\big) = \beta_{n,t} + \varepsilon_{n,t}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]}), \tag{2.87}$$

with (recall t_r from (2.34))

$$\beta_{n,t} = 1 - (1-\alpha_n)^{(t-t_r)\wedge r_n} \tag{2.88}$$

being the probability that the t^{th} edge is rewired when it is in the range of the random walk path, and $\varepsilon_{n,t}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]}) \geq 0$ is the contribution due to short-cuts. Note that $\beta_{n,t}$ is independent of $(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]})$, so that to verify Condition 2.3.4(D1), we only need to bound $\varepsilon_{n,t}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]})$.

To identify $\varepsilon_{n,t}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]})$, we write

$$\begin{aligned} &\varepsilon_{n,t}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]}) \\ &\quad = (1-\beta_{n,t})\,\mathbb{E}\Big[[1-(1-\alpha_n)^{\chi_*^{r_n}(t)}] \mid \mathsf{DSA}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]})\Big], \end{aligned} \tag{2.89}$$

with

$$\chi_*^{r_n}(t) = \sum_{k=1}^{(t-r_n)_+} S_{kt}^{r_n}, \tag{2.90}$$

where $\varepsilon_{n,t}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]})$ is the probability that the t^{th} edge is rewired due to a short-cut that puts it in the r_n-neighbourhood of the location of the random walk at some time $k < t - r_n$, but is not rewired due to a rewiring on the path of the random walk. The crux of the argument is to show that the event $\mathsf{DSA}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]})$ affects a negligible amount of half-edges. After that we are in a situation where we can once again apply configuration model estimates, as in (2.83).

The event $\mathsf{DSA}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]})$ implies certain restrictions on the pairing of half-edges for every $s \in [0, t-1]$. These restrictions can be of two kinds: they can pair two half-edges with certainty or with a probability that depends on the fine details of the rewiring dynamics. In the near-to-global case these probabilities are generally close to 1. Denote by H_s the (partially) random set of half-edges that are paired by the event $\mathsf{DSA}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]})$ at time$s$. The following observation is crucial:

Lemma 2.4.9 (Random pairings outside H_s). *Conditionally on* $\mathsf{DSA}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]})$, *the half-edges in* $H \setminus H_s$ *are paired and rewired randomly at any time* $s \in [0, t-1]$. *Furthermore,* $H_s \subseteq H'$, *where* $H' = (x_{[0,t-1]} \cup \bar{x}_{[0,t-1]} \cup \hat{x}_{[r]} \cup \tilde{x}_{[r]})$.

Proof. Since the graph is initially drawn according to the configuration model, and the configuration model is the stationary distribution of the graph dynamics, we see that on the set $H \setminus H_s$ the pairing is uniformly at random. Because the paired half-edges in H_s are fixed, they do not affect the half-edges in $H \setminus H_s$. Let us clarify the possible restrictions implied by $\mathsf{DSA}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]})$ at time s:

(a) Edges already traversed by the random walk can get stuck in the configuration seen by the random walk. More formally, edges $\{X_q, C_q(X_q)\}$ with $q \le s$ need not be a part of the near-set $\mathrm{Near}_{C_{p-1},r_n}(X_{p-1})$ for any time $p \ge s$. This concerns half-edges in $x_{[0,t-1]}, \bar{x}_{[0,t-1]}$ and $\hat{x}_{[r]}$.

(b) Edges that are traversed at time q with $q > s$ and $q \notin T$ must not get rewired before the random walk crosses them. This concerns half-edges in $x_{[0,t-1]}$ and $\bar{x}_{[0,t-1]}$.

(c) Edges that are traversed at time q with $q > s$ and $q \in T$ can (but need not) get rewired before the random walk crosses them. If they get rewired just before the random walk crosses them and near-sets at times $< q$ do not contain $\tilde{x}_q$, then $\{x_q, \bar{x}_q\}$ and $\{\hat{x}_q, \tilde{x}_q\}$ must remain paired until time q. This concerns half-edges in $x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}$ and $\tilde{x}_{[r]}$.

Observe that only the edges that consist of half-edges in $x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]}$ can be fixed. If we take the union of all these half-edges H', we get a crude upper estimate on H_s that is valid for all $s \in [0, t-1]$. □

Next we estimate the number of half-edges that are influenced by the restrictions implied by $\mathsf{DSA}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]})$:

Lemma 2.4.10 (Estimate of influenced half-edges). *Conditionally on* $\mathsf{DSA}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]})$, *$H_s$ satisfies the estimate $|H_s| = O(t)$ for any $s \in [0, t-1]$.*

Proof. In view of Lemma 2.4.9, it suffices to bound the number of half-edges in the sequences $x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]}$, namely, $|H'| = O(t)$. The sequences $x_{[0,t-1]}$ and $\bar{x}_{[0,t-1]}$ each contain $t-1$ half-edges by definition. The numbers of half-edges in $\hat{x}_{[r]}$ and $\tilde{x}_{[r]}$ depend on the set $T \subset [t-1]$ of times when the random walk steps over a rewired edge. Pick $T = [t-1]$ to see that $\hat{x}_{[r]}$ and $\tilde{x}_{[r]}$ both contain at most $t-1$ half-edges. Summing the four contributions, we see that indeed $|H'| = O(t)$. □

We are now ready to apply configuration-model estimates:

Lemma 2.4.11 (Bound on number of short-cuts). *Subject to Condition 2.3.5, conditionally on* $\mathsf{DSA}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]})$, *$\chi_*^{r_n}(t) = 0$ whp uniformly in $t = O(\log n)$, $r_n \le (1-\varepsilon)\rho_{\max} \log n$ for some $\varepsilon > 0$, and $x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]}$.*

Proof. Observe that $\chi_*^{r_n}(t) > 0$ implies the existence of a (s,i,j,k)-short-cut at some time $s \in [0, t-1]$. In Lemma 2.4.8 we proved a result about rarity of these short-cuts where we assumed only Condition 2.3.5. The statement of the current lemma furthermore assumes that the event $\mathsf{DSA}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]})$ occurs.

In Lemma 2.4.9 we have shown that at any time $s \in [0, t-1]$ the conditioning on $\mathsf{DSA}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]})$ only affects the pairing of some half-edges in H_s. In Lemma 2.4.11 we gave an estimate of $|H_s|$ for any $s \in [0, t-1]$. These two results bring us into the same setting as we had in the proof of Lemma 2.4.8, namely, we see that configuration model estimates hold (recall (2.83)). Therefore, by the same argument as above, given that $r_n, t = O(\log n)$, we once again claim that the probability of

$\chi_*^{r_n}(t) > 0$ is at most

$$r_n t^2 \Big(\max_{1 \le i \le t-1} \deg_H(x_i) \Big)^2 n^{-\varepsilon/2}. \tag{2.91}$$

Since Condition 2.3.4(D1) concerns sequences $x_{[0,t-1]}$ that are good, we have

$$\max_{1 \le i \le t-1} \deg_H(x_i) \le (\log n)^{2+\varepsilon}, \tag{2.92}$$

and so

$$r_n t^2 ((\log n)^{2+\varepsilon})^2 n^{-\varepsilon/2} = o(1), \tag{2.93}$$

as required. Note that $\chi_*^{r_n}(t) \le \chi^{r_n}(t)$ (compare (2.80) and (2.90)). □

Now we see that the contribution of the $\varepsilon_{n,t}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]})$ term in (2.87) is $O(n^{-\varepsilon/2})$ and therefore Condition 2.3.4(D1) holds. □

Proof of Condition 2.3.4(D3). Observe that, for $t = O(\log n)$,

$$\begin{aligned}
\mathbb{P}\left(X_{[0,t-1]} \text{ is bad}\right) &= \mathbb{P}\left(\exists\, 1 \le i \le t\colon\ \deg_H(X_i) > (\log n)^{2+\varepsilon}\right) \\
&\le \sum_{1 \le i \le t} \mathbb{P}\left(\deg_H(X_i) > (\log n)^{2+\varepsilon}\right) \\
&\le \sum_{1 \le i \le t} \frac{\mathbb{E}[\deg_H(X_i)]}{(\log n)^{2+\varepsilon}} \\
&\le t\, \frac{\max_{1 \le i \le t} \mathbb{E}[\deg_H(X_i)]}{(\log n)^{2+\varepsilon}} \\
&= O\left(\frac{1}{(\log n)^{1+\varepsilon}}\right),
\end{aligned} \tag{2.94}$$

where we use that $\max_{1 \le i \le t} \mathbb{E}[\deg_H(X_i)]$ is finite by Condition 2.3.5. Since $\varepsilon > 0$, Condition 2.3.4(D3) follows. □

Completion of the proof of Theorem 2.1.5(B) and Corollary 2.1.8. We already verified Condition 2.3.4, and have shown that $\mathbb{P}(\tau > t \mid \mathsf{SA}(t))$ is asymptotically equal to the expression in (2.81). Furthermore, by (2.75), $\mathsf{SA}(t)$ occurs whp, uniformly in $t = O(\log n)$. This completes the proof of Theorem 2.1.5(B). Finally, given Condition 2.3.1(R1), we can again use Corollary A.2.4, which combined with (A.3) yields Corollary 2.1.8.

□

§2.4.4 Global-to-global rewiring

In this section we focus on global-to-global rewiring. This choice was already explored in [9], [10], with the minor difference that in the present chapter the parameter α_n is the probability that an edge gets rewired per unit of time, while in [9], [10] it was the *fraction* of edges that get rewired per unit of time. This difference has no impact on the scaling of the mixing times. Global-to-global rewiring corresponds to the choice $K_t = L_t \equiv C_{t-1}$ (see Remark 2.4.3) for all t in Definition 2.4.2. Unlike for the previous examples, now the rewiring is *independent* of the position of the random walk, so the graph dynamics becomes Markovian.

As before, the use of Corollary A.2.4 depends on Condition 2.3.1(R1). The proof of Theorem 2.1.5(C) uses that for all x and ξ,

$$\mathbb{P}_{x,\xi}(\tau > t \mid \mathsf{SA}(t)) = \prod_{i=1}^{t}(1-\alpha_n)^{t-i} = \exp\big(-[1+o(1)]\,\tfrac{1}{2}\alpha_n t^2\big). \tag{2.95}$$

The first equality comes from the requirement that up to time t each of the half-edges on the future path of the random walk up must not get rewired. We thus obtain the scaling in Theorem 2.1.5(C) (again recall (2.75)). Given Condition 2.3.1(R1), we can again use Corollary A.2.4, which combined with (A.3) yields Corollary 2.1.9.

Irreducibility and aperiodicity of the rewiring was settled in [9]. The fact that the stationary distribution is the configuration model is settled by Proposition A.3.2, in combination with an argument analogous to Proposition 2.4.5. It remains to establish Condition 2.3.4.

Proposition 2.4.12 (Graph dynamics regularity for global-to-global rewiring). *Global-to-global rewiring satisfies the graph-dynamics regularity conditions formulated in Condition 2.3.4.*

Proof. Since any edge can get rewired at any time, we have

$$\mathbb{P}\big(I_t = 1 \mid \mathsf{DSA}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]})\big) = \alpha_n^t, \tag{2.96}$$

where we use that the edge crossed at time t has had exactly t opportunities to get rewired. Since this holds for any choice of $x_{[0,t-1]}$, $\bar{x}_{[0,t-1]}$, $\hat{x}_{[r]}$ and $\tilde{x}_{[r]}$, Conditions 2.3.4(D1) follows with zero error. Moreover, since a half-edge can get rewired to any half-edge except itself and its current pair, we know that $\mathbb{P}(C_t(x_{t-1}) \in \cdot \mid \mathsf{DSA}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]}) \cap \{I_t = 1\})$ is the uniform distribution on $H \setminus \{x_{t-1}, C_{t-1}(x_{t-1})\}$, which gives

$$d_{\mathrm{TV}}\big(\mathbb{P}\big(C_t(x_{t-1}) \in \cdot \mid \mathsf{DSA}(T, x_{[0,t-1]}, \bar{x}_{[0,t-1]}, \hat{x}_{[r]}, \tilde{x}_{[r]}) \cap \{I_t = 1\}\big), U_H(\cdot)\big) = \frac{2}{|H|}. \tag{2.97}$$

Since (2.97) holds for any choice of $x_{[0,t-1]}$, $\bar{x}_{[0,t-1]}$, $\hat{x}_{[r]}$, $\tilde{x}_{[r]}$, Condition 2.3.4(D2) also follows. □

Remark 2.4.13 (Comparison with previous results). The proof in [9] and [10] required a condition analogous to Condition 2.3.5, while in the present proof this is no longer needed. ♦

APPENDIX A

Appendices of Chapter 2

§A.1 Irreducibility and aperiodicity of local-to-global dynamics

In this section we show that the random walk with *local-to-global* rewiring is irreducible and aperiodic. This ensures that the total variation distance $\mathcal{D}_{x,\xi}(t)$ converges to 0 as $t \to \infty$ for fixed $x \in H$, $\xi \in \mathit{Conf}_H$ and $\alpha_n \in (0,1)$. Our proof builds on the proof of irreducibility of the switch chain on multigraphs given in [51].

Proposition A.1.1 (Irreducibility and aperiodicity). *The random walk with local-to-global rewiring $(X_t, C_t)_{t\in\mathbb{N}}$ (see Section 2.4.2) is irreducible and aperiodic for any initial state $(x,\xi) \in H \times \mathit{Conf}_H$ and any choice of $\alpha_n \in (0,1)$.*

Proof. Let $V = \{v_1, \dots, v_n\}$ and assume that $\deg(v_1) \le \deg(v_2) \le \cdots \le \deg(v_n)$. Identify the set of half-edges H with $[|H|] = \{1, \dots, |H|\}$, such that the half-edges $1, \dots, \deg(v_1)$ are associated to v_1, the half-edges $\deg(v_1)+1, \dots, \deg(v_1)+\deg(v_2)$ to v_2, and so on. Let $v'_1, \dots, v'_{2k} \in V$ be the odd-degree vertices. We fix a configuration $\xi_0 \in \mathit{Conf}_H$ such that each vertex has the maximum number of self-loops, i.e., each vertex $v \in V$ with even degree has $\frac{1}{2}\deg(v)$ self-loops, each vertex $v \in V$ with odd degree has $\frac{1}{2}(\deg(v)-1)$ self-loops, and there is exactly one edge between every pair of odd-degree vertices v'_{2i-1}, v'_{2i} for $i = 1, \dots, k$ (see Figure A.1). We will show that the pair $(1,\xi_0) \in H \times \mathit{Conf}_H$ is accessible from any pair $(x,\xi) \in H \times \mathit{Conf}_H$ by allowed moves for the random walk with local rewiring.

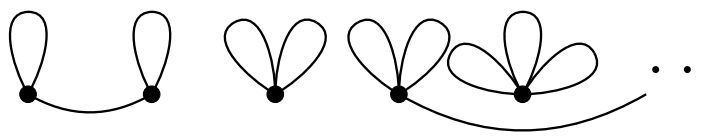


Figure A.1: The configuration ξ_0.

First we show that, for any $x \in H$, $(1,\xi_0)$ is accessible from (x,ξ_0), by considering two different scenarios:

(a) Suppose that x is on a self-loop and $\xi_0(x) = x'$. We first move to $(1,\xi_1)$ from $(1,\xi_0)$ by rewiring the half-edges $x, x', 1, 2$ where ξ_0 and ξ_1 agree on all the edges except that $\xi_1(1) = x'$ and $\xi_1(2) = x$. After that we again move to $(1,\xi_0)$ from $(1,\xi_1)$ by rewiring $1, 2, x, x'$ (see Figure A.2).

(b) Suppose that x is not on a self-loop, i.e., it is on an edge between two odd-degree vertices. We first move to (x',ξ_0) without rewiring, where $x' \in H$ is on a self-loop. After that we apply the procedure in item 1 to (x',ξ_0).

Next, we show that for any $(x,\xi) \in H \times \mathit{Conf}_H$ with $\xi \neq \xi_0$ we have access from (x,ξ) to (y,ξ_0), for some $y \in H$. To do this, we show that we can move from (x,ξ) to some $(y,\eta) \in H \times \mathit{Conf}_H$ such that the configuration η has more edges in common with ξ_0 than ξ has, i.e., $|\xi \cap \xi_0| < |\eta \cap \xi_0|$, by considering the two scenarios:

(a) Suppose that x is on an edge that is not in ξ_0, i.e., $\xi(x) \neq \xi_0(x)$. Then we move to (y,η) by rewiring the half-edges $x, \xi(x), \xi_0(x), \xi(\xi_0(x))$, where ξ and η agree

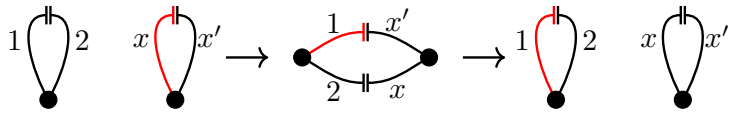


Figure A.2: Move from half-edge x on a self-loop to half-edge 1 in ξ_0. The red colour indicates the position of the walk.

on all the edges except that $\eta(x) = \xi_0(x)$ and $\eta(\xi(x)) = \xi(\xi_0(x))$ and $y \sim \xi_0(x)$. Since $\eta(x) = \xi_0(x)$, we have $|\xi \cap \xi_0| \leq |\eta \cap \xi_0| - 1$.

(b) Suppose that x is on an edge that is in ξ_0, i.e., $\xi(x) = \xi_0(x)$. Let $y \in H$ be a half-edge such that $\xi(y) \neq \xi_0(y)$, $\xi(x) = x'$ and $\xi(y) = y'$. Since $\deg(v) \geq 2$ for all $v \in V$, in the graph given by ξ there is a cycle of edges $\{y, y'\}, \{y_1, y_1'\}, \ldots, \{y_K, y_K'\}$ with $v(y') = v(y_1)$, $v(y_K') = v(y)$ and $v(y_i') = v(y_{i+1})$ for $i = 1, \ldots, K-1$. Let $\eta \in \mathit{Conf}_H$ be the configuration that agrees with ξ on all the edges except that $\eta(x) = y'$ and $\eta(y) = x'$, so that the edges $\{y_1, y_1'\}, \ldots, \{y_K, y_K'\}$ are present in η as well as in ξ. First we move from (x, ξ) to (y_1, η) by rewiring x, x', y, y'. Then we make K moves, from (y_i, η) to (y_{i+1}, η) for $i = 1, \ldots, K$, where $y_{K+1} = y$ without rewiring. After that we move from (y, η) to (y_1, ξ) by rewiring x, x', y, y', and finally we traverse the cycle again without rewiring to reach (y, ξ) from (y_1, ξ) (see Figure A.3). Now y is on an edge that is not in ξ_0, so by applying the procedure in item 1 we can increase the number of edges we have in common with ξ_0.

By applying these procedures, we can reduce the number of edges that are not in ξ_0. So, we can go from any $(x, \xi) \in H \times \mathit{Conf}_H$ to (y, ξ_0) for some $y \in H$, and then apply the above procedure to reach $(1, \xi_0)$.

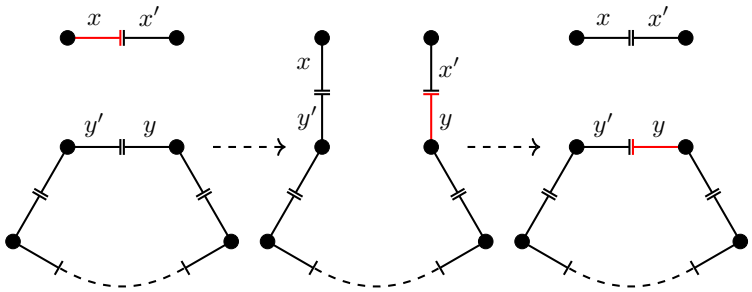


Figure A.3: Moving from (x, ξ) to (y, η) by using a cycle. The red colour indicates the position of the walk.

To show that we can access an arbitrary state (x, ξ) from $(1, \xi_0)$, we first note that we can access (y, ξ_0), for any y, from $(1, \xi_0)$ by relabelling the half-edges and using the first argument above. Then we see that we can access (x, ξ) from (y, ξ_0) for any y by using the above strategy of reducing the edges and using the cycles to move around. Hence, the Markov chain is irreducible. Since, by traversing the self-loop without rewiring, we can reach $(1, \xi_0)$ from itself in one step, we see that the Markov chain is

also aperiodic. □

§A.2 Cut-off in the static setting

In order to use the results of [21], we need to assume the conditions stated there:

Condition A.2.1 (Additional regularity of degrees).

(R1**) $d_{\max} \coloneqq \max_{v\in V} \deg(v) = n^{o(1)}$ *as* $n \to \infty$.

(R2**)

$$\frac{\lambda_2}{\lambda_1^3} = \omega\left(\frac{(\log\log|H|)^2}{\log|H|}\right), \qquad \frac{\lambda_2^{3/2}}{\lambda_3\sqrt{\lambda_1}} = \omega\left(\frac{1}{\sqrt{\log|H|}}\right), \quad n\to\infty,$$

where

$$\lambda_1 \coloneqq \frac{1}{|H|}\sum_{z\in H}\log(\deg_H(z)),$$
$$\lambda_m \coloneqq \frac{1}{|H|}\sum_{z\in H}|\log(\deg_H(z)) - \lambda_1|^m, \qquad m = 2,3.$$

(R3**) $\deg(v) \geq 3$ *for all* $v \in V$.

Conditions A.2.1(R1**) and (R2**) are technical and proof-generated. It might be possible to relax them via a truncation argument [28]. Condition A.2.1(R3**) ensures that the random walk does not behave deterministically and that the configuration model is connected whp. Note that (R1**) and (R3**) are considerably more stringent than (R2) and (R3) in Condition 2.3.1.

As shown in [21], the following holds:

Theorem A.2.2 (Scaling of static mixing time). *Subject to Condition A.2.1,*

$$\mathcal{D}^{\text{stat}}_{x,\xi}(t) = \begin{cases} 1-o_{\mathbb{P}}(1), & \text{if } \limsup_{n\to\infty} t/t^{\text{stat}}_{\text{mix}} < 1,\\ o_{\mathbb{P}}(1), & \text{if } \liminf_{n\to\infty} t/t^{\text{stat}}_{\text{mix}} > 1,\end{cases} \tag{A.1}$$

where

$$t^{\text{stat}}_{\text{mix}} \coloneqq [1+o_{\mathbb{P}}(1)]\, c_n^{\text{stat}} \log n \qquad \forall\, \varepsilon \in (0,1), \tag{A.2}$$

with $1/c_n^{\text{stat}} \coloneqq \frac{1}{|H|}\sum_{z\in H}\log(\deg_H(z)) \in (0,\infty)$.

If, in addition,

$$\lim_{n\to\infty} c_n^{\text{stat}} = c_*, \tag{A.3}$$

then Theorem A.2.2 yields Theorem 2.1.6.

Remark A.2.3. We are aware of the fact that Condition A.2.1 is in [21] used to prove a much stronger statement than Theorem A.2.2 that is related to an exact computation of the cut-off window. Therefore, if we are interested only in proving Theorem A.2.2, weaker conditions might suffice. ♦

Combining Theorem 2.1.4 and Theorem A.2.2, we obtain the following corollary:

Corollary A.2.4 (Link between static and dynamic). *Suppose $t = O(\log n)$. Subject to Condition 2.3.1(R1), Condition 2.3.4 and Condition A.2.1, the following holds whp in x and ξ:*

$$\mathcal{D}^{\mathrm{dyn}}_{x,\xi}(t) = \begin{cases} \mathbb{P}_{x,\xi}(\tau > t) + o_{\mathbb{P}}(1), & \text{if } \limsup_{n\to\infty} t/t^{\mathrm{stat}}_{\mathrm{mix}} < 1, \\ o_{\mathbb{P}}(1), & \text{if } \liminf_{n\to\infty} t/t^{\mathrm{stat}}_{\mathrm{mix}} > 1. \end{cases} \tag{A.4}$$

§A.3 Transition matrix for graph rewiring

Recall Definition 2.4.2, where we have introduced the general class of rewirings considered in this chapter. In this appendix we provide a general expression for the transition matrix of the graph dynamics. Furthermore, we explore the conditions required for this transition matrix to be doubly stochastic.

Proposition A.3.1 (Transition matrix for (K_t)-to-(L_t) rewiring). *The transition matrix for the rewirings in Definition 2.4.2 is*

$$Q_{K_t\to L_t} = \sum_{k=0}^{|K_t|} (1-\alpha_n)^{|K_t|-k} (\alpha_n)^k \sum_{\{e_1,\dots,e_k\}\in K_t} Q^{K_t\to L_t}_{\{e_1,\dots,e_k\}}. \tag{A.5}$$

The matrix element $Q^{K_t\to L_t}_X(\eta,\xi)$ that represents the rewiring of the edges in the set X that realises the transition from graph state η to graph state ξ is given by

$$Q^{K_t\to L_t}_X(\eta,\xi) = \begin{cases} \dfrac{1}{2^{|X|}\displaystyle\prod_{i=0}^{|X|-1}(|L_t| - |L_t\cap K_t| - i)} & \text{if } \xi \text{ is accessible from } \eta \\ & \text{by rewiring all edges in } X, \\ 0 & \text{otherwise.} \end{cases} \tag{A.6}$$

Observe that the matrix given by (A.5) is a sum of multiple terms. Let us explain the meaning of these terms through the example of the general term

$$(1-\alpha_n)^{|K_t|-k}\alpha_n^k \sum_{\{e_1,\dots,e_k\}\in K_t} Q^{K_t\to L_t}_{\{e_1,\dots,e_k\}}. \tag{A.7}$$

First, the factor $(1-\alpha_n)^{|K_t|-k}$ represents the probability of $|K_t|-k$ edges not getting rewired, and its counterpart α_n^k represents the probability of k edges getting rewired. The sum runs over all k-tuples from K_t, and the matrix $Q^{K_t\to L_t}_{\{e_1,\dots,e_k\}}$ represents the possible rewiring of the k-tuples of edges we are summing over.

The Markov chain transition matrix must be stochastic. Let us check this by an explicit computation. Take an arbitrary graph state η. In the row that lists the

probabilities of all the possible transitions from η, we get the following contributions:

$$\begin{aligned}\sum_{\xi\in Conf_H} Q_{K_t\to L_t}(\eta,\xi) &= \sum_{k=0}^{|K_t|}(1-\alpha_n)^{|K_t|-k}\alpha_n^k\binom{|K_t|}{k}\frac{2^k\prod_{i=0}^{k-1}(|L_t|-|L_t\cap K_t|-i)}{2^k\prod_{i=0}^{k-1}(|L_t|-|L_t\cap K_t|-i)}\\ &= \sum_{k=0}^{|K_t|}(1-\alpha_n)^{|K_t|-k}\alpha_n^k\binom{|K_t|}{k}\\ &= [(1-\alpha_n)+\alpha_n]^{|K_t|}=1.\end{aligned} \tag{A.8}$$

The combinatorial factor $\binom{|K_t|}{k}$ counts the different ways of choosing k-tuples from K_t. Since the entries in $Q_X^{K_t\to L_t}(\eta,\xi)$ are chosen to be the reciprocal of the number of accessible states, it is not surprising that they sum up to 1. The factor 2^k comes from the ability to break up an edge into two ordered sets of half-edges.

Observe that the matrix defined by (A.5) has a "binomial" structure, but that it is *not* of the form

$$Q_X^{K_t\to L_t}(\eta,\xi)=\prod_{e\in X}\left((1-\alpha_n)I+\alpha_n Q_{\{e\}}^{K_t\to L_t}\right). \tag{A.9}$$

Clearly, (A.9) would be correct if we would draw $e'\in R_t'$ in Definition 2.4.2 *with replacement*, when the state space for the rewiring of $|X|$ edges would have size $2^{|X|}\prod_{i=1}^{|X|}(|L_t|-|L_t\cap K_t|)$. In the current setting, where we draw e' *without replacement*, the state space for the rewiring of $|X|$ edges is smaller, namely, size $2^{|X|}\prod_{i=0}^{|X|-1}(|L_t|-|L_t\cap K_t|-i)$, due to the removal of already drawn edges.

While we have observed that the transition matrix is stochastic, it is doubly stochastic only subject to additional conditions. For the purpose of this chapter, we need the following fact:

Proposition A.3.2 (Double stochasticity of (K_t)-to-global rewiring transition matrix). *The transition matrix given in* (A.5) *is doubly stochastic for $L_t\equiv\xi$, in the sense that edges in L_t are generated by pairing ξ of the whole set of half-edges H (recall Remark 2.4.3).*

Proof. The proof is by explicit computation. Choose an arbitrary graph state ξ and count the contributions to the sum over the row corresponding to transitions leading to ξ:

$$\sum_{\xi\in Conf_H} Q_{K_t\to L_t}(\eta,\xi)=\sum_{k=0}^{|K_t|}(1-\alpha_n)^{|K_t|-k}\alpha_n^k\binom{|K_t|}{k}\frac{\prod_{i=0}^{k-1}(|H|-2|K_t|-2i)}{2^k\prod_{i=0}^{k-1}(|L_t|-|L_t\cap K_t|-i)}. \tag{A.10}$$

The term $(|H|-1-(2|K_t|-1)-2i)$ is based on the following observation. We are counting possible pairs of half-edges where we see a difference in ξ compared to η. This way we get the whole set of half-edges $|H|$, without the considered half-edge itself

and without all but one half-edge in K_t. Rewiring cannot create an edge between half-edges that gave rise to K_t, and the term -1 arises from the one half-edge from K_t the considered half-edge is paired with in ξ. The term $-2i$ again arises because we are drawing without replacement. Now observe that $2|L_t| = |H|$ and $L_t \cap K_t = K_t$. Apply the binomial theorem to get the claim. □

CHAPTER 3

Mixing of fast random walks on dynamic random permutations

This chapter is based on the following article:

L. Avena, R. van der Hofstad, F. den Hollander, and O. Nagy. Mixing of fast random walks on dynamic random permutations. *arXiv preprint arXiv:2403.00094*, 2024.

Abstract

We analyse the mixing profile of a random walk on a dynamic random permutation, focusing on the regime where the walk evolves much faster than the permutation. Two types of dynamics generated by random transpositions are considered: one allows for coagulation of permutation cycles only, the other allows for both coagulation and fragmentation. We show that for both types, after scaling time by the length of the permutation and letting this length tend to infinity, the total variation distance between the current distribution and the uniform distribution converges to a limit process that drops down in a single jump. This jump is similar to a one-sided cut-off, occurs after a random time whose law we identify, and goes from the value 1 to a value that is a strictly decreasing and deterministic function of the time of the jump, related to the size of the largest component in Erdős-Rényi random graphs. After the jump, the total variation distance follows this function down to 0.

§3.1 Introduction and main results

§3.1.1 Target

The goal of this chapter is to identify the mixing profile of a fast random walk on a *dynamic random permutation*, where fast means that the random walk instantly achieves local equilibrium, i.e., fully mixes on the cycle of the permutation it sits on before the next change in the permutation occurs. The focus is on two types of dynamics for the permutation, both starting from the identity permutation and consisting of successive applications of *random transpositions*. The first type – called *coagulative dynamics* – imposes the constraint that transpositions leading to fragmentation of a permutation cycle are ignored. The second type – called *coagulative-fragmentative dynamics* – does not impose this constraint. A major feature of dynamic random permutations is that they represent a *disconnected* geometry, which marks a departure from the setting that was considered in earlier work.

We show that for both dynamics, after scaling time by the length of the permutation and letting this length tend to infinity, the total variation distance between the current distribution and the uniform distribution converges to a limit process that makes a *single jump* down from the value 1 to a value on a deterministic curve and subsequently follows this curve on its way down to 0. The aforementioned curve is strictly decreasing in time and is related to the size of the largest component in the Erdős-Rényi random graph. The jump down to this curve, which is similar to a *one-sided cut-off*, occurs after a *random time* whose law we identify. This type of mixing profile is *different* from that of previously studied models. The law of the drop-down time and the function describing the deterministic curve are different for the two types of dynamics. Visual representations of the mixing profiles are given in Figs. 3.1 and 3.3, while simulations are shown in Figs. 3.2 and 3.4.

The model analysed in this chapter is a first step towards understanding the behaviour of a *simple random walk on a dynamic permutation*. This process is, despite its apparent simplicity, difficult to analyse in detail, especially when the stepping rate of the random walk is *commensurate* with the transposition rate of the dynamic permutation. A key tool in our analysis is Schramm's coupling [135]. While this coupling was used previously to study the cycle structure of a dynamic permutation at a *fixed* time, we adapt the arguments in a way that allows us to study the evolution of cycles over a time interval of *diverging length*.

The remainder of this section is organised as follows. Section 3.1.2 provides background and recalls earlier work. Section 3.1.3 fixes the setting and introduces relevant definitions and notations. Section 3.1.4 lists some preliminaries for Erdős-Rényi random graphs that are needed along the way. Section 3.1.5 introduces a graph process associated with the dynamics that serves as a tool for analysing the dynamics. Section 3.1.6 contains two main theorems, one for each type of dynamics, describing the evolution of the total variation distance between the current distribution of the random walk and its equilibrium distribution, which is the uniform distribution on $[n] = \{1, \dots, n\}$. Section 3.1.7 discusses the importance of the main theorems, places them in their proper context, and provides an outline of the remainder of the chapter.

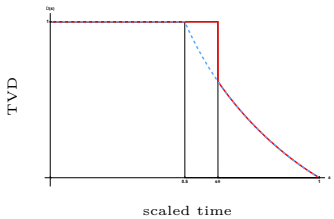


Figure 3.1: The red curve is a typical evolution of the total variation distance for an infinitely fast random walk on a coagulative dynamic permutation. *The blue curve is a plot of the deterministic function of the scaled time to which the total variation distance drops at a random time and subsequently sticks to.*

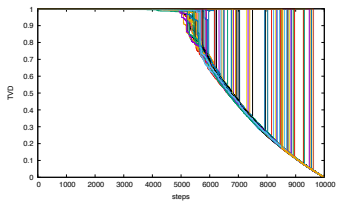


Figure 3.2: Simulations of the evolution of the total variation distance for 10^2 *different realisations of a* coagulative dynamic permutation *of* 10^4 *elements and an infinitely fast random walk on top. Each simulation run corresponds to a single coloured curve.*

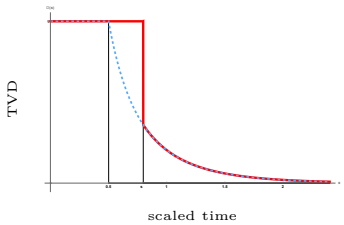


Figure 3.3: The same as Figure 3.1 for a coagulative-fragmentative dynamic permutation.

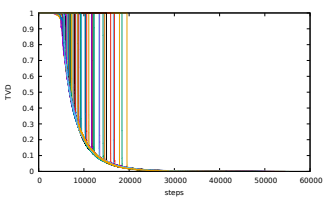


Figure 3.4: The same as Figure 3.2 for a coagulative-fragmentative permutation.

§3.1.2 Background and earlier work

While over the past years random walks on *static* random graphs have received a lot of attention, and the scaling properties of quantities like mixing times, cover times and metastable crossover times have been identified, much less is known about random walks on *dynamic* random graphs. In the static setting, a two-sided cut-off on scale $\log n$ has been established for a general class of *undirected* sparse graphs with good expansion properties [21, 26, 103]. Similar results have been obtained for *directed* sparse graphs [37, 38] and for graphs with a *community structure* [20].

In the dynamic setting, predominantly the focus has been on dynamic percolation, Erdős-Rényi random graphs with edges switching on and off randomly, and configuration models with random rewiring of edges. Both directed and undirected graphs have been considered, as well as backtracking and non-backtracking random walks. In [124, 123, 77] random walks on *dynamic percolation* clusters on a d-dimensional *discrete torus* were considered. Mixing times were identified for several parameter regimes controlling the rate of the random walk and the rate of the random graph dynamics. Similar results were obtained for dynamic percolation on the *complete graph* [138, 133]. Some further advances were achieved in [13], where general bounds on mixing times, hitting times, cover times and return times were derived for certain classes of dynamic random graphs under appropriate *expansion assumptions*. Non-backtracking random walks on configuration models that with high probability are connected were studied in a series of papers [9, 10, 11], which culminated in a general framework for studying mixing times of non-backtracking random walks on dynamic random graphs subject to mild regularity conditions. Mixing of random walks on *directed* configuration models was treated in [42].

Random permutations generated by random transpositions have attracted plenty of interest as well. An important starting point is [49], where a *cut-off* in the total variation distance was established after the application of $\frac{1}{2}n\log n + O(n)$ random transpositions. The *sharp* constant in front of the leading-order term was achieved with the help of *representation theory* for the symmetric group. This paper led to a flurry of follow-up work, of which we mention [135], where the structure of large cycles of an evolving random permutation was studied. Similar results were obtained in [24], including sharp control on the number of observed fragmentations. An important aspect of both [135] and [24] is the representation of an evolving random permutation, starting from the identity permutation, in terms of a random graph process that can be studied by using the theory of random graphs. For the *coagulative-fragmentative* dynamics considered in the present chapter, also called *transposition dynamics*, this graph process representation yields a graph-growth model that at every step adds an edge drawn uniformly at random. This graph-growth model is closely related to the standard "combinatorial" Erdős-Rényi model, whose study is by now a classical topic in the theory of random graphs (see, for example, [62] or [78]). Yet another important feature of [135] is the introduction of *Schramm's coupling* as a tool to study the cycle structure of evolving random permutations. In a follow-up article [27], a modified version of this coupling is used to study the mixing of dynamic permutations endowed with a more general dynamics, of which the transposition dynamics is a special case. We also mention [29], which contains a detailed account of Schramm's coupling. The

works cited above each highlight one particular facet of the random transposition model, but close relatives have been studied extensively under different names: *mean-field Tóth model* [142], the *interchange process* on the complete graph (see [6, 29, 74] and references therein), or *multi-urn Bernoulli–Laplace diffusion models* [128], where our setting corresponds to a particular choice of the model parameters.

The *coagulative* dynamics considered in the present chapter can be recast, in the spirit of [135], as a graph-valued random process that starts with an empty graph on n vertices and describes a forest that progressively merges into a spanning tree on n vertices through the addition of edges that do not create a cycle. The study of this process and its close relatives has a somewhat twisted history. It is similar to the *standard additive coalescent* (see [4] for an overview), but it is also interesting in its own right (see [3, 104]). Finally, there is a wealth of results on *minimal spanning trees* and *Kruskal's algorithm*, which is another closely related process. In particular, we mention [88], since this work implies some facts that we list in Section 3.2. We derive these facts independently, using different techniques in a different setting.

§3.1.3 Setting, definitions and notation

For $n \in \mathbb{N}$, let S_n denote the set of permutations of $[n]$, i.e., bijections from $[n]$ to itself. Recall that S_n endowed with the operation of permutation composition $\circ$ forms a group. Write $\gamma_v(\pi)$ to denote the cycle of the permutation π that contains the element v.

Definition 3.1.1 (Dynamic permutation). A sequence of permutations of $[n]$, denoted by $\Pi_n = (\Pi_n(t))_{t=0}^{t^{\max}}$ with time horizon $t^{\max} \in \mathbb{N}_0 \cup \{\infty\}$, is called a dynamic permutation. ■

Example 3.1.2 (Transpositions may fragment cycles or coagulate cycles). Pick $n = 7$ and consider the permutation

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 2 & 3 & 4 & 5 & 6 & 7 & 1 \end{pmatrix} \quad \text{with cycle structure} \quad (1,2,3,4,5,6,7).$$

The transposition $(1, 5)$ turns this into the permutation

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 6 & 3 & 4 & 5 & 2 & 7 & 1 \end{pmatrix} \quad \text{with cycle structure} \quad (1,6,7)(2,3,4,5).$$

Another application of the same transposition acts in reverse. Note that S_n is a *non-commutative group* for any $n \geq 3$. ♠

We consider two types of dynamic permutations:

Definition 3.1.3 (Coagulative dynamic permutation). $\Pi_n = (\Pi_n(t))_{t=0}^{n-1}$ is called a *coagulative dynamic permutation* (CDP) when $\Pi_n(0) = \mathrm{Id}$ (i.e., the identity permutation) and

$$\Pi_n(t) = \Pi_n(t-1) \circ (a, b), \qquad t \in [n-1], \tag{3.1}$$

where, for each $t \in [n-1]$, (a, b) is a random transposition sampled uniformly at random from the set of all transpositions of $[n]$ that satisfy the *constraint*

$$\gamma_a(\Pi_n(t-1)) \neq \gamma_b(\Pi_n(t-1)). \tag{3.2}$$

The latter guarantees that no cycle of $\Pi_n(t-1)$ is fragmented by the transposition (a, b). ■

Definition 3.1.4 (Coagulative-fragmentative dynamic permutation). $\Pi_n = (\Pi_n(t))_{t=0}^{\infty}$ is called a *coagulative-fragmentative dynamic permutation* (CFDP) when the same holds as in Definition 3.1.3, but without the constraint in (3.2). ■

Remark 3.1.5 (Time horizon for dynamic permutations and cycle structure). Since CDP starts from the identity permutation, it becomes a permutation with a single cycle after exactly $n-1$ steps. Once this happens, there is no permutation that satisfies (3.2) and the dynamics is trapped. CFDP has no traps and can evolve forever. The structure of cycles is random and their sizes, scaled by $1/n$, converge in distribution to the Poisson-Dirichlet distribution with parameter 1 (see e.g. [135, Theorem 1.1] for a precise statement). ♦

Our aim is to study mixing of fast random walks on both CDP and CFDP. To simplify our analysis, we work with *infinite-speed* random walks, as defined next:

Definition 3.1.6 (Infinite-speed random walk on Π_n). Fix Π_n and an element $v_0 \in [n]$. Recall that $\gamma_v(\Pi_n(t))$ is the cycle of $\Pi_n(t)$ that contains v. Formally, the infinite-speed random walk (ISRW) starting from v_0 is a sequence of probability distributions $(\mu^{n,v_0}(t))_{t\in\mathbb{N}_0}$ supported on $[n]$, with initial distribution at time $t = 0$ given by

$$\mu^{n,v_0}(0) = (\mu_w^{n,v_0}(0))_{w\in[n]}\,, \tag{3.3}$$

where $\mu_w^{n,v_0}(0)$, the mass at $w \in [n]$ at time $t = 0$, is given by

$$\mu_w^{n,v_0}(0) = \begin{cases} \frac{1}{|\gamma_w(\Pi_n(0))|}, & w \in \gamma_{v_0}(\Pi_n(0)), \\ 0, & w \notin \gamma_{v_0}(\Pi_n(0)), \end{cases} \tag{3.4}$$

and with distribution at a later time $t \in \mathbb{N}$ given by

$$\mu^{n,v_0}(t) = (\mu_w^{n,v_0}(t))_{w\in[n]}\,, \tag{3.5}$$

where

$$\mu_w^{n,v_0}(t) = \frac{1}{|\gamma_w(\Pi_n(t))|} \sum_{u\in\gamma_w(\Pi_n(t))} \mu_u^{n,v_0}(t-1). \tag{3.6}$$

Informally, the distribution of ISRW *spreads infinitely fast over the cycle in the permutation it resides on.* ■

In Appendix B.1 we show that the infinite-speed random walk arises as the limit of a standard random walk whose stepping rate relative to the rate of the permutation dynamics tends to infinity. Note that the evolution of the ISRW distribution is fully determined by the initial position of the random walk and the realisation of the dynamic permutation. See Figure 3.5 for an illustration.

Remark 3.1.7 (ISRW as a mass-spreading process). The reader may prefer to let go of the connection with the random walk and view the ISRW purely as a mass-spreading process. Such a change of perspective would change nothing in our arguments. ♦

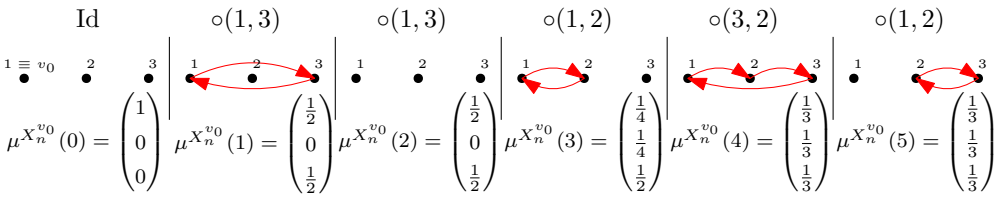


Figure 3.5: Example of an evolution of an ISRW on top of a CFDP with three elements starting from the identity permutation. The first row shows the transpositions that generate the next permutation. The second row is a graphical representation of the cycles of this permutation. The third row shows the evolution of the ISRW distribution, given that it started from the element 1.

§3.1.4 Preliminaries for Erdős-Rényi random graphs

The arguments in this chapter frequently make use of results on the structure of Erdős-Rényi random graphs. This section provides what is needed to state the main theorems in Section 3.1.6.

Definition 3.1.8 (Standard Erdős-Rényi multi-graph process). The *standard Erdős-Rényi multi-graph process* on n vertices is the discrete-time process $(G(n,t))_{t=0}^{t_{\max}}$ constructed as follows:

(a) $G(n,0)$ is the graph with n vertices and no edges.

(b) At each time $t \in \mathbb{N}_0$, pick an edge e_t uniformly at random from the $\binom{n}{2}$ possible edges, and let $G(n,t)$ be the graph obtained by adding e_t to $G(n,t-1)$.

Note that we do not allow for self-loops, but do allow for multiple edges. ■

Remark 3.1.9 (Versions and asymptotic equivalence). There are versions of the Erdős-Rényi multi-graph process that differ in how edges are deployed and whether or not multiple edges and self-loops are allowed. With respect to monotone properties, notably the expected size of connected components, the "random growth" $G(n,t)$ model described in Definition 3.1.8 is asymptotically equivalent to the "combinatorial" model $G(n,M)$ with $M = t$ edges at times $t = O(n)$, which in turn is asymptotically equivalent to the "bond percolation" model $G(n,p)$ with $p = M\binom{n}{2}^{-1}$. For details, see [62, Sections 1.1, 1.3]. Since we work on time scales of order n, we will use this asymptotic equivalence without further notice. ♦

Definition 3.1.8 allows for some natural modifications, of which one is important for the study of CDP:

Definition 3.1.10 (Cycle-free Erdős-Rényi graph process). The *cycle-free Erdős-Rényi graph process* on n vertices is the graph process $(G(n,t))_{t=0}^{t_{\max}}$ starting from the empty graph with n vertices, such that at each time t an edge is added that is chosen uniformly at random from the set of edges that do not create a cycle, a multi-edge or a self-loop. Thus, $G(n,t)$ is a forest for all $0 \leq t \leq t_{\max}$. ■

To understand the typical evolution of CDP, we make use of two *couplings*: one between CDP and cycle-free Erdős-Rényi graph processes, the other between cycle-free Erdős-Rényi graph processes and their standard counterparts. To explain how, we need to introduce three functions that describe key structural properties of these processes:

Definition 3.1.11 (Functions related to the structure of Erdős-Rényi random graphs).
(1) Define $\zeta\colon [0,\infty) \to [0,1)$ as $\zeta(u) = 0$ for $u \in [0, \frac{1}{2}]$ and as the unique positive solution of the equation $1 - \zeta(u) = \mathrm{e}^{-2u\zeta(u)}$ for $u \in (\frac{1}{2}, \infty)$. Note that ζ is non-decreasing and continuous on $[0,\infty)$, and analytic on $(\frac{1}{2},\infty)$.
(2) Define $\phi\colon [0,\infty) \to [0,1)$ as

$$\phi(v) = \int_0^v \mathrm{d}u\,[1-\zeta^2(u)], \qquad v \in [0,\infty). \tag{3.7}$$

Note that ϕ is strictly increasing and continuous on $[0,\infty)$, and hence has a well-defined inverse ϕ^{-1}. Furthermore, the function ϕ is properly normalised in the sense that $\phi(\infty) = 1$ (see Appendix B.2).
(3) Define $\eta\colon [0,1) \to [0,1)$ as

$$\eta(w) = \zeta(\phi^{-1}(w)), \qquad w \in [0,1). \tag{3.8}$$

■

The functions defined in Definition 3.1.11 are illustrated in Figure 3.6 and have the following interpretation:

(1) $\zeta(u)$ describes the expected size of the largest component of the Erdős-Rényi random graph at time un. For $u \in [0,\infty)$, denote by $|\mathscr{C}^{\mathrm{ER}}_{\max}(n,un)|$ the size of the largest connected component in the Erdős-Rényi random graph with n vertices and un edges, and $|\mathscr{C}^{\mathrm{ER}}_{\mathrm{sec}}(n,un)|$ the size of the second-largest connected component. Then, as $n \to \infty$,

$$\frac{|\mathscr{C}^{\mathrm{ER}}_{\max}(n,un)|}{n} \xrightarrow{\mathbb{P}} \zeta(u), \qquad \frac{|\mathscr{C}^{\mathrm{ER}}_{\mathrm{sec}}(n,un)|}{n} \xrightarrow{\mathbb{P}} 0. \tag{3.9}$$

(2) The function ϕ provides the link between the standard and the cycle-free Erdős-Rényi graph process (see Lemmas 3.2.4–3.2.5 below).

(3) $\eta(u)$ is the analogue of $\zeta(u)$ for the cycle-free Erdős-Rényi graph process at time un, $u \in [0,1]$ (see Lemma 3.2.6 below).

Note the change in behaviour of $\zeta, \phi, \phi^{-1}, \eta$ at $\frac{1}{2}$. Note that ϕ^{-1} blows up at 1.

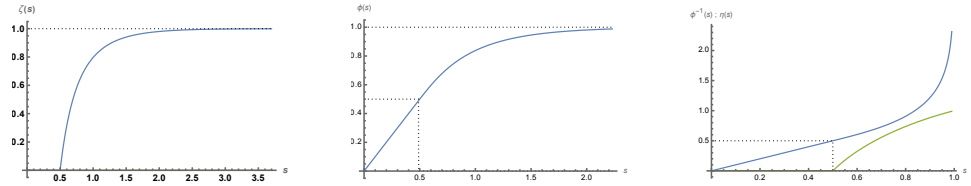


Figure 3.6: Graphs of the functions introduced in Definition 3.1.11: ζ, ϕ, respectively, ϕ^{-1} (upper curve), η (lower curve).

§3.1.5 Associated graph process

For any dynamic permutation starting from the identity permutation, define the associated graph process as follows:

Definition 3.1.12 (Graph process associated with Π_n). Let $\Pi_n = (\Pi_n(t))_{t=0}^{t_{\max}}$ with $t_{\max} \in \mathbb{N} \cup \{\infty\}$ be a dynamic permutation starting from the identity permutation. Construct the *associated graph process*, denoted by A_{Π_n}, as follows:

(a) At time $t = 0$, start with the empty graph on the vertex set $\mathcal{V} = [n]$.

(b) At times $t \in \mathbb{N}$, add the edge $\{a, b\}$, where a, b are such that $\Pi_n(t) = \Pi_n(t-1) \circ (a, b)$.

■

Associated graph processes were used in [135, 24] and follow-up articles to represent the evolution of a general dynamic permutation in terms of a dynamics generated by applying a single transposition at every time step.

A crucial role will be played by the first time when the support of the random walk distribution *intersects the largest connected component of the associated graph process*:

Definition 3.1.13 (Largest component of the associated graph process). Denote by $\mathscr{C}_{\max}(A_{\Pi_n}(t))$ the set of vertices in the largest connected component in the associated graph process at time t. If such a connected component is not unique, then take all the vertices in all the largest connected components. ■

Remark 3.1.14 (Possible non-uniqueness of the largest connected component). In situations where we employ Definition 3.1.13, the largest connected component is unique with high probability. Situations where it is not unique will be of no importance. ♦

Definition 3.1.15 (Drop-down time). Fix any $\varepsilon_n > 0$ such that $\varepsilon_n = \omega(n^{-1/3})$ and $\varepsilon_n = o(1)$ as $n \to \infty$. The *drop-down time* is defined as

$$T^{\Downarrow}_{n,v_0} = \inf\Big\{t > \tfrac{n}{2}[1+\varepsilon_n] \colon \operatorname{supp}(\mu^{n,v_0}(t)) \cap \mathscr{C}_{\max}(A_{\Pi_n}(t)) \neq \emptyset\Big\}. \tag{3.10}$$

■

Remark 3.1.16 (Drop-down time and hitting time of the largest permutation cycle). At first sight it might seem unintuitive that the time $T^{\Downarrow}_{n,v_0}$ from Definition 3.1.15

plays an important role. Given the diffusive nature of ISRW, an arguably more natural candidate would be the first time when the ISRW is supported on the largest *permutation cycle*. However, the above definition in terms of the associated graph process allows for a unified presentation of our results in different settings, even when the associated graph process at a single time *does not* provide all the information about the structure of permutation cycles. ♦

For CDP, the drop-down time is the first time when the cycle that contains v_0 merges with the largest cycle. For CFDP, however, this is not necessarily true because cycles fragment. We therefore define the drop-down time to be the first time when the random walk 'sees' the maximal component, see (3.10). Later, we will see that, in fact, afterwards the mass spreads over $\mathscr{C}_{\max}(A_{\Pi_n}(t))$ quickly.

Remark 3.1.17 (Properties of drop-down time). Clearly, $T^{\Downarrow}_{n,v_0}$ is random. However, if we condition on a particular realisation of Π_n, then $T^{\Downarrow}_{n,v_0}$ is a deterministic function of the starting point of the random walk. The role of ε_n is to ensure that $T^{\Downarrow}_{n,v_0}$ represents the first time in the *supercritical regime* when the *largest component* in the associated *Erdős-Rényi graph process* coincides with the support of the ISRW, see Section 3.2.1. The choice of ε_n ensures that the definition of the drop-down time *avoids the critical window*, which corresponds to $\frac{n}{2} + O(n^{2/3})$, yet covers the entire supercritical regime. ♦

§3.1.6 Main results

For convenience, we introduce the following shorthand notation:

Definition 3.1.18 (Total variation distance away from equilibrium). For $v \in [n]$, define

$$\mathcal{D}^{v_0}_n(t) = d_{\mathrm{TV}}\left(\mu^{n,v_0}(t), \mathsf{Unif}([n])\right), \qquad t \in \mathbb{N}_0. \tag{3.11}$$

■

Our main results are the following two theorems ($\xrightarrow{d}$ denotes convergence in distribution):

Theorem 3.1.19 (Mixing profile for ISRW on CDP).

(1) *Uniformly in* $v_0 \in [n]$,

$$\frac{T^{\Downarrow}_{n,v_0}}{n} \xrightarrow{d} s^{\Downarrow}, \tag{3.12}$$

where $s^{\Downarrow}$ *is the* $[0,1]$*-valued random variable with distribution (recall* (3.8)*)*

$$\mathbb{P}(s^{\Downarrow} \le s) = \eta(s), \qquad s \in [0,1]. \tag{3.13}$$

(2) *Uniformly in* $v_0 \in [n]$,

$$\left(\mathcal{D}^{v_0}_n(sn)\right)_{s\in[0,1]} \xrightarrow{d} \left(1 - \eta(s)\mathbb{1}_{\{s>s^{\Downarrow}\}}\right)_{s\in[0,1]} \quad \textit{in the Skorokhod } M_1\textit{-topology.} \tag{3.14}$$

Theorem 3.1.20 (Mixing profile for ISRW on CFDP).

(1) *Uniformly in* $v_0 \in [n]$,

$$\frac{T^{\Downarrow}_{n,v_0}}{n} \xrightarrow{d} u^{\Downarrow}, \tag{3.15}$$

where $u^{\Downarrow}$ *is the non-negative random variable with distribution (recall Definition 3.1.11(1))*

$$\mathbb{P}(u^{\Downarrow} \leq u) = \zeta(u), \qquad u \in [0, \infty). \tag{3.16}$$

(2) *Uniformly in* $v_0 \in [n]$,

$$(\mathcal{D}^{v_0}_n(un))_{u\in[0,\infty)} \xrightarrow{d} \big(1 - \zeta(u)\mathbb{1}_{\{u>u^{\Downarrow}\}}\big)_{u\in[0,\infty)} \quad \textit{in the Skorokhod } M_1\textit{-topology.} \tag{3.17}$$

The proofs of these theorems are given in Sections 3.2 and 3.3, respectively.

§3.1.7 Discussion

1. Despite the similarity of Theorems 3.1.19–3.1.20, the latter is *far more delicate*. For CDP, mixing is simply induced by the ISRW entering the ever-growing largest cycle. For CFDP, the presence of fragmentations breaks the direct link between the dynamic permutation and its associated graph process: *a single connected component may carry more than one permutation cycle*. Specifically, the largest component of the associated graph process carries a large number of permutation cycles and, at the drop-down time, the distribution of the ISRW is supported on only one of them. It is not a priori clear how many steps the dynamics needs to spread out the ISRW distribution over all the elements that lie on the largest component of the associated graph process. Therefore, a major hurdle in the proof of Theorem 3.1.20 is to show that such *local mixing* happens on time scale $o(n)$. We actually show a stronger statement, namely, that local mixing occurs on an arbitrarily small but diverging time scale (see Section 3.3.3 for details). The core of the proof is to show that on the largest component over time there is a diverging count of appearances of permutation cycles that span almost the entire largest component of the associated graph process.

2. Theorems 3.1.19–3.1.20 extend our earlier results for the total variation distance of a (non-backtracking) random walk on a configuration model subject to random rewirings [11]. There we assumed that *all the degrees are at least three*, which corresponds to a *supercritical* configuration model that with high probability is connected (see [79, Chapter 4]). Our model with evolving permutation cycles is closely related to the setting where *all the degrees are two*, which in turn corresponds to a special kind of configuration model that with high probability is disconnected (see Figure 3.7). In this setting, even small perturbations of the degree sequence can lead to significantly different behaviour (see [57] for details). In Appendix B.5 we comment further on the connection between permutations and degree-two graphs. More concretely, we show that in the setting of dynamic degree-two graphs with rewiring, we obtain an ISRW-mixing profile analogous to the one described in Theorem 3.1.20 (see Theorem B.5.3).

We stress that in the present work the starting configuration is fixed to be the identity permutation, which would correspond to a graph with only self-loops, whereas in our previous work the starting configuration was sampled from the configuration model.

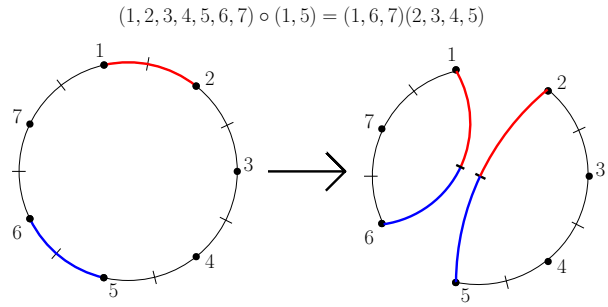


Figure 3.7: Dynamic permutations are similar to rewirings in the configuration model, where all degrees are two. Recall Example 3.1.2. Consider the permutation $\Pi(0) = (1, 2, 3, 4, 5, 6, 7)$*, which consists of a single cycle and corresponds to a degree-two graph that has a single connected component. Apply the transposition* $(1, 5)$ *to get a new permutation* $\Pi(1) = \Pi(0) \circ (1, 5) = (1, 6, 7)(2, 3, 4, 5)$*, which consists of two cycles and corresponds to a degree-two graph that has two connected components, obtained by sampling the edges* $(1, \Pi(0)(1)) = (1, 2)$ *and* $(5, \Pi(0)(5)) = (5, 6)$ *and rewiring them.*

3. The mixing profile in Theorems 3.1.19–3.1.20 is unusual: the total variation distance makes a *single jump* down from the value 1 to a value on a deterministic curve and subsequently follows this curve on its way down to 0. This jump, which is similar to a *one-sided cut-off*, occurs after a *random time*. The law of the drop-down time and the function describing the deterministic curve depend on the choice of dynamics.

4. The pathwise statements in part (2) of Theorems 3.1.19–3.1.20 imply the following pointwise statements ($\sim$ denotes equality in distribution):

$$\begin{aligned} \mathcal{D}_n^{v_0}(sn) &\xrightarrow{d} 1 - \eta(s)Y(s), \quad s \in [0,1], \quad \text{with } Y(s) \sim \mathsf{Bernoulli}(\eta(s)), \\ \mathcal{D}_n^{v_0}(sn) &\xrightarrow{d} 1 - \zeta(s)\bar{Y}(s), \quad s \in [0,\infty), \quad \text{with } \bar{Y}(s) \sim \mathsf{Bernoulli}(\zeta(s)). \end{aligned} \tag{3.18}$$

Through the function ϕ plotted in Figure 3.6, we can view the two mixing profiles as a continuous deformation of one another. Slower mixing for CFDP is intuitive: fragmentation slows down the mixing, while coagulation induces it.

5. Note the similarities between the mixing profiles described by Theorems 3.1.19–3.1.20. Both feature a single macroscopic jump at a random time to a deterministic curve that depends on the choice of the dynamics. We expect this type of behaviour to occur for any permutation dynamics whose associated graph process exhibits scaling behaviour similar to that of the Erdős-Rényi graph process. A class of graph processes that fits this criterion is the class of *Achlioptas processes with bounded-size rules* (see [127] or [139]).

6. We can formulate conjectures about finite-speed random walks as well. Settings where the random walk rate dominates are easy to handle. If the random walk is fast enough to ensure local mixing (e.g. $\gg n^2$ steps of the random walk occur for every step

of the random permutation), then our theorems should remain the same with negligible error terms. In this regime, the mixing is fully driven by the underlying geometry. However, once these rates are commensurate, we would have to deal with random walk distributions that are *partially mixed over cycles*, meaning that the distribution of the random walk would not be uniform over its supporting cycle before this cycle is affected by the permutation dynamics.

7. Dynamic permutations are a natural model for discrete dynamic random environments, which typically are *disconnected* but nonetheless allow for interaction between their constitutive elements. We believe this setting to be interesting for other stochastic processes on random graphs as well, such as the *voter model* or *the contact process*.

Organisation of the chapter. Section 3.2 starts by establishing a link between CDP and cycle-free Erdős-Rényi random graphs. A *coupling* construction is employed to describe the *cyclic structure* of a typical CDP. These results are used to prove Theorem 3.1.19. Section 3.3 deals with CFDP, where the main problem is that the associated graph process provides weaker control over permutation cycles than for CDP. After this discrepancy is settled, we employ arguments analogous to those in Section 3.2 to prove Theorem 3.1.20.

Appendices B.1–B.5 contain supplementary material that is not needed in Sections 3.2–3.3. Namely, Appendix B.1 shows that the ISRW arises as a fast-speed limit of the standard random walk. Appendix B.2 proves that the laws of the jump-down times in Theorems 3.1.19–3.1.20 are properly normalised. Appendix B.3 contains the key coupling that is used to study the cycle structure of CFDP, which is technical and of interest in itself. This coupling is needed in Section 3.3. Appendix B.4 contains a technical computation that is needed in Section 3.3 as well. Finally, Appendix B.5 elucidates the connection between random permutations and graphs with all the degrees equal to two and extends Theorem 3.1.20 to the setting of dynamic degree-two graphs.

§3.2 Coagulative dynamic permutations

In this section, we establish a link between dynamic permutations and evolving graphs. To do so, we couple a CDP with a cycle-free Erdős-Rényi graph process (Section 3.2.1), and couple the latter with the standard Erdős-Rényi graph process (Section 3.2.2) by making use of well-known results on the structure of connected components of Erdős-Rényi random graphs (recall Section 3.1.4). We use the couplings to prove Theorem 3.1.19 (Section 3.2.3).

§3.2.1 Representation via associated graph process

Note that for the dynamics generated by transpositions sampled uniformly at random from the set of all transpositions of n elements, the associated graph process is equal in distribution to the Erdős-Rényi process defined in Definition 3.1.8. In the setting of *coagulative* dynamic transpositions, this leads us to the following observation:

Lemma 3.2.1 (Representation of CDP as cycle-free graph process). *If Π_n is a CDP, then its associated graph process A_{Π_n} is the cycle-free Erdős-Rényi graph process defined in Definition 3.1.10.*

Proof. Recall that the change between two successive permutations in a CDP is generated by applying a *single* transposition. Furthermore, note that the only transpositions causing a split of a permutation cycle are the ones that transpose two elements from the same cycle. Recall Definition 3.1.12, and note that if $A_{\Pi_n}(t)$ is a forest, then its connected components correspond to cycles of $\Pi_n(t)$. Furthermore, observe that cycle-splitting transpositions correspond to edges that join two vertices from the same connected component. Thus, if $A_{\Pi_n}(t)$ is a forest, then any transposition causing a fragmentation of a permutation cycle corresponds to an edge that creates a cycle in the associated graph process.

Observe that the associated graph process always starts as a forest. Since fragmentations of permutation cycles are not allowed, there can be no edges that lead to graph cycles in the associated graph process. Since the associated graph process for a dynamic permutation with no constraints is the Erdős-Rényi graph process, the associated graph process for a CDP is the Erdős-Rényi graph process constrained to be a forest (see Definition 3.1.10). □

§3.2.2 Connected components of the cycle-free Erdős-Rényi graph process

• **Coupling of Erdős-Rényi graph processes.** We construct a coupling of the standard and the cycle-free Erdős-Rényi graph process that allows us to study the structure of the connected components of the cycle-free process.

Definition 3.2.2 (Coupling between cycle-free and standard Erdős-Rényi graph process). Let $G_n = (G_n(t))_{t\in\mathbb{N}_0}$ be the Erdős-Rényi graph process on $[n]$ defined in Definition 3.1.8, and denote the edge set of $G_n(t)$ by $\mathcal{E}_{G_n(t)}$. Based on G_n, construct a graph-valued process $F_n = (F_n(t))_{t\in\mathbb{N}_0}$ as follows:

(a) $F_n(0)$ is the empty graph with vertex set $[n]$.

(b) At times $t \in \mathbb{N}$, define $e^\star(t) = \mathcal{E}_{G_n(t)} \setminus \mathcal{E}_{G_n(t-1)}$, which is the edge added at time t to $G_n(t)$.

  (a) Construct the candidate graph at time t, defined as $F_n^\star(t) = (\mathcal{V}, \mathcal{E}_{F_n(t-1)} \cup \{e^\star(t)\})$.
  (b) If $F_n^\star(t)$ is a forest, then set $F_n(t) = F_n^\star(t)$.
  (c) Otherwise, set $F_n(t) = F_n(t-1)$.

Define the *effective time* $\tau_n(t)$ of the coupled process $(F_n(t))_{t\in\mathbb{N}_0}$ by setting $\tau_n(0) = 0$ and, recursively for $t \in \mathbb{N}$,

$$\tau_n(t) = \begin{cases} \tau_n(t-1) + 1, & \text{if } F_n(t) \neq F_n(t-1), \text{ i.e., the proposed edge has been accepted,} \\ \tau_n(t-1), & \text{if } F_n(t) = F_n(t-1), \text{ i.e., the proposed edge has been rejected.} \end{cases} \tag{3.19}$$

Note that $\tau_n(t)$ is a random variable because it is a function of a random graph process.

We suppress the dependence of $\tau_n(t)$ on F_n, since we will never work with more than one set of coupled processes at a time. ■

Remark 3.2.3 (Relation between F_n and A_{Π_n}). By the definition of the coupling, if there are edge-rejections at times $\{t, t+1, t+2, \dots, t+k\}$, then a string of $k+1$ copies of the same graph is observed in F_n, i.e., $F_n(t-1) = F_n(t) = F_n(t+1) = \cdots = F_n(t+k)$. On the other hand, the associated graph process A_{Π_n} is by construction a sequence of graphs such that no two graphs are the same. To recover A_{Π_n} from F_n, from every string of copies of the same graph choose only one copy of that graph. ♦

The reason why this construction is useful to control the connected components of the cycle-free Erdős-Rényi graph process is stated in the following lemma:

Lemma 3.2.4 (Connected components of F_n). *Let H be a graph with vertex set $\mathcal{V}$, and define $\boldsymbol{\mathcal{CC}}(H)$ to be the partition of $\mathcal{V}$ induced by the connected components of H. Let G_n, F_n be as in Definition 3.2.2. Then, at every time $t \in \mathbb{N}_0$, $\boldsymbol{\mathcal{CC}}(G_n(t)) = \boldsymbol{\mathcal{CC}}(F_n(t))$.*

Proof. Note that any edge creating a cycle does not influence the size of the connected components. □

• **Effective time.** To use the above observation, we need to control the effective time $\tau_n(t)$. The following lemma shows that with high probability and after scaling by $1/n$, there is a simple relation between the standard time t and the effective time $\tau_n(t)$:

Lemma 3.2.5 (Effective time of a cycle-free Erdős-Rényi graph process). *Let G_n, F_n and τ_n be as in Definition 3.2.2, and ϕ as in Definition 3.7. Then, for any for $u \in [0, \infty)$,*

$$\frac{\tau_n(un)}{n} \xrightarrow{\mathbb{P}} \phi(u). \tag{3.20}$$

Proof. The proof of Lemma 3.2.5 consists of two separate lines of argument. First, we show that the left-hand side in (3.20) concentrates around a deterministic quantity. Afterwards, the value of this quantity is computed.

Part 1: Concentration of the associated martingale. Observe that

$$\tau_n(t) = t - \sum_{s=0}^{t} \mathbb{1}_{R(s)}, \qquad t \in \mathbb{N}_0, \tag{3.21}$$

where $\mathbb{1}_{R(s)}$ is the edge-rejection indicator at time s. Let $\mathcal{F} = (\mathcal{F}_n(t))_{t\in\mathbb{N}_0}$ with $\mathcal{F}_n(t) = \sigma((G_n(q))_{q=0}^t)$ be the natural filtration with respect to the Erdős-Rényi graph process. By the construction of the coupling, a rejection occurs whenever there is an edge that creates a cycle within a connected component. Therefore

$$\mathbb{1}_{R(0)} = 0, \qquad \big[\mathbb{1}_{R(s)} \mid \mathcal{F}_n(s-1)\big] \sim \mathsf{Bernoulli}(p_s), \qquad s \in \mathbb{N}, \tag{3.22}$$

where the success probabilities are given by

$$p_s = \sum_{\mathscr{C} \in \boldsymbol{\mathcal{CC}}(G_n(s-1))} \frac{|\mathscr{C}|(|\mathscr{C}|-1)}{n(n-1)}, \qquad s \in \mathbb{N}, \tag{3.23}$$

i.e., the number of edges that can join two vertices from the same connected component at time $s-1$ divided by the total number of edges. Introduce the shorthand notation

$$\mathbb{E}_t[\cdot] = \mathbb{E}\left[\cdot \mid \mathcal{F}_n(t)\right], \tag{3.24}$$

and define two sequences of random variables $(D_t)_{t\in\mathbb{N}_0}$ and $(S_t)_{t\in\mathbb{N}_0}$ such that

$$\begin{aligned} S_0 &= D_0 = 0, \\ D_t &= \mathbb{1}_{R(t)} - \mathbb{E}_{t-1}\left[\mathbb{1}_{R(t)}\right], \qquad t \in \mathbb{N}, \\ S_t &= \sum_{s=0}^{t} D_s = \sum_{s=0}^{t} \mathbb{1}_{R(s)} - \sum_{s=0}^{t} \mathbb{E}_{s-1}\left[\mathbb{1}_{R(s)}\right], \qquad t \in \mathbb{N}. \end{aligned} \tag{3.25}$$

Note that, for any $t \in \mathbb{N}_0$,

$$\begin{aligned} \mathbb{E}[S_t] &\le t < \infty, \\ \mathbb{E}_t\left[S_{t+1}\right] &= \mathbb{E}_t\left[\mathbb{1}_{R(t+1)} - \mathbb{E}_t\left[\mathbb{1}_{R(t+1)}\right]\right] + \mathbb{E}_t\left[S_t\right] = S_t, \\ |S_t - S_{t-1}| &= |D_t| \le 1. \end{aligned} \tag{3.26}$$

Hence, $(S_t)_{t\in\mathbb{N}_0}$ is a martingale with bounded differences with respect to the natural filtration of the Erdős-Rényi graph process. Using the Azuma-Hoeffding inequality, we can estimate

$$\mathbb{P}\left(|S_t| \ge \varepsilon\right) \le 2\exp\left(-\frac{\varepsilon^2}{2t}\right). \tag{3.27}$$

Pick $t = un$, $u \in [0,\infty)$, and $\varepsilon_n = n^{\frac{1+\delta}{2}}$, $c > 0$, $\delta \in (0,1)$. Introduce the event

$$\Xi(un) = \{|S_{un}| < n^{\frac{1+\delta}{2}}\}. \tag{3.28}$$

By (3.27),

$$\mathbb{P}\left(\Xi^{\mathrm{c}}(un)\right) \le 2\exp\left(-\frac{c^2 n^{\delta}}{2u}\right) = o(1) \tag{3.29}$$

and hence $\mathbb{P}\left(\Xi(un)\right) = 1 - o(1)$. By the definition of S_t in (3.25), we see that, on the event $\Xi(un)$,

$$\mathbb{1}_{\Xi(un)}\left[\frac{1}{n}\sum_{s=0}^{un}\mathbb{1}_{R(s)} - \frac{1}{n}\sum_{s=0}^{un}\mathbb{E}_{s-1}\left[\mathbb{1}_{R(s)}\right]\right] = \mathbb{1}_{\Xi(un)}\,o(1), \qquad \text{with } |o(1)| \le n^{\frac{1+\delta}{2}}, \tag{3.30}$$

which establishes the concentration of $\tau_n(un)/n$ as $n \to \infty$.

Part 2: Computation of limit. We compute

$$\frac{1}{n}\sum_{s=0}^{un}\mathbb{1}_{R(s)} = \frac{\left(\mathbb{1}_{\Xi(un)} + \mathbb{1}_{\Xi^{\mathrm{c}}(un)}\right)}{n}\sum_{s=0}^{un}\mathbb{1}_{R(s)} \tag{3.31}$$

$$= \frac{1}{n}\left[\mathbb{1}_{\Xi(un)}\left(\sum_{s=0}^{un}\mathbb{E}_{s-1}\left[\mathbb{1}_{R(i)}\right] + o(n)\right) + \mathbb{1}_{\Xi^{\mathrm{c}}(un)}\sum_{s=0}^{un}\mathbb{1}_{R(s)}\right]. \tag{3.32}$$

1. Observe that

$$\frac{1}{n}\mathbb{1}_{\Xi^{\mathrm{c}}(un)}\sum_{s=0}^{un}\mathbb{1}_{R(s)} \xrightarrow{\mathbb{P}} 0, \tag{3.33}$$

because $\sum_{s=0}^{un}\mathbb{1}_{R(s)} \le un$ and $\mathbb{1}_{\Xi^{\mathrm{c}}(un)} \xrightarrow{\mathbb{P}} 0$. To understand the first summand in (3.32), we need to introduce another event. Fix a sequence ε_n such that $\varepsilon_n = o(1)$ and $\varepsilon_n = \omega(n^{-1/3})$. For any $n \in \mathbb{N}$, $t \in \mathbb{N}_0$, define the Erdős-Rényi *typicality* event

$$\begin{aligned}\Omega_n(un) = \big\{|\mathscr{C}^{\mathrm{ER}}_{\max}(n, vn)| \in (n[\zeta(v) - \varepsilon_n], n[\zeta(v) + \varepsilon_n])\ \forall\, v \in [\tfrac{1}{2}, u]\big\} \\ \cap \big\{0 \le \max_{s\le un}|\mathscr{C}^{\mathrm{ER}}_{\mathrm{sec}}(n, s)| \le n\varepsilon_n\big\}.\end{aligned} \tag{3.34}$$

Then we can write

$$\mathbb{1}_{\Xi(un)}\sum_{s=0}^{un}\mathbb{E}_{s-1}\left[\mathbb{1}_{R(s)}\right] = \mathbb{1}_{\Xi(un)}\left(\mathbb{1}_{\Omega_n(un)} + \mathbb{1}_{\Omega_n^{\mathrm{c}}(un)}\right)\sum_{s=0}^{un}\mathbb{E}_{s-1}\left[\mathbb{1}_{R(s)}\right]. \tag{3.35}$$

Again, we see that

$$\frac{1}{n}\mathbb{1}_{\Omega_n^{\mathrm{c}}(un)}\sum_{s=0}^{un}\mathbb{E}_{s-1}\left[\mathbb{1}_{R(s)}\right] \xrightarrow{\mathbb{P}} 0, \tag{3.36}$$

because $\sum_{s=0}^{un}\mathbb{E}_{s-1}[\mathbb{1}_{R(s)}] \le un$ and $\mathbb{1}_{\Omega_n^{\mathrm{c}}(un)} \xrightarrow{\mathbb{P}} 0$.

2. It remains to compute $\frac{1}{n}\mathbb{1}_{\Xi(un)}\mathbb{1}_{\Omega_n(un)}\sum_{s=0}^{un}\mathbb{E}_{s-1}[\mathbb{1}_{R(s)}]$, which is the only term that will be non-zero after we take the limit $n \to \infty$ in (3.31). Recall that $\mathbb{E}_{s-1}[\mathbb{1}_{R(s)}] \sim \mathsf{Bernoulli}(p_s)$, where the success probabilities p_s were introduced in (3.23).

Since $|\mathcal{V}| = n$, we have the following bounds:

$$p_s = \sum_{\mathscr{C}\in\boldsymbol{\mathcal{CC}}(G_n(s-1))}\frac{|\mathscr{C}|(|\mathscr{C}|-1)}{n(n-1)} \le \begin{cases}\frac{|\mathscr{C}^{\mathrm{ER}}_{\max}(n,n/2)|}{n}, & 0 \le s \le \frac{1}{2}n,\\ \frac{|\mathscr{C}^{\mathrm{ER}}_{\max}(n,s)|^2}{n^2} + \max\limits_{0\le s\le un}\frac{|\mathscr{C}^{\mathrm{ER}}_{\mathrm{sec}}(n,s)|}{n} & \frac{1}{2}n < s < un,\end{cases} \tag{3.37}$$

where the last bound is uniform in $s \le un$. The first line, for times below $n/2$, holds since

$$\sum_{\mathscr{C}\in\boldsymbol{\mathcal{CC}}(G_n(s-1))}\frac{|\mathscr{C}|(|\mathscr{C}|-1)}{n(n-1)} \le \frac{|\mathscr{C}^{\mathrm{ER}}_{\max}(n,s)|}{n} \le \frac{|\mathscr{C}^{\mathrm{ER}}_{\max}(n,n/2)|}{n}. \tag{3.38}$$

The second line separates the contribution of the maximal component and all the other components, and bound the non-maximal component similarly as in the first line.

In the *supercritical* regime, we separately describe the contribution of the unique largest component and give an upper bound only on the probability of rejection due to the other components. On the Erdős-Rényi typicality event $\Omega_n(un)$ (recall (3.34)), the size of all the components in the subcritical and critical regime and all the components but the largest one in the supercritical regime can be uniformly bounded by $|\mathscr{C}| \le Zn^{2/3}$, where Z is a positive random variable. From (3.22), (3.34) and (3.37), it follows that, on the event $\Omega_n(un)$,

$$\mathbb{1}_{\Omega_n(un)}\sum_{s=0}^{un}\mathbb{E}_{s-1}\left[\mathbb{1}_{R(s)}\right] = \mathbb{1}_{\Omega_n(un)}\sum_{s=0}^{un}p_s\begin{cases}\le u|\mathscr{C}^{\mathrm{ER}}_{\max}(n,n/2)| \le \varepsilon_n u, & 0 \le u \le \frac{1}{2},\\ = \sum_{s=\frac{n}{2}}^{un}(\zeta(i/n)+\varepsilon_n)^2 + \mathcal{R}(un), & u > \frac{1}{2},\end{cases} \tag{3.39}$$

where we use that $\zeta(\frac{1}{2}) = 0$. Since $|\mathscr{C}^{\mathrm{ER}}_{\max}(n, n/2)| \leq \varepsilon_n n$ on $\Omega_n(un)$, the remainder term $\mathcal{R}(un)$ can be bounded as (recall (3.37))

$$\mathcal{R}(un) \leq un \max_{0 \leq s \leq un} \frac{|\mathscr{C}^{\mathrm{ER}}_{\mathrm{sec}}(n, s)|}{n} \leq \varepsilon_n un = o(n). \tag{3.40}$$

3. Before wrapping up, let us note that

$$\frac{o(n) + \sum_{s=0}^{un} \zeta(s/n)^2}{n} \to \int_0^u \mathrm{d}v\, \zeta^2(v), \tag{3.41}$$

because ζ^2 is bounded and hence Riemann integrable over compact intervals $[0, u]$, and $\frac{1}{n}\sum_{s=0}^{un} \zeta(s/n)^2$ is a Riemann sum of ζ^2 over a regular partition of $[0, u]$ into subintervals of length $1/n$. This allows us to finish our previous computation, namely,

$$\begin{aligned}
&\frac{1}{n}\sum_{s=0}^{un} \mathbb{1}_{R(s)} \\
&\quad = \frac{1}{n}\left[\mathbb{1}_{\Xi(un)}\left(\mathbb{1}_{\Omega_n(un)} + \mathbb{1}_{\Omega_n^{\mathrm{c}}(un)}\right)\left(\sum_{s=0}^{un} \mathbb{E}_{s-1}\left[\mathbb{1}_{R(s)}\right] + o(n)\right) + \mathbb{1}_{\Xi^{\mathrm{c}}(un)} \sum_{s=0}^{un} \mathbb{1}_{R(s)}\right] \\
&\quad = \frac{1}{n}\left(o(n) + \mathbb{1}_{\Xi(un)}\mathbb{1}_{\Omega_n(un)} \sum_{s=0}^{un} \mathbb{E}_{s-1}\left[\mathbb{1}_{R(s)}\right]\right) \xrightarrow{d} \int_0^u \mathrm{d}v\, \zeta^2(v) = u - \phi(u)
\end{aligned} \tag{3.42}$$

(recall (3.7)), from which the desired result follows. □

• **Mapping between times.** The main purpose of Lemma 3.2.5 is to show that on time scales of order n there is a function ϕ (recall Definition 3.1.11) capturing the correspondence, in the limit as $n \to \infty$, between the times at which the standard Erdős-Rényi graph process and its cycle-free counterpart have certain quantities distributed equally, notably, the sizes of their connected components. Since ϕ is strictly monotone, it admits a proper inverse, which allows us to relate the cycle-free graph process to the standard Erdős-Rényi graph growth process.

• **Largest component.** To conclude the analysis of the cycle-free graph process, we combine the above results to obtain a characterisation of the largest component of the cycle-free Erdős-Rényi graph process:

Lemma 3.2.6 (Size of the largest component). *For $s \in [0, 1]$, let $|\mathscr{C}^{\mathrm{cfER}}_{\max}(n, sn)|$ be the size of the largest component of the cycle-free Erdős-Rényi graph process on n vertices at time sn. Then*

$$\frac{|\mathscr{C}^{\mathrm{cfER}}_{\max}(n, sn)|}{n} \xrightarrow{\mathbb{P}} \eta(s), \qquad \eta(s) = \zeta(\phi^{-1}(s)). \tag{3.43}$$

Proof. In Lemma 3.2.4 we have established that for every $n \in \mathbb{N}$ the connected components of the cycle-free graph $G^{\mathrm{cf}}_n(t)$ correspond exactly to the connected components of the standard Erdős-Rényi graph process at some random time, denoted

by $\tau_n^{-1}(t)$. In Lemma 3.2.5 we have established that $\tau_n^{-1}(t) = n\phi^{-1}(t/n) + o_{\mathbb{P}}(n)$ and $|\mathscr{C}_{\max}^{\mathrm{ER}}(n, sn)| = n\zeta(s) + o_{\mathbb{P}}(n)$. Hence

$$|\mathscr{C}_{\max}^{\mathrm{cfER}}(n, sn)| \stackrel{d}{=} |\mathscr{C}_{\max}^{\mathrm{ER}}(n, n(\phi^{-1}(s) + o_{\mathbb{P}}(1)))| = n\zeta(\phi^{-1}(s) + o_{\mathbb{P}}(1)) + o_{\mathbb{P}}(n), \quad (3.44)$$

from which the claim follows. □

§3.2.3 Drop-down time and mixing profile

As stated in Theorem 3.1.19, for CDP the mixing profile exhibits a cut-off at a *random* time. From that moment onwards, the total variation distance follows a *deterministic* curve that is related to the typical structure of CDP. The following lemma gives the distribution of the *drop-down* time and settles Theorem 3.1.19(1):

Lemma 3.2.7 (Limit distribution of drop-down time for ISRW on CDP). *Recall $T_{n,v_0}^{\Downarrow}$ from Definition 3.1.15. There exists a $[0,1]$-valued random variable $s^{\Downarrow}$ with a distribution function $\mathbb{P}(s^{\Downarrow} \leq s) = \eta(s)$, $s \in [0,1]$, such that*

$$\frac{T_{n,v_0}^{\Downarrow}}{n} \stackrel{d}{\to} s^{\Downarrow}. \quad (3.45)$$

Proof. By the arguments in the proof of Lemma 3.2.1, the sizes of the connected components of the cycle-free associated graph process exactly correspond to the sizes of the permutation cycles of the CDP at a given time t. Therefore we must study the probability that a uniform vertex lies on the largest component of a cycle-free Erdős-Rényi graph process.

Let $\mathbb{P}_n$ denote the law of CDP on $[n]$ and $\mathbb{P}_n^{\mathrm{cfER}}$ the law of the associated graph process, which is a cycle-free Erdős-Rényi graph process (recall Definition 3.1.12 and Lemma 3.2.1). Fix a sequence ε_n such that $\varepsilon_n = o(1)$ and $\varepsilon_n = \omega(n^{-1/3})$, and for $n \in \mathbb{N}$, $t \geq 0$ define the *typicality* event

$$\Omega_n^{\mathrm{cfER}}(t) = \left\{|\mathscr{C}_{\max}^{\mathrm{cfER}}(n,t)| \in n(\eta(t/n) - \varepsilon_n, \eta(t/n) + \varepsilon_n,)\right\} \cap \left\{|\mathscr{C}_{\mathrm{sec}}^{\mathrm{cfER}}(n,t)| \leq n\varepsilon_n\right\}. \quad (3.46)$$

For $s \in [0, \frac{1}{2})$, we have $\mathbb{P}_n(T_{n,v_0}^{\Downarrow} \leq sn) = 0$ by the definition of $T_{n,v_0}^{\Downarrow}$. For $s \in [\frac{1}{2}, 1]$, by Lemma 3.2.1,

$$\begin{aligned}\mathbb{P}_n(T_{n,v_0}^{\Downarrow} \leq sn) &= \mathbb{P}_n^{\mathrm{cfER}}\big(v_0 \in \mathscr{C}_{\max}^{\mathrm{cfER}}(sn)\big) \\ &= \mathbb{P}_n^{\mathrm{cfER}}\big(\{v_0 \in \mathscr{C}_{\max}^{\mathrm{cfER}}(sn)\} \cap \Omega_n^{\mathrm{cfER}}(sn)\big) + \mathbb{P}_n^{\mathrm{cfER}}\big(\{v_0 \in \mathscr{C}_{\max}^{\mathrm{cfER}}(sn)\} \cap [\Omega_n^{\mathrm{cfER}}(sn)]^{\mathrm{c}}\big).\end{aligned} \quad (3.47)$$

Since the event $\Omega_n^{\mathrm{cfER}}(sn)$ occurs with probability $1-o(1)$ and $\mathbb{P}_n^{\mathrm{cfER}}(\{v_0 \in \mathscr{C}_{\max}^{\mathrm{cfER}}(sn)\} \cap \Omega_n^{\mathrm{cfER}}(sn)) = \eta(s) + o(1)$, we see that, for any $s \in [0,1]$,

$$\mathbb{P}_n(T_{n,v_0}^{\Downarrow} \leq sn) \stackrel{n\to\infty}{\to} \eta(s). \quad (3.48)$$

Since η is continuous, non-negative and non-decreasing on $[0,1]$ such that $\int_0^1 \mathrm{d}u\, \eta(u) = 1$, (3.48) defines a proper distribution function (recall Definition 3.1.11). See Appendix B.2 for a detailed computation. □

With the above results in hand, we are ready to prove the pointwise version of Theorem 3.1.19(2), characterising the mixing profile of ISRW on CDP:

Lemma 3.2.8 (Pointwise limit of mixing profile for ISRWs on CDP). *Uniformly in* $v_0 \in [n]$,

$$\mathcal{D}_n^{v_0}(sn) \xrightarrow{d} 1 - \eta(s)Y(s), \qquad s \in [0,1), \tag{3.49}$$

where $Y(s) \sim \mathsf{Bernoulli}(\eta(s))$.

Proof. Given a permutation π, let

$$|\gamma_{\max}(\pi)| = \max\{|\gamma_v(\pi)| \colon\ v \in [n]\} \tag{3.50}$$

denote the size of the largest cycle of π. For every $n \in \mathbb{N}$,

$$\mathcal{D}_n^{v_0}(sn) = d_{\mathrm{TV}}\Big(\mathsf{Unif}([n]), \mathsf{Unif}([\gamma_{\max}(sn)])\Big) = 1 - \frac{|\gamma_{sn}(v_0)|}{n} \tag{3.51}$$

by Definition 3.1.6 and the definition of total variation distance. Using Lemma 3.2.1, we see that

$$\begin{aligned}\mathcal{D}_n^{v_0}(sn) &\stackrel{d}{=} 1 - \frac{1}{n}\left(|\mathscr{C}_{v_0}(sn)|\mathbb{1}_{\Omega_n^{\mathrm{cfER}}(sn)} + |\mathscr{C}_{v_0}(sn)|\mathbb{1}_{[\Omega_n^{\mathrm{cfER}}(sn)]^{\mathrm{c}}}\right)\\ &= 1 - \frac{1}{n}\left(|\mathscr{C}_{v_0}(sn)|\mathbb{1}_{\Omega_n^{\mathrm{cfER}}(sn)}\left(\mathbb{1}_{\{T^{\Downarrow}_{n,v_0} \le sn\}} + \mathbb{1}_{\{T^{\Downarrow}_{n,v_0} > sn\}}\right) + |\mathscr{C}_{v_0}(sn)|\mathbb{1}_{[\Omega_n^{\mathrm{cfER}}(sn)]^{\mathrm{c}}}\right).\end{aligned} \tag{3.52}$$

Standard results for the size of Erdős-Rényi connected components (recall (3.9)) imply that

$$\mathbb{1}_{[\Omega_n^{\mathrm{cfER}}(sn)]^{\mathrm{c}}} \xrightarrow{\mathbb{P}} 0, \qquad \mathbb{1}_{\Omega_n^{\mathrm{cfER}}(sn)} \xrightarrow{\mathbb{P}} 1, \qquad \frac{|\mathscr{C}_{v_0}(sn)|\mathbb{1}_{\Omega_n^{\mathrm{cfER}}(sn)}\mathbb{1}_{\{T^{\Downarrow}_{n,v_0} > sn\}}}{n} \xrightarrow{\mathbb{P}} 0, \tag{3.53}$$

where the last limit follows from the fact that at times $0 \le sn < T^{\Downarrow}_{n,v_0}$ the initial vertex v_0 lies on a non-largest component, and hence the numerator scales as $o(n)$. Similarly, the size of the largest Erdős-Rényi component has a well-known limit (recall (3.9)), on the event $\{T^{\Downarrow}_{n,v_0} < sn\}$, namely,

$$\frac{|\mathscr{C}_{v_0}(sn)|\mathbb{1}_{\Omega_n^{\mathrm{cfER}}(sn)}}{n} \xrightarrow{\mathbb{P}} \eta(s). \tag{3.54}$$

Finally, the only random variable that converges to a non-degenerate random variable is

$$\mathbb{1}_{\{T^{\Downarrow}_{n,v_0} \le sn\}} \xrightarrow{d} Y(s), \qquad Y(s) \sim \mathsf{Bernoulli}(\eta(s)), \tag{3.55}$$

which follows from Lemma 3.2.7. Note that $\mathcal{D}_n^{v_0}(sn)$ is a sum of several random variables, and we established the convergence of each in (3.52)–(3.55). Hence, the claim follows via Slutsky's theorem. □

To conclude this section, we use Lemma 3.2.8 to prove the pathwise convergence part of Theorem 3.1.19:

Proof of Theorem 3.1.19(2). Observe that, for every $n \in \mathbb{N}$, any realisation of $\mathcal{D}_n^{v_0}(\cdot)$ is a monotone càdlàg path on the compact set $[0,1]$. In this special situation, the pointwise convergence proven in Lemma 3.2.8 implies pathwise convergence in the Skorokhod M_1-topology. For details, see [151, Corollary 12.5.1]. □

§3.3 Coagulative-fragmentative dynamic permutations

In Section 3.2, for CDP it took effort to control the structure of the associated graph process, while the mixing profile was obtained via an easy argument. For CFDP the opposite is true: the associated graph process, introduced in Section 3.2.1, is the Erdős-Rényi graph process defined in Definition 3.1.8 (which is one of the key facts used in [135]), while the link between the cycles of the underlying permutation and the connected components of the associated graph process is far less clear. Indeed, each non-tree connected component of the associated graph process may represent *multiple* permutation cycles, which brings a *substructure* into the problem that needs to be controlled. Moreover, it is not a priori clear whether or not this substructure influences the mixing profile, since *immediately* after the drop-down time the distribution of the ISRW is uniform over a component that spans only a random fraction of the largest component of the associated graph process.

The key result in this section is that ISRW on CFDP exhibits *fast local mixing on the largest component of the associated graph process upon drop-down.* After scaling, this leads to results that are qualitatively similar to those obtained for CDP, namely, the occurrence of a *single* jump in the total variation distance, from 1 to a deterministic value on a curve related to the largest component of the associated graph process, at a *random time* whose distribution is again connected to the largest component of the associated graph process. The scaled time now takes values in $[0,\infty)$ instead of $[0,1]$.

In Section 3.3.1 we identify the drop-down time and prove Theorem 3.1.20(1). In Section 3.3.2 we show that the support of ISRW lies on a single permutation cycle before the drop-down time. In Section 3.3.3 we prove fast local mixing after the drop-down time. In Section 3.3.4 we identify the mixing profile and prove Theorem 3.1.20(2).

Remark 3.3.1 (Permutation elements and graph vertices representing them). Throughout this section we will (with a slight abuse of notation) identify the vertices in the associated graph process with the permutation elements they represent. ♦

§3.3.1 Drop-down time

Recall that the central object for the identification of the limit distribution of the drop-down time for CDP in Section 3.2.3 was the function η (recall Definition 3.1.11), which describes the size of the largest component in the cycle-free Erdős-Rényi graph process. In the setting of CFDP, we formulate a result for $T_{n,v_0}^{\Downarrow}$ analogous to Lemma 3.2.7, with

the role of η taken over by ζ, which describes the size of the largest component in the standard Erdős-Rényi graph process:

Lemma 3.3.2 (Limiting distribution of drop-down time for ISRW on CFDP). *Recall $T^{\Downarrow}_{n,v_0}$ from Definition 3.1.15. There exists a $[0,\infty)$-valued random variable $u^{\Downarrow}$ with distribution function $\mathbb{P}(u^{\Downarrow} \leq u) = \zeta(u)$, $u \in [0,\infty)$, such that*

$$\frac{T^{\Downarrow}_{n,v_0}}{n} \xrightarrow{d} u^{\Downarrow}. \tag{3.56}$$

Proof. The proof is the same as that of Lemma 3.2.7, but uses the laws of CFDP and its associated graph processes, and uses ζ in place of η. □

§3.3.2 Drop-down in a single permutation cycle

In principle, it could happen that the ISRW support has experienced fragmentation before the drop-down time, which would significantly complicate our analysis. The main point of this section is to show that, with high probability, this *does not* occur.

Lemma 3.3.3 (ISRW support lies on a single permutation cycle before $T^{\Downarrow}_{n,v_0}$). *Fix $\varepsilon_n > 0$ such that $\varepsilon_n = \omega(n^{-1/3})$ and $\varepsilon_n = o(1)$ as $n \to \infty$. Let $\Omega^{(\mathrm{SC})}(t)$ denote the event that the support of the ISRW at time t lies on a single permutation cycle. Then, uniformly in v_0 and $t = cn$ with $c \in (\frac{1}{2}, \infty)$,*

$$\mathbb{P}\left(\Omega^{(\mathrm{SC})}(t) \mid T^{\Downarrow}_{n,v_0} \geq t\right) = 1 - o(1). \tag{3.57}$$

Proof. Recall the associated graph process introduced in Definition 3.1.12, and the fact that the associated graph process of CFDP is equal in distribution to the standard Erdős-Rényi graph process. As explained in the proof of Lemma 3.2.1, tree components in the associated graph process correspond to permutation cycles that have never experienced fragmentation. The idea of the proof is to show that, conditionally on the event $\{T^{\Downarrow}_{n,v_0} \geq t\}$, the event $\Omega^{\mathrm{tree}}(t)$ that ISRW at time t is supported on a single tree-component in the associated graph process occurs with high probability. Observe that $\Omega^{\mathrm{tree}}(t) \subseteq \Omega^{(\mathrm{SC})}(t)$. First we condition on the event $\{T^{\Downarrow}_{n,v_0} > t\}$. Afterwards, we extend to the event $\{T^{\Downarrow}_{n,v_0} \geq t\}$.

Recall that in Lemma 3.3.2 we identified the limiting distribution of $T^{\Downarrow}_{n,v_0}/n$. Since

$$\begin{aligned}\mathbb{P}\left(\Omega^{\mathrm{tree}}(t) \mid T^{\Downarrow}_{n,v_0} > t\right) &= 1 - \mathbb{P}\left(\left[\Omega^{\mathrm{tree}}(t)\right]^{\mathrm{c}} \mid T^{\Downarrow}_{n,v_0} > t\right) \\ &= 1 - \frac{\mathbb{P}\left(\left[\Omega^{\mathrm{tree}}(t)\right]^{\mathrm{c}} \cap \{T^{\Downarrow}_{n,v_0} > t\}\right)}{\mathbb{P}(T^{\Downarrow}_{n,v_0} > t)},\end{aligned} \tag{3.58}$$

and the denominator is bounded away from 0 (recall Lemma 3.3.2), it suffices to show that $\mathbb{P}([\Omega^{\mathrm{tree}}(t)]^{\mathrm{c}} \cap \{T^{\Downarrow}_{n,v_0} > t\}) = o(1)$. By the law of total probability, we can take the sum over all possible realisations of the underlying dynamics to obtain

$$\mathbb{P}\left(\left[\Omega^{\mathrm{tree}}(t)\right]^{\mathrm{c}} \cap \{T^{\Downarrow}_{n,v_0} > t\}\right) = \mathbb{E}\left[\mathbb{P}\left(\left[\Omega^{\mathrm{tree}}(t)\right]^{\mathrm{c}} \cap \{T^{\Downarrow}_{n,v_0} > t\} \mid (\Pi_n(t))^t_{s=0}\right)\right]. \tag{3.59}$$

By [87, Theorem 5.10], with high probability the connected components of the associated graph process at time t consist of the *unique largest component, unicyclic connected*

components and trees. By Definition 3.1.15, conditionally on $\{T^{\Downarrow}_{n,v_0} > t\}$, the support of the ISRW in the associated graph process *does not* lie on the largest component. It therefore lies, with high probability, on either a unicyclic component or a tree. Denote by $N^{\mathrm{uc}}(t)$ the number of vertices in an Erdős-Rényi graph process that are in unicyclic connected components at time t, and recall from (3.9) that $\mathscr{C}^{\mathrm{ER}}_{\max}(n,t)$ denotes the size of the largest component of an Erdős-Rényi graph on n vertices with t edges. It follows that

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{P}\left(\left[\Omega^{\mathrm{tree}}(t)\right]^{\mathrm{c}} \cap \{T^{\Downarrow}_{n,v_0} > t\} \;\middle|\; (\Pi_n(t))^t_{s=0}\right)\right] \\
&= \mathbb{E}\left[\mathbb{P}\left(\left[\Omega^{\mathrm{tree}}(t)\right]^{\mathrm{c}} \cap \{v_0 \notin \mathscr{C}_{\max}(A_{\Pi_n}(t))\} \;\middle|\; (\Pi_n(t))^t_{s=0}\right)\right] + o(1) \\
&\leq \mathbb{E}\left[\min\left(\frac{N^{\mathrm{uc}}(t)}{n - |\mathscr{C}^{\mathrm{ER}}_{\max}(n,t)|}, 1\right)\right] + o(1).
\end{aligned} \tag{3.60}$$

From [87, Theorem 5.11] it follows that $N^{\mathrm{uc}}(t) = O_{\mathbb{P}}(n^{2/3})$ uniformly in $t = cn$ with $c \in (\frac{1}{2}, \infty)$. Since the size of the largest component is $\zeta(\frac{t}{n})n + o_{\mathbb{P}}(n)$ (recall Definition 3.1.11) and the number of vertices is n, it follows that the number of vertices outside the largest component at time t is $(1-\zeta(\frac{t}{n}))n + o_{\mathbb{P}}(n) = \Theta_{\mathbb{P}}(n)$, again uniformly in $t = cn$ with $c \in (\frac{1}{2}, \infty)$. This gives

$$\mathbb{E}\left[\min\left(\frac{N^{\mathrm{uc}}(t)}{n - |\mathscr{C}^{\mathrm{ER}}_{\max}(n,t)|}, 1\right)\right] = o(1). \tag{3.61}$$

Putting the above estimates together, we get

$$\mathbb{P}\left(\Omega^{(\mathrm{SC})}(t) \mid T^{\Downarrow}_{n,v_0} > t\right) = 1 - o(1). \tag{3.62}$$

It remains to show that this estimate holds not only conditionally on $\{T^{\Downarrow}_{n,v_0} > t\}$, but also conditionally on $\{T^{\Downarrow}_{n,v_0} \geq t\}$. Given that the ISRW support lies on a single cycle before time $T^{\Downarrow}_{n,v_0}$, at time $T^{\Downarrow}_{n,v_0}$ this cycle merges with exactly one other cycle whose elements are represented by the vertices in the largest component of the associated graph process, and so the ISRW remains supported on a single permutation cycle. It therefore follows that

$$\mathbb{P}\left(\Omega^{(\mathrm{SC})}(t) \mid T^{\Downarrow}_{n,v_0} \geq t\right) = 1 - o(1). \tag{3.63}$$

□

§3.3.3 Local mixing upon drop-down

The main difference with the setting in Section 3.2 is that each non-tree connected component of the associated graph process of CFDP may represent *multiple* permutation cycles. We show that, after scaling of time, this fine structure is not felt because the distribution of the ISRW rapidly becomes uniform over the elements of the permutation represented by the vertices of the relevant connected component of the associated graph process. A consequence of this *fast mixing* is the occurrence of the same phenomenon as observed for CDP, namely, at time $T^{\Downarrow}_{n,v_0}$ there is a single drop in the total variation distance.

• **Local mixing.** To formalise the arguments, we first introduce the notion of local mixing on the largest component of the associated graph process:

Definition 3.3.4 (Local mixing time). Consider an ISRW with distribution μ^{n,v_0} started from the element v_0 and running on top of CFDP Π_n (recall Definition 3.1.4), and let A_{Π_n} be the associated graph process. For $\varepsilon \in (0,1)$, define the stopping time

$$T^{LM,\varepsilon}_{n,v_0} = \min\left\{t > T^{\Downarrow}_{n,v_0} \colon\ d_{\mathrm{TV}}\left(\mu^{n,v_0}(t), \mathsf{Unif}(\mathscr{C}_{\max}(A_{\Pi_n}(t)))\right) < \varepsilon\right\}. \tag{3.64}$$

■

At time $T^{LM,\varepsilon}_{n,v_0}$, the ISRW is well mixed on the giant $\mathscr{C}_{\max}(A_{\Pi_n}(t))$. In the following statements we illustrate and quantify the influence of large-enough permutation cycles on ISRW-mixing. Below we play with three parameters n, ε, δ and take limits in the order $n \to \infty$, $\varepsilon \downarrow 0$ and $\delta \downarrow 0$. We also play with a time scale a_n satisfying $\lim_{n\to\infty} a_n = \infty$ and $a_n = o(n)$. Along the way we need some facts established in Appendices B.3 and B.4 that require more stringent conditions on a_n, namely, $a_n = o(n^{1/26})$, respectively, $a_n = o(n^{1/3})$. We summarise this by saying that a_n grows slowly enough.

We will often use the following Erdős-Rényi typicality event, which occurs with high probability:

Definition 3.3.5 (Erdős-Rényi typicality event). Take ε_n such that $\varepsilon_n = o(1)$ and $\varepsilon_n = \omega(n^{-1/3})$. Define the following event

$$\Omega^{(\mathrm{ER})}_n(t) = \left\{|\mathscr{C}^{\mathrm{ER}}_{\max}(n,t)| = n(\zeta(\tfrac{t}{n}) - \varepsilon_n, \zeta(\tfrac{t}{n}) + \varepsilon_n,)\right\} \cap \left\{|\mathscr{C}^{\mathrm{ER}}_{\mathrm{sec}}(n,t)| \le n\varepsilon_n\right\}. \tag{3.65}$$

Note that this event is different from the event $\Omega_n(un)$ defined in (3.34). ■

Definition 3.3.6 (Events $\mathcal{M}_1(\varepsilon,\delta)$, $\mathcal{M}_2(\varepsilon)$). Denote by $\mathfrak{X}^{(n)}_1(t)$ the normalised size of the largest cycle at time t (see (B.14)). Recall the event $\Omega^{(\mathrm{SC})}(t)$ from Lemma 3.3.3, the Erdős-Rényi typicality event $\Omega^{(\mathrm{ER})}_n(t)$ from Definition 3.3.5 (which both occur with high probability for any $t = cn$ with $c \in (\frac{1}{2}, \infty)$), and introduce the abbreviation $M = |\mathscr{C}_{\max}(A_{\Pi_n}(T^{\Downarrow}_{n,v_0}))|$. Define the events

$$\begin{aligned}\mathcal{M}_1(\varepsilon,\delta) &= \left\{|\operatorname{supp}(\mu^{n,v_0}(T^{\Downarrow}_{n,v_0}))| > \varepsilon M\right\} \cap \Omega^{(\mathrm{SC})}(T^{\Downarrow}_{n,v_0}) \cap \Omega^{(\mathrm{ER})}_n(T^{\Downarrow}_{n,v_0}),\\ \mathcal{M}_2(\varepsilon) &= \left\{\exists\, t_L \in (T^{\Downarrow}_{n,v_0}, T^{\Downarrow}_{n,v_0} + a_n) \colon\ \mathfrak{X}^{(n)}_1(t_L) > 1 - \varepsilon^2\right\}.\end{aligned} \tag{3.66}$$

■

Lemma 3.3.7 (Mixing induced by a single large cycle). *Recall Definition 3.3.6. Let $(a_n)_{n\in\mathbb{N}}$ be such that $\lim_{n\to\infty} a_n = \infty$ slowly enough. Then*

$$\left\{T^{LM,\varepsilon}_{n,v_0} \in (T^{\Downarrow}_{n,v_0}, T^{\Downarrow}_{n,v_0} + a_n)\right\} \supseteq \mathcal{M}_1(\varepsilon,\delta) \cap \mathcal{M}_2(\varepsilon). \tag{3.67}$$

Furthermore, on the event $\mathcal{M}_1(\varepsilon,\delta) \cap \mathcal{M}_2(\varepsilon) \cap \Omega^{(\mathrm{SC})}(T^{\Downarrow}_{n,v_0})$, there exists a $t_L \in (T^{\Downarrow}_{n,v_0}, T^{\Downarrow}_{n,v_0} + a_n)$ such that

$$1 - \frac{1}{n}|\mathscr{C}_{\max}(A_{\Pi_n}(t_L))| - \varepsilon \le \mathcal{D}^{v_0}_n(t_L) \le 1 - \frac{1}{n}|\mathscr{C}_{\max}(A_{\Pi_n}(t_L))| + \varepsilon. \tag{3.68}$$

Proof. Recall that the event $\Omega^{(\mathrm{SC})}(T^{\Downarrow}_{n,v_0}) \subset \mathcal{M}_1(\varepsilon,\delta)$ implies that all the mass of the ISRW-distribution enters the giant component on a single cycle. Therefore the event $\mathcal{M}_1(\varepsilon,\delta)$ implies that

$$\forall\, u \in \mathrm{supp}(\mu^{n,v_0}(T^{\Downarrow}_{n,v_0}))\colon \mu^{n,v_0}_u(T^{\Downarrow}_{n,v_0}) \le \frac{1}{\varepsilon M}, \tag{3.69}$$

where we recall that $M = |\mathscr{C}_{\max}(A_{\Pi_n}(T^{\Downarrow}_{n,v_0}))|$. The event $\mathcal{M}_1(\varepsilon,\delta) \cap \mathcal{M}_2(\varepsilon)$ indicates that a cycle of size at least $(1-\varepsilon^2)|\mathscr{C}_{\max}(A_{\Pi_n}(T^{\Downarrow}_{n,v_0}))|$ has appeared by time $T^{\Downarrow}_{n,v_0} + a_n$. We denote this large permutation cycle by $\mathfrak{X}^{(n)}_1(t_L)$. This cycle necessarily contains some mass of the ISRW-distribution because, due to the event $\mathcal{M}_1(\varepsilon,\delta)$, the mass was initially spread out over a cycle that is larger than the region not covered by $\mathfrak{X}^{(n)}_1(t_L)$. We compute the effect of the event $\mathcal{M}_1(\varepsilon,\delta) \cap \mathcal{M}_2(\varepsilon)$ on the decay of the total variation distance. The worst possible scenario is when the $\varepsilon^2 M$ elements not covered by $\mathfrak{X}^{(n)}_1(t_L)$ each carry mass $1/(\varepsilon M)$. Note that the definition of ISRW requires that the remaining mass is spread out uniformly over $\mathfrak{X}^{(n)}_1(t_L)$. A simple calculation (see Appendix B.4) shows that, at time t_L (introduced in the definition of the event $\mathcal{M}_2(\varepsilon)$) and for n large enough,

$$d_{\mathrm{TV}}\Big(\mu^{n,v_0}(t_L), \mathsf{Unif}(\mathscr{C}_{\max}(A_{\Pi_n}(t_L)))\Big) < \varepsilon, \tag{3.70}$$

and (3.67) follows.

To prove (3.68) we use that, for probability mass functions $p = (p_x)_{x\in\mathcal{X}}$ and $q = (q_x)_{x\in\mathcal{X}}$,

$$d_{\mathrm{TV}}(p,q) = \sum_{x\in\mathcal{X}} [\, p_x - (p_x \wedge q_x)\,]. \tag{3.71}$$

• For the upper bound in (3.68), we use the triangle inequality to estimate

$$\begin{aligned} &d_{\mathrm{TV}}\left(\mu^{n,v_0}(t_L), \mathsf{Unif}([n])\right) \\ &\le d_{\mathrm{TV}}\left(\mu^{n,v_0}(t_L), \mathsf{Unif}(\mathscr{C}_{\max}(A_{\Pi_n}(t_L)))\right) + d_{\mathrm{TV}}\Big(\mathsf{Unif}(\mathscr{C}_{\max}(A_{\Pi_n}(t_L))), \mathsf{Unif}([n])\Big). \end{aligned} \tag{3.72}$$

Note that, by (3.71),

$$d_{\mathrm{TV}}\Big(\mathsf{Unif}(\mathscr{C}_{\max}(A_{\Pi_n}(t_L))), \mathsf{Unif}([n])\Big) = 1 - \frac{1}{n}|\mathscr{C}_{\max}(A_{\Pi_n}(t_L))|, \tag{3.73}$$

while, by (3.70),

$$d_{\mathrm{TV}}\left(\mu^{n,v_0}(t_L), \mathsf{Unif}(\mathscr{C}_{\max}(A_{\Pi_n}(t_L)))\right) \le \varepsilon. \tag{3.74}$$

Combing (3.73)–(3.74) with (3.72), we get the upper bound in (3.68).

• For the lower bound in (3.68), we note that

$$\begin{aligned} &d_{\mathrm{TV}}\left(\mu^{n,v_0}(t_L), \mathsf{Unif}([n])\right) \\ &\ge d_{\mathrm{TV}}\left(\mu^{n,v_0}(t_L), \mathsf{Unif}(\mathscr{C}_{\max}(A_{\Pi_n}(t_L)))\right) - d_{\mathrm{TV}}\Big(\mathsf{Unif}(\mathscr{C}_{\max}(A_{\Pi_n}(t_L))), \mathsf{Unif}([n])\Big). \end{aligned} \tag{3.75}$$

Combining (3.73)–(3.74) with (3.75), we get the lower bound in (3.68). □

Before proceeding we make the following observation. Since $\frac{1}{n}T^{\Downarrow}_{n,v_0} \xrightarrow{d} u^{\Downarrow}$ by Lemma 3.3.2, and $\mathbb{P}(u^{\Downarrow} \le \frac{1}{2}+\delta) = 2\delta(1+o(1))$ for $\delta > 0$ small enough, we note that $\mathbb{P}(\frac{1}{n}T^{\Downarrow}_{n,v_0} \le \frac{1}{2}+\delta)$ can be made arbitrarily small by picking $\delta > 0$ small. Furthermore, for $\delta > 0$ small enough, $\mathscr{C}_{\max}(A_{\Pi_n}((\frac{1}{2}+\delta)n)) \ge \delta n$ with high probability, which follows from the properties of the Erdős-Rényi giant, specifically from the fact that $\zeta'(\frac{1}{2}) = 2$. Thus, (3.69) implies that, on the event

$$\Big\{T^{\Downarrow}_{n,v_0} > (\tfrac{1}{2}+\delta)n,\ |\mathscr{C}_{\max}(A_{\Pi_n}((\tfrac{1}{2}+\delta)n))| \ge \delta n\Big\}, \tag{3.76}$$

we have

$$\begin{aligned} &\forall\, u \in \mathrm{supp}(\mu^{n,v_0}(T^{\Downarrow}_{n,v_0})): \quad \mu^{n,v_0}_u(T^{\Downarrow}_{n,v_0}) \le \frac{1}{\varepsilon\delta n}, \\ &\forall\, w \notin \mathscr{C}_{\max}(A_{\Pi_n}(T^{\Downarrow}_{n,v_0})): \quad \mu^{n,v_0}_w(T^{\Downarrow}_{n,v_0}) = 0, \end{aligned} \tag{3.77}$$

where the last equality holds on the event $\Omega^{\mathrm{tree}}(T^{\Downarrow}_{n,v_0})$.

Note that for times $t \ge T^{\Downarrow}_{n,v_0}$, just as for CDP,

$$\mu^{n,v_0}_w(t) = 0 \quad \forall\, w \notin \mathscr{C}_{\max}(A_{\Pi_n}(t)) \qquad \forall\, t \ge T^{\Downarrow}_{n,v_0}, \tag{3.78}$$

since, by the construction of the associated graph process, the support of $\mu^{n,v_0}_u(T^{\Downarrow}_{n,v_0})$ always lies on a single connected component of the associated graph process. The uniform bounds above will prove to be essential below.

Lemma 3.3.7 allows us to quantify the probability of ε-mixing after a single appearance of a cycle of size $(1-\varepsilon^2)M$:

Proposition 3.3.8. *Fix $\delta > 0$. Let $(a_n)_{n\in\mathbb{N}}$ be such that $\lim_{n\to\infty} a_n = \infty$ slowly enough. Then there exists a function $\varepsilon \mapsto f(\varepsilon)$ satisfying $\lim_{\varepsilon\downarrow 0} f(\varepsilon) = 0$ such that*

$$\mathbb{P}\Big(T^{LM,\varepsilon}_{n,v_0} \notin (T^{\Downarrow}_{n,v_0}, T^{\Downarrow}_{n,v_0}+a_n), T^{\Downarrow}_{n,v_0} \ge (\tfrac{1}{2}+\delta)n\Big) \le f(\varepsilon) + o(1), \qquad n\to\infty. \tag{3.79}$$

Consequently, on the event that $T^{\Downarrow}_{n,v_0} \ge (\frac{1}{2}+\delta)n$, the conclusion of (3.68) *fails with probability at most $f(\varepsilon)$.*

Proof. We will derive an upper bound for the probability of the event $(\mathcal{M}_1(\varepsilon,\delta)^{\mathrm{c}} \cup \mathcal{M}_2(\varepsilon)^{\mathrm{c}}) \cap \{T^{\Downarrow}_{n,v_0} \ge (\frac{1}{2}+\delta)n\}$, which by (3.67) includes the event in the left-hand side of (3.79). To do so, we will work with a further sub-event.

Denote the number of vertices in $\mathscr{C}_{\max}(A_{\Pi_n}(t))$ that are in cycles of size smaller than εn by $S(\varepsilon n, t)$. We use [29, Lemma 2.4], which states that for any $t > cn$ with $c > \frac{1}{2}$ there exists a $C > 0$ such that, for any $\varepsilon \in (0,1)$ and n large enough,

$$\mathbb{E}\left[S(\varepsilon n, t)\right] < C\varepsilon \log(\tfrac{1}{\varepsilon})\, n. \tag{3.80}$$

Define the event

$$\mathcal{M}_3(\varepsilon, t) = \{S(\varepsilon n, t) < \sqrt{\varepsilon}\, n\}. \tag{3.81}$$

Recall that the mass at $T^{\Downarrow}_{n,v_0}$ enters the largest component of the associated graph process on a cycle that belongs to a uniform element. Trivially,

$$\begin{aligned} &(\mathcal{M}_1(\varepsilon,\delta)^{\mathrm{c}} \cup \mathcal{M}_2(\varepsilon)^{\mathrm{c}}) \cap \{T^{\Downarrow}_{n,v_0} \ge (\tfrac{1}{2}+\delta)n\} \\ &\quad \subseteq \Big((\mathcal{M}_1(\varepsilon,\delta)^{\mathrm{c}} \cup \mathcal{M}_2(\varepsilon)^{\mathrm{c}}) \cap \mathcal{M}_3(\varepsilon, T^{\Downarrow}_{n,v_0}) \cap \{T^{\Downarrow}_{n,v_0} \ge (\tfrac{1}{2}+\delta)n\}\Big) \cup \mathcal{M}_3(\varepsilon, T^{\Downarrow}_{n,v_0})^{\mathrm{c}}. \end{aligned} \tag{3.82}$$

We estimate the probability of these events one by one. First, use the Markov inequality and (3.80) to estimate, for n large enough,

$$\mathbb{P}\left(\mathcal{M}_3(\varepsilon, T^{\Downarrow}_{n,v_0})^{\mathrm{c}}\right) \le C\sqrt{\varepsilon}\log(\tfrac{1}{\varepsilon}). \tag{3.83}$$

Second, estimate

$$\mathbb{P}\left(\mathcal{M}_1(\varepsilon,\delta)^{\mathrm{c}} \cap \mathcal{M}_3(\varepsilon, T^{\Downarrow}_{n,v_0})\right) \le \mathbb{P}\left(\mathcal{M}_1(\varepsilon,\delta)^{\mathrm{c}} \mid \mathcal{M}_3(\varepsilon, T^{\Downarrow}_{n,v_0})\right) \le \sqrt{\varepsilon}, \tag{3.84}$$

with the last inequality following from the definition of $\mathcal{M}_1(\varepsilon,\delta)$. Third, the key estimate stated in Proposition B.3.10, whose proof turns out to be rather delicate, yields that, for $\delta > 0$ fixed,

$$\mathbb{P}\left(\mathcal{M}_2(\varepsilon)^{\mathrm{c}}, T^{\Downarrow}_{n,v_0} \ge (\tfrac{1}{2}+\delta)n\right) = o(1). \tag{3.85}$$

Indeed, the key event that is estimated in Proposition B.3.10 is

$$\mathcal{E}_n(c,\varepsilon,\kappa) = \left\{\exists\,(t_k)_{k=1}^{\kappa} \in (cn, cn+a_n)\colon\ \mathfrak{X}_1^{(n)}(t_k-1) < 1-\varepsilon,\ \mathfrak{X}_1^{(n)}(t_k) \ge 1-\varepsilon\right\}, \tag{3.86}$$

which states that there are at least $\kappa \in \mathbb{N}$ times in the interval $(cn, cn+a_n)$ such that the size of the maximal cycle crosses $(1-\varepsilon)$ upwards, i.e., $\mathfrak{X}_1^{(n)}(t_k) \ge 1-\varepsilon$. Proposition B.3.10 states that $\mathcal{E}_n(c,\varepsilon,\kappa)$ occurs with high probability for all $c \in (1/2,\infty)$, $\kappa \in \mathbb{N}$ and $\varepsilon > 0$. We apply Corollary B.3.12, which is a consequence of Proposition B.3.10, to obtain (3.85).

Combining (3.82)–(3.85), we find that there exists a $C > 0$ such that

$$\begin{aligned}
&\mathbb{P}\big((\mathcal{M}_1(\varepsilon,\delta)^{\mathrm{c}} \cup \mathcal{M}_2(\varepsilon)^{\mathrm{c}}) \cap \{T^{\Downarrow}_{n,v_0} \ge (\tfrac{1}{2}+\delta)n\}\big)\\
&\le \mathbb{P}\left(\mathcal{M}_3(\varepsilon, T^{\Downarrow}_{n,v_0})^{\mathrm{c}}\right) + \mathbb{P}\left(\mathcal{M}_1(\varepsilon,\delta)^{\mathrm{c}} \cap \mathcal{M}_3(\varepsilon, T^{\Downarrow}_{n,v_0})\right) + \mathbb{P}\left(\mathcal{M}_2(\varepsilon)^{\mathrm{c}}, T^{\Downarrow}_{n,v_0} \ge (\tfrac{1}{2}+\delta)n\right)\\
&\le C\sqrt{\varepsilon}\log(\tfrac{1}{\varepsilon}) + \sqrt{\varepsilon} + o(1) < \varepsilon^{1/3}
\end{aligned} \tag{3.87}$$

for ε small enough, which in turn decays to 0 as $\varepsilon \to 0$. □

• **Adaptation of Lemma 3.3.7 and Proposition 3.3.8.** Finally, we adapt Lemma 3.3.7 and Proposition 3.3.8. Note that Lemma 3.3.7 is true at time $t = cn$ when we replace the events $\mathcal{M}_1(\varepsilon,\delta), \mathcal{M}_2(\varepsilon)$ by (compare with (3.66))

$$\begin{aligned}
\mathcal{M}_1'(cn,\varepsilon,\delta) &= \left\{|\operatorname{supp}(\mu^{n,v_0}(cn))| > \varepsilon n\right\} \cap \Omega^{(\mathrm{SC})}(T^{\Downarrow}_{n,v_0}),\\
\mathcal{M}_2'(cn,\varepsilon,\delta) &= \left\{\exists\, t_L \in (cn, cn+a_n)\colon\ \mathfrak{X}_1^{(n)}(t_L) > 1-\tfrac{\varepsilon^2}{\delta}\right\}.
\end{aligned} \tag{3.88}$$

Here, we recall the event $\Omega^{(\mathrm{SC})}(t)$ from Lemma 3.3.3 (which occurs with high probability for any $t = cn$ with $c \in (\frac{1}{2},\infty)$), and the extra factor $1/\delta$ is added to accommodate the extra factor $1/\delta$ in the first line of (3.77). It remains to redo the calculations in the proofs of Lemma 3.3.7 and Proposition 3.3.8 with these modified events. Take $t = cn$ with $c \in (\frac{1}{2},\infty)$, and define

$$T^{LM,\varepsilon}_{n,v_0}(t) = \min\left\{s > t\colon\ d_{\mathrm{TV}}\big(\mu^{n,v_0}(s), \mathsf{Unif}(\mathscr{C}_{\max}(t))\big) < \varepsilon\right\}. \tag{3.89}$$

We start by adapting Lemma 3.3.7:

Lemma 3.3.9 (Mixing induced by a single large cycle). *Let $(a_n)_{n\in\mathbb{N}}$ be such that $\lim_{n\to\infty} a_n = \infty$ slowly enough, and let $c \in (1/2, \infty)$. Then, for any $\delta \in (0, c - \frac{1}{2})$,*

$$\begin{aligned}\big\{T^{LM,\varepsilon}_{n,v_0}(cn) \in (cn, cn + a_n)\big\} \cap \{(\tfrac{1}{2} + \delta)n \le T^{\Downarrow}_{n,v_0} \le cn\} \\ \supseteq \mathcal{M}'_1(cn, \varepsilon, \delta) \cap \mathcal{M}'_2(cn, \varepsilon, \delta) \cap \{(\tfrac{1}{2} + \delta)n \le T^{\Downarrow}_{n,v_0} \le cn - a_n\}.\end{aligned} \tag{3.90}$$

Furthermore, on the event $\mathcal{M}_1(\varepsilon, \delta) \cap \mathcal{M}_2(\varepsilon) \cap \{(\frac{1}{2}+\delta)n \le T^{\Downarrow}_{n,v_0} \le cn - a_n\} \cap \Omega^{(\mathrm{SC})}(T^{\Downarrow}_{n,v_0})$ there exists a $t_L \in (cn, cn + a_n)$ such that

$$1 - \frac{1}{n}|\mathscr{C}_{\max}(A_{\Pi_n}(t_L))| - \varepsilon \le \mathcal{D}^{v_0}_n(t_L) \le 1 - \frac{1}{n}|\mathscr{C}_{\max}(A_{\Pi_n}(t_L))| + \varepsilon. \tag{3.91}$$

Proof. The main ingredient in the proof of Lemma 3.3.7 was (3.69). Recall the extension of (3.69) in (3.77). With (3.77) in hand, we can simply follow the proof of Lemma 3.3.7. □

We continue by adapting Proposition 3.3.8:

Proposition 3.3.10. *Let $(a_n)_{n\in\mathbb{N}}$ be such that $\lim_{n\to\infty} a_n = \infty$ slowly enough, and let $c > \frac{1}{2}$. Then, for any $\delta \in (0, c - \frac{1}{2})$, with $\varepsilon \mapsto f(\varepsilon)$ as in Proposition 3.3.8,*

$$\mathbb{P}\Big(T^{LM,\varepsilon}_{n,v_0}(cn) \notin (cn, cn + a_n), (\tfrac{1}{2} + \delta)n \le T^{\Downarrow}_{n,v_0} \le cn - a_n\Big) \le f(\varepsilon) + o(1), \quad n \to \infty. \tag{3.92}$$

Consequently, on the event that $(\frac{1}{2} + \delta)n \le T^{\Downarrow}_{n,v_0} \le cn - a_n$, the conclusion of (3.91) fails with probability at most $f(\varepsilon)$.

Proof. We follow the proof of Proposition 3.3.8, which relies on the inclusion in Lemma 3.3.7. Instead, we now rely on the inclusion in Lemma 3.3.9. Recall from the proof of Lemma 3.3.7 that $S(\varepsilon n, t)$ denotes the number of vertices in $\mathscr{C}_{\max}(A_{\Pi_n}(t))$ that are in cycles of size smaller than εn, and that, by (3.80), $\mathbb{E}[S(\varepsilon n, t)] < C\varepsilon \log(\frac{1}{\varepsilon})\, n$.

Recall $T^{LM,\varepsilon}_{n,v_0}$ from (3.64). Define the event

$$\mathcal{M}'_3(\varepsilon) = \{T^{LM,\varepsilon}_{n,v_0} \in (T^{\Downarrow}_{n,v_0}, T^{\Downarrow}_{n,v_0} + a_n)\}. \tag{3.93}$$

Trivially,

$$\begin{aligned}&\Big(\mathcal{M}'_1(cn, \varepsilon, \delta)^{\mathrm{c}} \cup \mathcal{M}'_2(cn, \varepsilon, \delta)^{\mathrm{c}}\Big) \cap \Big\{(\tfrac{1}{2} + \delta)n \le T^{\Downarrow}_{n,v_0} \le cn - a_n\Big\} \\ &\quad\subseteq \Big(\mathcal{M}'_3(\varepsilon) \cap \Big(\mathcal{M}'_1(cn, \varepsilon, \delta)^{\mathrm{c}} \cup \mathcal{M}'_2(cn, \varepsilon, \delta)^{\mathrm{c}}\Big) \cap \Big\{(\tfrac{1}{2} + \delta)n \le T^{\Downarrow}_{n,v_0} \le cn - a_n\Big\}\Big) \\ &\qquad\cup \Big(\mathcal{M}'_3(\varepsilon)^{\mathrm{c}} \cap \{(\tfrac{1}{2} + \delta)n \le T^{\Downarrow}_{n,v_0} \le cn - a_n\}\Big).\end{aligned} \tag{3.94}$$

We estimate the probability of these events one by one. First, for n large enough,

$$\begin{aligned}\mathbb{P}\Big(\mathcal{M}'_3(\varepsilon)^{\mathrm{c}}, (\tfrac{1}{2} + \delta)n \le T^{\Downarrow}_{n,v_0} \le cn - a_n\Big) &\le \mathbb{P}\Big(\mathcal{M}'_3(\varepsilon)^{\mathrm{c}}, (\tfrac{1}{2} + \delta)n \le T^{\Downarrow}_{n,v_0}\Big) \\ &\le f(\varepsilon) + o(1),\end{aligned} \tag{3.95}$$

where the last inequality follows from Proposition 3.3.8.

Second, if $T^{LM,\varepsilon}_{n,v_0} \in (T^{\Downarrow}_{n,v_0}, T^{\Downarrow}_{n,v_0} + a_n)$ and $T^{\Downarrow}_{n,v_0} \geq (\frac{1}{2}+\delta)n$, then

$$\mathbb{P}\Big(\mathcal{M}'_1(cn,\varepsilon,\delta)^{\mathrm{c}} \;\Big|\; \mathcal{M}'_3(\varepsilon), (\tfrac{1}{2}+\delta)n \leq T^{\Downarrow}_{n,v_0} \leq cn - a_n\Big) = 0. \tag{3.96}$$

Indeed, $\mathcal{M}'_1(cn,\varepsilon,\delta)^{\mathrm{c}}$ and $T^{\Downarrow}_{n,v_0} \leq cn - a_n$ imply that $|\operatorname{supp}(\mu^{n,v_0}(T^{\Downarrow}_{n,v_0} + a_n))| \leq \varepsilon n$. By an application of (3.71) with $\mathcal{X} = [n]$, $p = \mathsf{Unif}([n])$ (for which $p_v = \frac{1}{n}$ for all $v \in [n]$) and $q_v = \mu^{n,v_0}(T^{\Downarrow}_{n,v_0} + a_n)$, this implies that

$$\mathcal{D}^{v_0}_n(T^{\Downarrow}_{n,v_0} + a_n) \geq 1 - \varepsilon. \tag{3.97}$$

However, the latter is incompatible with (3.68) when $T^{\Downarrow}_{n,v_0} \geq (\frac{1}{2}+\delta)n$, since

$$\begin{aligned}\mathcal{D}^{v_0}_n(T^{\Downarrow}_{n,v_0} + a_n) &\leq \mathcal{D}^{v_0}_n(t_L) \leq 1 - \frac{1}{n}|\mathscr{C}_{\max}(A_{\Pi_n}(t_L))| + \varepsilon \\ &\leq 1 - \frac{1}{n}|\mathscr{C}_{\max}(A_{\Pi_n}((\tfrac{1}{2}+\delta)n))| + \varepsilon \\ &\leq 1 - 2\delta + o(1) + \varepsilon < 1 - \varepsilon,\end{aligned} \tag{3.98}$$

where the second inequality uses the definition of the event $\mathcal{M}'_3(\varepsilon)$, and the last inequality is valid for ε small enough depending on δ. Third, apply the key estimate stated in Proposition B.3.10 (see the explanation below (3.86)), to get

$$\mathbb{P}\Big(\mathcal{M}'_2(cn,\varepsilon,\delta), (\tfrac{1}{2}+\delta)n \leq T^{\Downarrow}_{n,v_0} \leq cn - a_n\Big) = o(1). \tag{3.99}$$

Combining (3.90), (3.94)–(3.96) and (3.99), we obtain

$$\begin{aligned}&\mathbb{P}\Big(T^{LM,\varepsilon}_{n,v_0}(cn) \notin (cn, cn + a_n), (\tfrac{1}{2}+\delta)n \leq T^{\Downarrow}_{n,v_0} \leq cn - a_n\Big) \\ &\quad\leq \mathbb{P}\Big((\mathcal{M}'_1(cn,\varepsilon,\delta)^{\mathrm{c}} \cup \mathcal{M}'_2(cn,\varepsilon,\delta)^{\mathrm{c}}) \cap \{(\tfrac{1}{2}+\delta)n \leq T^{\Downarrow}_{n,v_0} \leq cn - a_n\}\Big) \\ &\quad\leq \mathbb{P}\Big(\mathcal{M}'_3(\varepsilon)^{\mathrm{c}}, (\tfrac{1}{2}+\delta)n \leq T^{\Downarrow}_{n,v_0} \leq cn - a_n\Big) \\ &\qquad + \mathbb{P}\Big(\mathcal{M}'_1(cn,\varepsilon,\delta)^{\mathrm{c}} \cap \mathcal{M}'_3(\varepsilon), (\tfrac{1}{2}+\delta)n \leq T^{\Downarrow}_{n,v_0} \leq cn - a_n\Big) \\ &\qquad + \mathbb{P}\Big(\mathcal{M}'_2(cn,\varepsilon,\delta)^{\mathrm{c}}, (\tfrac{1}{2}+\delta)n \leq T^{\Downarrow}_{n,v_0} \leq cn - a_n\Big) \\ &\quad\leq f(\varepsilon) + o(1) + 0 + o(1) = f(\varepsilon) + o(1),\end{aligned} \tag{3.100}$$

where the first inequality uses the inclusion in Lemma 3.3.9. □

§3.3.4 Mixing profile

Like in the case of CDP, the results on the mixing profile are established in two steps. First we establish pointwise convergence, afterwards we extend to process convergence. The following lemma settles Theorem 3.1.20(2):

Lemma 3.3.11 (Pointwise convergence of the mixing profile for ISRW on CFDP). *Uniformly in $v_0 \in [n]$,*

$$\mathcal{D}^{v_0}_n(sn) \xrightarrow{d} 1 - \zeta(s)W(s), \qquad s \in [0,\infty), \tag{3.101}$$

where $W(s) \sim \mathsf{Bernoulli}(\zeta(s))$.

Proof. Fix $s \in [0,\infty)$ and split the random variable $\mathcal{D}_n^{v_0}(sn) - 1$ as

$$\begin{aligned}\mathcal{D}_n^{v_0}(sn) - 1 = [\mathcal{D}_n^{v_0}(sn) - 1] \left(\mathbb{1}_{\{T^{\Downarrow}_{n,v_0} > sn\}} + \mathbb{1}_{\{T^{\Downarrow}_{n,v_0} \le sn\}} \right) \\ \times \left(\mathbb{1}_{\Omega_n^{(\mathrm{ER})}(sn)} + \mathbb{1}_{[\Omega_n^{(\mathrm{ER})}(sn)]^{\mathrm{c}}} \right) \left(\mathbb{1}_{\Omega^{\mathrm{tree}}(sn)} + \mathbb{1}_{[\Omega^{\mathrm{tree}}(sn)]^{\mathrm{c}}} \right),\end{aligned} \tag{3.102}$$

where the event $\Omega^{\mathrm{tree}}(sn)$ is defined in the proof of Lemma 3.3.3, and we recall the Erdős-Rényi typicality event (see (3.65))

$$\Omega_n^{(\mathrm{ER})}(t) = \left\{ |\mathscr{C}^{\mathrm{ER}}_{\max}(n,t)| = n(\zeta(\tfrac{t}{n}) - \varepsilon_n, \zeta(\tfrac{t}{n}) + \varepsilon_n,) \right\} \cap \left\{ |\mathscr{C}^{\mathrm{ER}}_{\mathrm{sec}}(n,t)| \le n\varepsilon_n \right\}. \tag{3.103}$$

Because $\Omega_n^{(\mathrm{ER})}(sn)$ and $\Omega^{\mathrm{tree}}(sn)$ both occur with high probability, the terms containing the indicators $\mathbb{1}_{[\Omega_n^{\mathrm{ER}}(sn)]^{\mathrm{c}}}$ and $\mathbb{1}_{[\Omega^{\mathrm{tree}}]^{\mathrm{c}}}$ converge to 0 in probability, and hence

$$\mathcal{D}_n^{v_0}(sn) - 1 = [\mathcal{D}_n^{v_0}(sn) - 1] \left(\mathbb{1}_{\{T^{\Downarrow}_{n,v_0} > sn\}} + \mathbb{1}_{\{T^{\Downarrow}_{n,v_0} \le sn\}} \right) \mathbb{1}_{\Omega_n^{(\mathrm{ER})}(sn)} \mathbb{1}_{\Omega^{\mathrm{tree}}(sn)} + o_{\mathbb{P}}(1). \tag{3.104}$$

To deal with the first term in (3.104), we note that

$$\begin{aligned}&[\mathcal{D}_n^{v_0}(sn) - 1] \mathbb{1}_{\{T^{\Downarrow}_{n,v_0} > sn\}} \mathbb{1}_{\Omega_n^{(\mathrm{ER})}(sn)} \mathbb{1}_{\Omega^{\mathrm{tree}}(sn)} \\ &\quad \stackrel{d}{=} \left[\left(1 - \frac{O_{\mathbb{P}}(n^{2/3})}{n} \right) - 1 \right] \mathbb{1}_{\{T^{\Downarrow}_{n,v_0} > sn\}} \mathbb{1}_{\Omega_n^{(\mathrm{ER})}(sn)} \mathbb{1}_{\Omega^{\mathrm{tree}}(sn)} \xrightarrow{\mathbb{P}} 0,\end{aligned} \tag{3.105}$$

since, on the above events, the distribution of ISRW is uniform over a single permutation cycle *outside* of the largest component of the associated graph process, whose size is $O_{\mathbb{P}}(n^{2/3})$.

To deal with the second term in (3.104), which only contributes when $s > \frac{1}{2}$, we use Lemma 3.3.2. For $\delta > 0$ sufficiently small and a_n as in Proposition 3.3.10, we split

$$\mathbb{1}_{\{T^{\Downarrow}_{n,v_0} \le sn\}} = \mathbb{1}_{\left\{ (\frac{1}{2}+\delta)n \le T^{\Downarrow}_{n,v_0} \le sn - a_n \right\}} + \mathbb{1}_{\left\{ sn - a_n < T^{\Downarrow}_{n,v_0} \le sn \right\}} + \mathbb{1}_{\left\{ T^{\Downarrow}_{n,v_0} < (\frac{1}{2}+\delta)n \right\}}. \tag{3.106}$$

We rely on (3.91) in Lemma 3.3.9, which holds with high probability due to Proposition 3.3.10. (It is here that we need $T^{\Downarrow}_{n,v_0} \ge (\frac{1}{2}+\delta)n$, since this appears as an assumption in Proposition 3.3.10.) We claim that

$$\sup_{t \in \mathbb{N}} \left| \frac{1}{n} |\mathscr{C}_{\max}(A_{\Pi_n}(t))| - \zeta(\tfrac{t}{n}) \right| = o_{\mathbb{P}}(1). \tag{3.107}$$

Indeed, (3.107) holds because $\frac{1}{n}|\mathscr{C}_{\max}(A_{\Pi_n}(sn))| \xrightarrow{\mathbb{P}} \zeta(s)$ for all $s > 0$ fixed, $s \mapsto \zeta(s)$ is non-decreasing and continuous, and $s \mapsto \frac{1}{n}|\mathscr{C}_{\max}(A_{\Pi_n}(sn))|$ is non-decreasing. By (3.91) and (3.107), we obtain, for all $s > \frac{1}{2} + \delta$ and on the event $\{T^{LM,\varepsilon}_{n,v_0}(cn) \in (sn, sn + a_n), (\frac{1}{2}+\delta)n \le T^{\Downarrow}_{n,v_0} \le sn - a_n\}$, that there exists a $t_L \in (sn, sn + a_n)$ such that

$$1 - \zeta(\tfrac{t_L}{n}) - \varepsilon - o_{\mathbb{P}}(1) \le \mathcal{D}_n^{v_0}(t_L) \le 1 - \zeta(\tfrac{t_L}{n}) + \varepsilon + o_{\mathbb{P}}(1). \tag{3.108}$$

Since $\varepsilon > 0$ is arbitrary, we conclude that, on the event $\{T^{LM,\varepsilon}_{n,v_0}(sn) \in (sn, sn+a_n), (\frac{1}{2}+\delta)n \le T^{\Downarrow}_{n,v_0} \le sn - a_n\}$, there exists a $t_L \in (sn, sn + a_n)$ such that

$$\mathcal{D}_n^{v_0}(t_L) = 1 - \zeta(\tfrac{t_L}{n}) + o_{\mathbb{P}}(1). \tag{3.109}$$

Since the above is true for all $s > \frac{1}{2} + \delta$, and $t \mapsto \mathcal{D}_n^{v_0}(t)$ is non-increasing, while $s \mapsto 1 - \zeta(s)$ is non-increasing and continuous, (3.109) implies that, for all $c > \frac{1}{2} + \delta$ and on the event $\{(\frac{1}{2} + \delta)n \leq T^{\Downarrow}_{n,v_0} \leq sn - a_n\}$,

$$\mathcal{D}_n^{v_0}(sn) = 1 - \zeta(s) + o_{\mathbb{P}}(1). \tag{3.110}$$

Since $\mathbb{1}_{\Omega_n^{(\mathrm{ER})}(sn)} \mathbb{1}_{\Omega^{\mathrm{tree}}(sn)} \xrightarrow{\mathbb{P}} 1$, it follows that

$$\begin{aligned}[1 - \mathcal{D}_n^{v_0}(sn)] \mathbb{1}_{\{(\frac{1}{2}+\delta)n \leq T^{\Downarrow}_{n,v_0} \leq sn - a_n\}} \mathbb{1}_{\Omega_n^{(\mathrm{ER})}(sn)} \mathbb{1}_{\Omega^{\mathrm{tree}}(sn)} - \zeta(s) \mathbb{1}_{\{\frac{1}{2}+\delta \leq \frac{1}{n} T^{\Downarrow}_{n,v_0} \leq s - \frac{a_n}{n}\}} \\ = o_{\mathbb{P}}(1).\end{aligned} \tag{3.111}$$

By Lemma 3.3.2 and Slutsky's theorem, we thus conclude that

$$[1 - \mathcal{D}_n^{v_0}(sn)] \mathbb{1}_{\{T^{\Downarrow}_{n,v_0} \leq sn - a_n\}} \mathbb{1}_{\Omega_n^{(\mathrm{ER})}(sn)} \mathbb{1}_{\Omega^{\mathrm{tree}}(sn)} \xrightarrow{d} \zeta(s) \mathbb{1}_{\{\frac{1}{2}+\delta \leq u^{\Downarrow} \leq s\}}. \tag{3.112}$$

Finally, by Lemma 3.3.2,

$$\mathbb{P}(sn - a_n < T^{\Downarrow}_{n,v_0} \leq sn) + \mathbb{P}(\tfrac{1}{n} T^{\Downarrow}_{n,v_0} < \tfrac{1}{2} + \delta) \to \zeta(\tfrac{1}{2} + \delta), \tag{3.113}$$

which tends to 0 as $\delta \downarrow 0$. The claim in (3.101) follows by combining (3.104)–(3.106) and (3.112)–(3.113). □

Finally, an argument based on monotonicity and a growing sequence of compact intervals settles Theorem 3.1.20 and concludes this section.

Proof of Theorem 3.1.20(2). Observe that for every $n \in \mathbb{N}$, any realisation of $\mathcal{D}_n^{v_0}(\cdot)$ is a monotone càdlàg path on the set $[0, \infty)$. The pointwise convergence proven in Lemma 3.3.11 implies, by [151, Corollary 12.5.1.], pathwise convergence in the Skorokhod M_1-topology on any compact set $[0, t]$ such that $t > 0$ is with probability 1 a continuity point of the limiting process. But the latter is true for any $t > 0$ because the limiting process has almost surely *one* point of discontinuity, whose position is distributed randomly according to the non-atomic distribution identified in Lemma 3.3.2. Taking a sequence $(t_k)_{k \in \mathbb{N}}$ of such continuity points with $t_k \to \infty$ as $k \to \infty$, we also obtain pathwise convergence in the Skorokhod M_1-topology on the non-compact set $[0, \infty)$. For details, see [151, p. 414]. □

APPENDIX B

Appendices of Chapter 3

§B.1 Infinite-speed random walk as limit of finite-speed random walk

It is natural to ask what the relation is between an infinite-speed random walk (ISRW) and a finite-speed standard random walk.

Definition B.1.1 (Finite-speed random walk on Π_n). Let Π_n be a dynamic permutation. Fix an element v_0. Recall that $\gamma_{v_0}(0)$ is the cycle of $\Pi_n(0)$ that contains v_0. Pick $\rho \in \mathbb{N}$ as the speed ratio between the evolution of the random walk and the random graph. More formally, let Π_n satisfy the condition that

$$\forall\, n \in \mathbb{N}_0\ \forall\, i\ \in \{0, \dots, \rho - 1\} \colon\ \Pi_n(\rho n) \equiv \Pi_n(\rho n + i). \tag{B.1}$$

Denote by $\Pi_n(t)[i]$ the image of the element i under the permutation $\Pi_n(t)$. The finite-speed random walk is the random process $(Y_n^{v_0}(t))_{t\in\mathbb{N}_0}$ on $[n]$ that has $Y_n^{v_0}(0) = v_0$ and, for any time $t > 0$,

$$\begin{aligned}
&\mathbb{P}(Y_n^{v_0}(t) = i \,|\, Y_n^{v_0}(t-1) = j) \\
&\quad = \begin{cases} 1, & \text{if } \Pi_n(t)[i] = j \text{ and } \Pi_n(t)[j] = i \text{ (which is the case for 1- and 2-cycles),} \\ \frac{1}{2}, & \text{if either } \Pi_n(t)[i] \neq j \text{ and } \Pi_n(t)[j] = i \text{ or } \Pi_n(t)[i] = j \text{ and } \Pi_n(t)[j] \neq i, \\ 0, & \text{otherwise.} \end{cases}
\end{aligned} \tag{B.2}$$

i.e., a simple symmetric random walk on the elements of any given cycle, where the underlying permutation changes once every ρ steps. ■

With this definition in hand, we can explain what we mean by saying that ISRW has infinite speed:

Proposition B.1.2 (ISRW arises as limit). *Let Π_n be a dynamic permutation. Recall $(\mu^{n,v_0}(t))_{t\in\mathbb{N}_0}$ defined in Definition 3.1.6, and let $Y_n^{v_0} = \{Y_n^{v_0}(t)\}_{t\in\mathbb{N}_0}$ be a finite-speed random walk with speed ratio ρ on the dynamic permutation Π_n^Y, with distribution $\mu^{Y_n^{v_0}}(t)$ at time t. Note that $v_0 \in [n]$ is the starting element for both the ISRW and the finite-speed random walk. Finally, suppose that the following relation holds between Π_n and Π_n^Y:*

$$\forall\, n \in \mathbb{N}_0\ \forall\, i\ \in \{0, \dots, \rho - 1\} \colon\ \Pi_n(n + i) \equiv \Pi_n^Y(n). \tag{B.3}$$

Then, with high probability, for any $i \in \mathbb{N}$,

$$d_{\mathrm{TV}}\left(\mu^{n,v_0}(i), \mu^{Y_n^{v_0}}(\rho i)\right) \overset{\rho\to\infty}{\to} 0. \tag{B.4}$$

Proof. It is well-known that a simple random walk on a circle of size m mixes in $O(m^2)$ steps. Pick $\varepsilon_{RW} > 0$, and pick $\rho(\varepsilon_{RW})$ larger than the ε_{RW}-mixing time of a simple random walk on a circle of size n. Denote by $Y_n^{v_0}(\rho(\varepsilon_{RW})i)|_{\mathcal{C}^Y_{\rho i}(v)}$ the restriction of $Y_n^{v_0}(\rho(\varepsilon_{RW})i)$ to the cycle $\mathcal{C}^Y_{\rho i}(v) \in \Pi^Y_n(\rho i)$, and let $c\,\mathsf{Unif}(S)$ be the uniform distribution multiplied component-wise by the constant c. Then

$$d_{\mathrm{TV}}\left(Y_n^{v_0}(\rho(\varepsilon_{RW})i)|_{\mathcal{C}^Y_{\rho i}(v)}, \frac{1}{\sum_{v\in\mathcal{C}_i(v)}\mu_v^{n,v_0}(i)}\mathsf{Unif}([|\mathcal{C}_i(v)|])\right) < \varepsilon_{RW}. \tag{B.5}$$

Since, by definition,

$$\mu^{n,v_0}(i)|_{\mathcal{C}_i(v)} = \frac{1}{\sum_{v\in\mathcal{C}_i(v)}\mu_v^{n,v_0}(i)}\,\mathsf{Unif}([|\mathcal{C}_i(v)|]), \tag{B.6}$$

the claim follows. □

§B.2 Normalisation of the jump-time distribution

In this appendix we return to Definition 3.1.11 and show that $\lim_{s\to\infty}\phi(s) = 1$, i.e., the laws of the jump-down times in Theorems 3.1.19–3.1.20 are normalised. The proof amounts to showing that

$$\int_{\frac12}^{\infty} \mathrm{d}s\,[1-\zeta^2(s)] = \tfrac12, \tag{B.7}$$

where $\zeta(u)$ is the unique solution of the equation

$$\mathrm{e}^{-2s\zeta(s)} = 1-\zeta(s), \qquad s \in \left[\tfrac12,\infty\right). \tag{B.8}$$

To evaluate (B.7), introduce a new variable $u = \zeta(s)$ and write

$$\int_{\frac12}^{\infty} \mathrm{d}s\,[1-\zeta^2(s)] = \int_0^1 \mathrm{d}u\,\frac{\mathrm{d}s}{\mathrm{d}u}\,(1-u^2). \tag{B.9}$$

To compute the Jacobian $\frac{\mathrm{d}s}{\mathrm{d}u}$, we take the logarithm of (B.8) and differentiate:

$$s = -\frac{1}{2u}\log(1-u), \qquad \frac{\mathrm{d}s}{\mathrm{d}u} = \frac{1}{2u(1-u)} + \frac{\log(1-u)}{2u^2}. \tag{B.10}$$

This gives

$$\begin{aligned}\int_0^1 \mathrm{d}u\,\frac{\mathrm{d}s}{\mathrm{d}s}\,(1-u^2) &= \int_0^1 \mathrm{d}u\,(1+u)\frac{1}{2u}\left[1-\frac{1-u}{u}\sum_{k\in\mathbb{N}}\frac{u^k}{k}\right]\\ &= \frac12\int_0^1 \mathrm{d}u\,\left[(1+u)-(1-u^2)\sum_{l\in\mathbb{N}_0}\frac{u^l}{l+2}\right] = \frac12\int_0^1 \mathrm{d}u\sum_{l\in\mathbb{N}_0} c_l u^l\end{aligned} \tag{B.11}$$

with the coefficients c_l given by

$$c_l = \delta_{0l} + \delta_{1l} - \frac{1}{l+2}\mathbb{1}_{\{l\geq 0\}} + \frac{1}{l}\mathbb{1}_{\{l\geq 2\}}, \qquad l \in \mathbb{N}_0. \tag{B.12}$$

The sum has radius of convergence 1, and so term by term integration gives

$$\begin{aligned}\frac{1}{2}\int_0^1 \mathrm{d}u \sum_{l\in\mathbb{N}_0} c_l u^l &= \frac{1}{2}\left[\frac{1}{2}\times 1 + \frac{2}{3}\times\frac{1}{2} + \sum_{l\geq 2}\frac{1}{l+1}\left(\frac{1}{l}-\frac{1}{l+2}\right)\right] \\ &= \frac{1}{4}+\frac{1}{6}+\frac{1}{2}\sum_{l\geq 2}\left[\frac{1}{l(l+1)}-\frac{1}{(l+1)(l+2)}\right] = \frac{1}{4}+\frac{1}{6}+\frac{1}{12} = \frac{1}{2},\end{aligned} \tag{B.13}$$

as required.

§B.3 Cycle structure of coagulative-fragmentative dynamic permutations and Schramm's coupling

In [135], Schramm introduced a remarkable coupling to study the cycle structure of CFDP. The coupling is realised between the part of the dynamic permutation that is supported on elements that lie in the largest component of the associated graph process and an independent $\mathsf{PoiDir}(1)$-sample, both seen as partitions of the unit interval $[0,1]$ evolving via coagulative-fragmentative dynamics. The main idea of the coupling is to evolve both partitions according to the same dynamics, but to rearrange and *match* components of the partition that become close in size. A detailed account of Schramm's coupling is given in [29, Section 5], to which we refer the reader. Below we provide a short overview of the main features.

Appendix B.3.1 provides a short summary of Schramm's coupling and collects a number of estimates that are needed later on. Appendix B.3.2 proves a key proposition (Proposition B.3.10 below) stating that large cycles recur arbitrarily often in any time interval of diverging length.

§B.3.1 Short summary of Schramm's coupling

Schramm's coupling is defined for a pair $\vec{Y}(t), \vec{Z}(t)$ of evolving partitions of the unit interval $[0,1]$ into countable many subintervals. The dynamics of the evolution is chosen such that the Poisson-Dirichlet distribution $\mathsf{PoiDir}(1)$ is invariant under the dynamics (see e.g. [143]). As we will see below, such a dynamics corresponds to the coagulative-fragmentative dynamics considered in Chapter 3.

Definition B.3.1 (Abstract version of Schramm's coupling). Take two partitions $\vec{Y}(0), \vec{Z}(0)$ of the unit interval $[0,1]$ into countably many subintervals. A single step of the coupling proceeds as follows:

(a) If there is a subinterval $a \in \vec{Y}(0)$ that has the same length as some other subinterval $b \in \vec{Z}(0)$, then declare a and b to be *matched*. (Note that the relation of being matched is symmetric.)

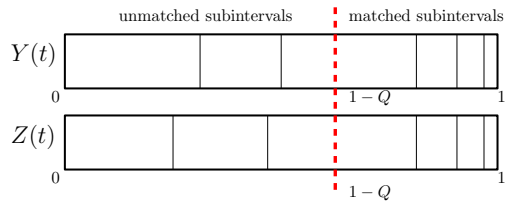


Figure B.1: Illustration of the setup for Schramm's coupling defined in Definition B.3.1.

(b) Reorder the subintervals within $\vec{Y}(0)$ and $\vec{Z}(0)$ as follows:

  (a) Let Q be the total length of all the matched intervals. In both $\vec{Y}(0)$ and $\vec{Z}(0)$, place within $(1-Q, 1]$ all the *matched* subintervals, ordered by their size such that the longest matched subinterval is on the left.
  (b) In both $\vec{Y}(0)$ and $\vec{Z}(0)$, place within $[0, 1-Q)$ the respective *unmatched* subintervals from the respective partitions, once again ordered by their size such that the longest matched subinterval is on the left.

(c) Sample $U, U' \sim \mathsf{Unif}([0,1))$ and use these random variables to evolve the partitions $\vec{Y}(0), \vec{Z}(0)$ as follows:

  (a) Call the subintervals $h_1 \in \vec{Y}(0)$ or $h_2 \in \vec{Z}(0)$ *highlighted* if U falls into these subintervals after the reordering described above.
  (b) If U' falls into a subinterval $g_1 \in \vec{Y}(0)$ such that $g_1 \neq h_1$, then merge h_1 and g_1. Do the same for h_2 and $\vec{Z}(0)$.
  (c) If U' falls into a highlighted subinterval $h_1 \in \vec{Y}(0)$, then split h_1 at U'. Do the same for h_2 and $\vec{Z}(0)$.

(d) Call $\vec{Y}(1)$ and $\vec{Z}(1)$ the new partitions that are obtained by the reordering and the application of the dynamics, and repeat.

See Figure B.1 for an illustration. ■

The construction specified in Definition B.3.1 has to be modified slightly to fit the setting of CFDP. First, we need to convert a permutation into a partition of the unit interval.

Definition B.3.2 (Cycle structure on a dynamic permutation). Denote by $|\gamma^{(i)}(t)|$ the size of the i^{th}-largest permutation cycle at time $t \geq 0$, with the following conventions: if there is only a finite number $k \in \mathbb{N}$ of permutation cycles, then $|\gamma^{(j)}(t)| = 0$ for all $j > k$, while if two or more elements have the same size, then their ordering (among each other) is arbitrary. Define

$$\vec{\mathfrak{x}}^{(n)}(t) = \left(\mathfrak{x}_i^{(n)}(t)\right)_{i\in\mathbb{N}} = \left(\frac{|\gamma^{(i)}(t)|}{|\mathscr{C}_{\max}(A_{\Pi_n}(t))|}\right)_{i\in\mathbb{N}}, \tag{B.14}$$

i.e., the ordered sequence of sizes of permutation cycles normalised by the size of $\mathscr{C}_{\max}(A_{\Pi_n}(t))$. Abusing notation, we denote by $\vec{\mathfrak{x}}^{(n)}(t)$ also the partition of the interval $[0, n/|\mathscr{C}_{\max}(A_{\Pi_n}(t)|)]$ into subintervals induced by $\vec{\mathfrak{x}}^{(n)}(t)$. ■

Remark B.3.3 (Normalisation of $\vec{\mathfrak{X}}^{(n)}(t)$). Note that, with this normalisation, the sum of all the elements in $\vec{\mathfrak{X}}^{(n)}(t)$ is $n/|\mathscr{C}_{\max}(A_{\Pi_n}(t))|$, which is roughly $\frac{1}{\zeta(t/n)}$, and the permutation cycles with elements on the largest component of the associated graph process (which is unique with high probability) correspond approximately to the subinterval $[0,1]$ of the partition induced by $\vec{\mathfrak{X}}^{(n)}(t)$. ♦

All the details required to modify Schramm's coupling to the setting of CFDP are explained in [29, Sections 5.2 and 5.3], and we will not reproduce them here in full. The main modifications of the coupling, this time between $\vec{\mathfrak{X}}^{(n)}(t)$ seen as a partition of $[0, n/|\mathscr{C}_{\max}(A_{\Pi_n}(t))|]$ and a random partition of the interval $[0,1]$ distributed according to $\mathsf{PoiDir}(1)$, are as follows:

(a) Allow for *approximate* matching of components, i.e., allows for a margin of order $O(n^{-1/2})$.

(b) Markers U, U' that generate the dynamics are sampled uniformly from the interval $[0, n/|\mathscr{C}_{\max}(A_{\Pi_n}(t))|]$, but if one of them falls outside the interval $[0,1]$, then the move is *not carried out* in the initially $\mathsf{PoiDir}(1)$-distributed partition.

(c) A *forbidden* set $F(t)$ is the set of points that U, U' must avoid for the coupling to be successful. This set takes care of the possible errors that may arise due to, for example, the discrete nature of the permutations or the growing size of the largest component in the associated graph process.

Remark B.3.4 (Limiting distribution of the cycle structure $\vec{\mathfrak{X}}^{(n)}$). In [135, Theorem 1.1], Schramm's coupling in the form adapted to CFDP was used to show that, at any time $t \geq cn$ with $c > \frac{1}{2}$, the restriction of $\vec{\mathfrak{X}}^{(n)}(t)$ to the interval $[0,1]$ has the distributional limit

$$\vec{\mathfrak{X}}^{(n)}_{|[0,1]}(t) \xrightarrow{d} \mathsf{PoiDir}(1), \tag{B.15}$$

where $\mathsf{PoiDir}(1)$ is the Poisson-Dirichlet distribution with parameter 1, which is the *unique* invariant distribution of $\vec{\mathfrak{X}}^{(n)}$ w.r.t. the permutation dynamics [135]. ♦

Note that the largest cycle $\mathfrak{X}^{(n)}_1(cn)$, for $c > 1/2$, is typically large (see e.g. [83]). In the lemma stated below, which is adapted from [29, Section 5.3], we collect results that give us control over $\vec{\mathfrak{X}}^{(n)}(t)$. To understand the limitations in the statement of this lemma, we need to introduce some extra notation:

Definition B.3.5 (Events relating to the success of Schramm's coupling). Let $c > \frac{1}{2}$, fix $\beta > 0$ and set $T_\beta = \lceil \beta^{-1/2} \rceil - 1$ and $\hat{I}_{c,\beta} = [cn, cn + T_\beta]$. Consider Schramm's coupling

$$\big(\vec{\mathfrak{X}}^{(n)}(t), \vec{Z}(t)\big)_{t \in \hat{I}_{c,\beta}}$$

with $\vec{Z}(0) \sim \mathsf{PoiDir}(1)$. For $t \in \hat{I}_{c,\beta}$, define

- $N_\beta(\vec{\mathfrak{X}}^{(n)}(t))$ to be the number of unmatched blocks in $\vec{\mathfrak{X}}^{(n)}(t)$ whose sizes are larger than β. Analogously for $N_\beta(\vec{Z}(t))$.
- $\overline{N}_\beta(t) = N_\beta(\vec{\mathfrak{X}}^{(n)}(t)) + N_\beta(\vec{Z}(t))$.

- $\sigma(\beta, \vec{\mathfrak{X}}^{(n)}(t))$ to be the total length of the blocks of size smaller than β in the partition $\vec{\mathfrak{X}}^{(n)}(t)$. Analogously for $\vec{Z}(t)$.
- $\overline{\beta}(t) = \beta + \sigma(\beta, \vec{\mathfrak{X}}^{(n)}(t)) + \sigma(\beta, \vec{Z}(t))$.
- $\mathfrak{X}_1^{(n,\mathrm{UM})}(t)$ to be the largest unmatched segment in $\vec{\mathfrak{X}}^{(n)}(t)$.

Using the above notation, define the following events:

$$\mathcal{A}_1(t) = \{\overline{\beta}(t) \le \beta^{3/4}\}, \qquad \mathcal{A}_2(t) = \{\overline{N}_\beta(t) \le \beta^{-1/4}\}. \tag{B.16}$$

■

The next definition captures the key regularity event used in the upcoming arguments.

Definition B.3.6 (Dynamics regularity event for Schramm's coupling). Define the event[1] $\mathcal{A}_3(\hat{I}_{c,\beta})$ by requiring that for any time $t \in \hat{I}_{c,\beta}$ the following three facts hold:

(a) U, U' sampled at time t do not fall in the forbidden set $F(t)$.

(b) The dynamics does not split a component of $\vec{\mathfrak{X}}^{(n)}(t)$ of size $\le \sqrt{n}/|\mathscr{C}_{\max}(A_{\Pi_n}(t))|$.

(c) If the dynamics induces a fragmentation in one of the partitions, then it does so also in the other partition.

■

Since the event $\mathcal{A}_3(\hat{I}_{c,\beta})$ is crucial for the success of Schramm's coupling, we will need the following quantitative estimate (see also [29, eq. (5.7)]) to get uniform control over the time scale on which Schramm's coupling remains successful with high probability.

Lemma B.3.7 (Probability bound for $\mathcal{A}_3(\hat{I}_{c,\beta})$). *Let $c > \frac{1}{2}$ and $\beta > 0$. Then*

$$\mathbb{P}\left(\mathcal{A}_3^{\mathrm{c}}(\hat{I}_{c,\beta})\right) \le 16\beta^{-1}n^{-1/13}. \tag{B.17}$$

With this notation and information in hand, we can now state the following key lemma, adapted from [29, Section 5.3], which gives us control over the cycle structure $\vec{\mathfrak{X}}^{(n)}(t)$:

Lemma B.3.8 ($\mathsf{PoiDir}(1)$ approximation of $\vec{\mathfrak{X}}^{(n)}(t)$). *Fix $c > \frac{1}{2}$ and $\beta > 0$. Let $\vec{\mathfrak{X}}^{(n)}(t)$ be as in Definition B.3.2. Consider $(\vec{Z}(t))_{t\in\hat{I}_{c,\beta}}$ such that $\vec{Z}(cn) \sim \mathsf{PoiDir}(1)$ is sampled independently of anything else, and at later times the evolution of $\vec{Z}(t)$ is governed by the dynamics of the underlying permutation (see [29, Section 5]). Consider Schramm's coupling $(\vec{\mathfrak{X}}^{(n)}(t), \vec{Z}(t))_{t\in\hat{I}_{c,\beta}}$, and let $q \sim \mathsf{Unif}([cn, \dots, cn + T_\beta])$ be a uniform random variable in $\hat{I}_{c,\beta}$, independent of anything else. For $\delta \in (0,1)$, define the event*

$$\mathscr{D}_{\delta,\beta} = \left\{\left\|\vec{\mathfrak{X}}^{(n)}(q) - \vec{Z}(q)\right\|_\infty < \delta\right\}. \tag{B.18}$$

[1] See [29, p. 44], where this event is denoted by $\mathcal{G}$ and the three defining properties correspond to the three items before Eq. (5.7) therein.

Recall the events in (3.34) *and in Definition B.3.5. There exist constants* $C, C' > 0$ *such that, for* n *sufficiently large, the following estimate is valid uniformly in* β *and* δ*:*

$$\begin{aligned}\mathbb{P}(\mathscr{D}^{\mathrm{c}}_{\delta,\beta}) &\leq \mathbb{P}\left(\Omega^{\mathcal{C}}_n(cn)\right) + \mathbb{P}\left(\Omega_n(cn) \cap \mathcal{A}^{\mathrm{c}}_1(cn)\right) \qquad \text{(B.19)}\\ &\quad + \mathbb{P}\left(\Omega_n(cn) \cap \mathcal{A}^{\mathrm{c}}_2(cn)\right) + \mathbb{P}\left(\Omega_n(cn) \cap \mathcal{A}^{\mathrm{c}}_3(\hat{I}_{c,\beta})\right)\\ &\quad + \mathbb{P}\left(\left\{\mathfrak{X}^{(n,\mathrm{UM})}_1(q) \geq \delta/2\right\} \cap \Omega_n(cn) \cap \mathcal{A}_1(cn) \cap \mathcal{A}_2(cn) \cap \mathcal{A}_3(\hat{I}_{c,\beta})\right)\\ &\leq o(1) + 3C\beta^{1/4}\log\left(\beta^{-1}\right) + 2C'\beta^{1/4}\log^2\left(\beta^{-1}\right) \qquad \text{(B.20)}\\ &\quad + 16\beta^{-1}n^{-1/13} + \frac{2C_1}{\delta\log(\beta^{-3/4})}.\end{aligned}$$

The proof of this lemma can be extracted from the various arguments and statements presented in [135, Section 3]. Alternatively, the precise bounds in (B.19) and (B.20) are explicitly derived in [29, Section 5]. In particular, the first inequality in (B.19) is obtained by applying a union bound in the second term of the first bound in the last display of the proof of [29, Theorem 1.1, p. 45]. The quantitative bounds on the non-trivial terms in (B.20) can be found in, respectively, the second to last display in the proof of [29, Theorem 1.1, p. 45] for the terms involving $\mathcal{A}^{\mathrm{c}}_1$ and $\mathcal{A}^{\mathrm{c}}_2$, Lemma B.3.7 for the term involving $\mathcal{A}^{\mathrm{c}}_3$, and [29, Corollary 5.7] for the last term related to the maximal unmatched block $\mathfrak{X}^{(n,\mathrm{UM})}_1$.

§B.3.2 Recurrence of large cycles

We will make use of Schramm's coupling in the proof of Proposition B.3.10 below, which is at the very core of our argument in the proofs of Propositions 3.3.8 and 3.3.10. To this aim, we first use the quantitative estimate in Lemma B.3.8 to derive the following technical lemma, stating that the Poisson-Dirichlet approximation of the underlying permutation dynamics is good with high probability over a properly chosen time interval of diverging length.

Lemma B.3.9 (Pathwise approximation property of Schramm's coupling). *Fix* $\varepsilon \in (0, \frac{1}{2})$ *and* $c > \frac{1}{2}$*. Take any time interval of the form* $I_{c,n} = [cn, cn + a_n]$ *with* $\lim_{n\to\infty} a_n = \infty$ *and* $a_n = o(n^{1/26})$*, and denote by* $\mathcal{F}^{\mathrm{SC}} = (\mathcal{F}^{\mathrm{SC}}_t)_{t\in I_c}$ *the natural filtration of Schramm's coupling. For* $a, b \in I_{c,n}$*, define the event that for all* $t \in [a, b]$ *the cycle structure of the underlying permutation* $\vec{\mathfrak{X}}^{(n)}(t)$ *is well-approximated by the process* $\vec{Z}(t)$ *introduced above. More precisely, define*

$$\Omega^{\mathrm{WA}}(a, b, \varepsilon) = \left\{\forall\, t \in [a, b]\colon\ \left\|\vec{\mathfrak{X}}^{(n)}(t) - \vec{Z}(t)\right\|_\infty < \varepsilon\right\} \in \mathcal{F}^{\mathrm{SC}}_b, \qquad \text{(B.21)}$$

pick a sequence $(d_n)_{n\in\mathbb{N}}$ *such that* $d_n = o(\log a_n)$ *and* $d_n \to \infty$ *as* $n \to \infty$*. Then*

$$\mathbb{P}\left(\Omega^{\mathrm{WA}}(q_n, q_n + \sqrt{d_n}, \varepsilon)\right) = 1 - o(1) \qquad \text{(B.22)}$$

for $q_n \sim \mathsf{Unif}(I_{c,n})$*.*

Proof. We will use the estimates in Appendix B.3.1. In particular, with the choice of a_n and d_n as in the statement of the lemma, we set $\delta = \delta_n = \varepsilon/d_n$, and $\beta = \beta_n$ such that $a_n = \lceil \beta_n^{-1/2} \rceil - 1$, and note that Lemma B.3.7 with $\hat{I}_{c,\beta} = I_{c,n}$ guarantees that

$$\mathbb{P}\left(\mathcal{A}_3(I_{c,n})\right) \geq 1 - o(1) \tag{B.23}$$

as soon as $a_n = o(n^{1/26})$.

The key event $\mathcal{A}_3(I_{c,n})$, occurring with high probability, guarantees regularity of the dynamics uniformly over the time interval $I_{c,n}$. On this event, we can find a (much smaller) random subinterval $[q_n, q_n + \sqrt{d_n}\,]$, still of diverging length, for which the Poisson-Dirichlet approximation up to the threshold ε is valid with high probability. Indeed, with the notation as in (B.18), we can bound

$$\begin{aligned}
\mathbb{P}\big(\Omega^{\mathrm{WA}}(q_n, q_n + \sqrt{d_n}, \varepsilon)\big) &= \mathbb{P}\Big(\forall\, t \in [q_n, q_n + \sqrt{d_n}]\colon \ \Big\|\vec{\mathfrak{X}}^{(n)}(t) - \vec{Z}(t)\Big\|_\infty < \varepsilon\Big) \\
&\geq \mathbb{P}\big(\mathscr{D}_{\varepsilon/d_n,\beta_n} \cap \mathcal{A}_3(I_{c,n})\big) \\
&= \mathbb{P}\big(\mathscr{D}_{\varepsilon/d_n,\beta_n}\big) - \mathbb{P}\big(\mathscr{D}_{\varepsilon/d_n,\beta_n} \cap \mathcal{A}_3^{\mathrm{c}}(I_{c,n})\big) \\
&\geq 1 - \mathbb{P}\big(\mathscr{D}^{\mathrm{c}}_{\varepsilon/d_n,\beta_n}\big) - o(1) \geq 1 - o(1),
\end{aligned}$$

where the first inequality is true because, on the event $\mathcal{A}_3(I_{c,n})$, if at time $q_n \in I_{c,n}$ we have $\|\vec{\mathfrak{X}}^{(n)}(q_n) - \vec{Z}(q_n)\|_\infty < \varepsilon/d_n$, then the next $\sqrt{d_n}$ (coagulation or fragmentation) moves of the dynamics can only increase the sup-norm in a bounded way. To see why, observe that on the event $\mathcal{A}_3(I_{c,n})$ only the following can occur:

- Due to Definition B.3.6[1.], we can leave out the effect of fragmentations of tiny components. Furthermore, Definition B.3.6[3.] guarantees consistency of moves between the two coupled cycle structures.

- Coagulation or fragmentation within the *matched* components can increase the sup-norm by at most $\frac{C}{\sqrt{n}}$, which is the margin allowed by the approximate matching rule. This discrepancy can be made arbitrarily small by taking n large enough. This follows from Definition B.3.6[1.] and the definition of a forbidden set $F(t)$.

- On the event $\mathscr{D}_{\varepsilon/d_n,\beta_n}$, the largest unmatched component satisfies $\mathfrak{X}_1^{(n,\mathrm{UM})}(q_n) \leq \frac{\varepsilon}{2d_n}$. This guarantees that, during the interval $[q_n, q_n + \sqrt{d_n}]$, the coagulation or fragmentation moves involving at least one *unmatched* component can, in the worst case, increase the sup-norm by at most $\sum_{k=1}^{\sqrt{d_n}} \frac{k\varepsilon}{2d_n} \leq \varepsilon$.

The other inequalities follow, respectively, from (B.23) and (B.20), with the choice of parameters as here. □

Using the pathwise approximation result in Lemma B.3.9, we can finally state the core proposition of this appendix (which was used to obtain (3.96)):

Proposition B.3.10 (Infinite crossings of level $1-\varepsilon$ for $\mathfrak{X}_1^{(n)}$). *Fix $\varepsilon \in (0, \frac{1}{2})$ and $c > \frac{1}{2}$. Take any time interval of the form $I_{c,n} = [cn, cn + a_n]$ with $\lim_{n\to\infty} a_n = \infty$ and*

$a_n = o(n^{1/26})$. *Consider the event* $\mathcal{E}_n(c,\varepsilon,\kappa)$ *that the process* $(\mathfrak{X}_1^{(n)}(t) - (1-\varepsilon))_{t=cn}^{cn+a_n}$ *changes sign* $\kappa \in \mathbb{N}$ *times along an increasing sequence of times* $(t_k)_{k=1}^{\kappa}$, *i.e.,*

$$\mathcal{E}_n(c,\varepsilon,\kappa) = \left\{\exists\,(t_k)_{k=1}^{\kappa}\colon\ \mathfrak{X}_1^{(n)}(t_k-1) < 1-\varepsilon,\ \mathfrak{X}_1^{(n)}(t_k) \geq 1-\varepsilon\right\}. \tag{B.24}$$

For any $\kappa \in \mathbb{N}$,

$$\lim_{n\to\infty} \mathbb{P}(\mathcal{E}_n(c,\varepsilon,\kappa)) = 1. \tag{B.25}$$

Proof. The idea behind the proof is that if the coupling described above holds, then the behaviour of the $\mathsf{PoiDir}(1)$-sample will induce the desired large cycles in the dynamic permutation.

Let $I_{c,n} = [cn, cn+a_n]$, and denote by $\mathcal{F}^{\mathrm{SC}} = (\mathcal{F}_t^{\mathrm{SC}})_{t\in I_c}$ the natural filtration of Schramm's coupling. For any $a,b \in I_{c,n}$, $\varepsilon \in (0,\frac{1}{2})$ and $\kappa \in \mathbb{N}$, define the event that large cycles in the Z process *recur* more that κ times:

$$\Omega^{\mathrm{ZR}}(a,b,\varepsilon,\kappa) = \left\{\#\{t\in[a,b]\colon\ Z_1(t) > 1-\varepsilon\} > \kappa\right\} \in \mathcal{F}_b^{\mathrm{SC}}. \tag{B.26}$$

The key observation is that, for any $0 < \lambda < \varepsilon$,

$$\left\{\#\{t\in[a,b]\colon\ \mathfrak{X}_1^{(n)}(t) > 1-\varepsilon\} > \kappa\right\} \supseteq \Omega^{\mathrm{WA}}(a,b,\lambda) \cap \Omega^{\mathrm{ZR}}(a,b,\varepsilon+\lambda,\kappa). \tag{B.27}$$

This is true because if the supremum norm between $\mathfrak{X}^{(n)}(t)$ and $Z(t)$ is at all times bounded by λ, then every occurrence of a cycle larger than $1-\varepsilon+\lambda$ in $Z(t)$ induces a cycle of size at least $1-\varepsilon$ in $\mathfrak{X}^{(n)}(t)$.

From Lemma B.3.9, it follows that for $q_n \sim \mathsf{Unif}(I_{c,n})$ with $d_n \to \infty$ and $d_n = o(\log a_n)$, the well-approximation event $\Omega^{\mathrm{WA}}(q_n, q_n+\sqrt{d_n}, \varepsilon/2)$ occurs with high probability. Recall that, at the beginning of the coupling, $Z(cn) \sim \mathsf{PoiDir}(1)$ independently of everything else, and the dynamics of the coupling is such that $\mathsf{PoiDir}(1)$ is invariant. Denote by $\mathcal{P}_1$ the space of size-ordered countable partitions of the unit interval $[0,1]$ into subintervals, which is the space over which the measure $\nu = \mathsf{PoiDir}(1)$ is defined. Introduce the notation

$$L_\varepsilon = \{P \in \mathcal{P}_1\colon\ P_1 > 1-\varepsilon\}, \tag{B.28}$$

where P_1 denotes the first, and therefore the largest, element of P. By [109, Theorem 1.2], for any $\varepsilon \in (0,\frac{1}{2})$,

$$\nu(L_\varepsilon) = -\log(1-\varepsilon) > 0. \tag{B.29}$$

Recall from [135] that the evolution of Z is a time-homogeneous Markov process and that $\nu = \mathsf{PoiDir}(1)$ is the unique invariant measure of this process. Since, for any $\varepsilon \in (0,\frac{1}{2})$, the set L_ε has a strictly positive ν-measure and the starting point of the process is sampled according to ν, it follows that the *hitting* time of L_ε is finite a.s. and has finite expectation. Furthermore, the *return* times to L_ε are all finite a.s. and, by the *Kac recurrence time lemma* [125, Theorem 4.6], have finite expectation, equal to $1/\nu(L_\varepsilon)$. Consequently, for any fixed number of occurrences $\kappa \in \mathbb{N}$, any fixed threshold for cycle sizes $\varepsilon \in (0,\frac{1}{2})$, random times $q_n \sim \mathsf{Unif}(I_{c,n})$ independent of anything else, and any $d_n \to \infty$,

$$\lim_{n\to\infty} \mathbb{P}\big(\Omega^{\mathrm{ZR}}(q_n, q_n+d_n, \varepsilon, \kappa)\big) = 1. \tag{B.30}$$

Therefore, for $q_n \sim \mathsf{Unif}(I_{c,n})$, any $\kappa \in \mathbb{N}$ and any sequence $(d_n)_{n\in\mathbb{N}}$ such that $d_n = o(\log a_n)$ and $d_n \to \infty$ as $n \to \infty$,

$$\mathbb{P}\Big(\Omega^{\mathrm{WA}}(q_n, q_n + \sqrt{d_n}, \varepsilon) \cap \Omega^{\mathrm{ZR}}(q_n, q_n + \sqrt{d_n}, \tfrac{\varepsilon}{2}, \kappa)\Big) = 1 - o(1), \tag{B.31}$$

which by (B.27) yields the desired result. □

Remark B.3.11 (Extension of Proposition B.3.10 to the degree-two graphs with rewiring). In [69] and [29, Section 5.2] it is noted that Schramm's coupling can be adapted to the setting of coagulation-fragmentation dynamics that keep the measure $\mathsf{PoiDir}(\theta)$, $\theta \in (0, 1]$, invariant. An example is a dynamic graph model with all degrees equal to two and endowed with a rewiring dynamics, which corresponds to a coagulation-fragmentation dynamics with invariant measure $\mathsf{PoiDir}(1/2)$ (see [120]). Since $\nu_\theta(L_\varepsilon) > 0$ for any $\theta, \varepsilon \in (0, 1)$ and $\nu_\theta = \mathsf{PoiDir}(\theta)$, the proof of Proposition B.3.10 can be adapted to the aforementioned situation. ♦

In conclusion of this section, we state a corollary of the previous proposition, which is useful in Section 3.3:

Corollary B.3.12. *In the setting of Proposition B.3.10,*

$$\mathbb{P}\left(\mathcal{E}_n^{\mathrm{c}}(\tfrac{T^{\Downarrow}_{n,v_0}}{n}, \varepsilon, \kappa) \cap \{T^{\Downarrow}_{n,v_0} \geq (\tfrac{1}{2} + \delta)n\}\right) = o(1). \tag{B.32}$$

Proof. First, let us rewrite the desired expression as

$$\begin{aligned}\mathbb{P}\left(\mathcal{E}_n^{\mathrm{c}}(\tfrac{T^{\Downarrow}_{n,v_0}}{n}, \varepsilon, \kappa) \cap \{T^{\Downarrow}_{n,v_0} \geq (\tfrac{1}{2} + \delta)n\}\right) \\ = \mathbb{E}\left[\mathbb{P}\left(\mathcal{E}_n^{\mathrm{c}}(\tfrac{T^{\Downarrow}_{n,v_0}}{n}, \varepsilon, \kappa) \mid T^{\Downarrow}_{n,v_0}\right) \mathbb{1}_{\{T^{\Downarrow}_{n,v_0} \geq (\frac{1}{2}+\delta)n\}}\right].\end{aligned} \tag{B.33}$$

Since the inner part of the right-hand side is bounded between 0 and 1, by dominated convergence theorem it is enough to show that

$$\mathbb{P}\left(\mathcal{E}_n^{\mathrm{c}}(\tfrac{T^{\Downarrow}_{n,v_0}}{n}, \varepsilon, \kappa) \mid T^{\Downarrow}_{n,v_0}\right) \mathbb{1}_{\{T^{\Downarrow}_{n,v_0} \geq (\frac{1}{2}+\delta)n\}} = o_{\mathbb{P}}(1). \tag{B.34}$$

The core observation is that arguments based on Schramm coupling from Proposition B.3.10 work even under the conditioning in (B.34). From the perspective of the cycle structure, on the event $\{T^{\Downarrow}_{n,v_0} \geq (\frac{1}{2} + \delta)n\}$, at time $T^{\Downarrow}_{n,v_0}$ the giant component of the associated graph process is enlarged by a connected component whose size can be, with high probability, uniformly bounded by $C \log^2 n$, for some $C > 0$. But the effect of these moves is already captured by the forbidden set construction (recall Definition B.3.6 and the discussion below Remark B.3.3), and a quantitative bound on the effect of these moves is expressed in Lemma B.3.7. Therefore, by repeating the arguments from Lemma B.3.9 and Proposition B.3.10, we obtain (B.34), which yields the desired expression.

□

§B.4 Mixing upon dropdown on the largest cycle

Recall that for two laws μ, ν defined on the same countable probability space

$$d_{\mathrm{TV}}(\mu,\nu) = \frac{1}{2}\sum_{v} |\mu_v - \nu_v|. \tag{B.35}$$

Fix $n \in \mathbb{N}$. Recall the short-hand notation $M = |\mathscr{C}_{\max}(A_{\Pi_n}(T^{\Downarrow}_{n,v_0}))|$. We wish to show that, on the events (recall (3.66))

$$\begin{aligned} \mathcal{M}_1(\varepsilon,\delta) &= \big\{|\operatorname{supp}(\mu^{n,v_0}(T^{\Downarrow}_{n,v_0}))| > \varepsilon M\big\} \cap \Omega^{(\mathrm{SC})}(T^{\Downarrow}_{n,v_0}) \cap \Omega^{(\mathrm{ER})}_n(T^{\Downarrow}_{n,v_0}), \\ \mathcal{M}_2(\varepsilon) &= \big\{\exists\, t_L \in (T^{\Downarrow}_{n,v_0}, T^{\Downarrow}_{n,v_0} + a_n)\colon\ \mathfrak{X}^{(n)}_1(t_L) > 1-\varepsilon^2\big\}, \end{aligned} \tag{B.36}$$

the following estimate (recall (3.70)) is valid:

$$d_{\mathrm{TV}}\Big(\mu^{n,v_0}(t_L), \mathsf{Unif}(\mathscr{C}_{\max}(A_{\Pi_n}(t_L)))\Big) < \varepsilon. \tag{B.37}$$

On the event $\mathcal{M}_1(\varepsilon,\delta)$, the ISRW-distributiuon at time $T^{\Downarrow}_{n,v_0}$ is uniform over a single cycle supported on the largest component of the associated graph process (due to $\Omega^{(\mathrm{SC})}(T^{\Downarrow}_{n,v_0})$), which is larger than εM (due to $\{|\operatorname{supp}(\mu^{n,v_0}(T^{\Downarrow}_{n,v_0}))| > \varepsilon M\}$) by assumption. Therefore,

$$\mu^{n,v_0}_v(T^{\Downarrow}_{n,v_0}) \begin{cases} \le \frac{1}{\varepsilon M}, & v \in \operatorname{supp}(\mu^{n,v_0}(T^{\Downarrow}_{n,v_0})), \\ = 0, & \text{otherwise.} \end{cases} \tag{B.38}$$

We will use (B.38) to provide us with an upper bound on the mass carried by the elements on the giant component of the associated graph process.

On the event $\mathcal{M}_1(\varepsilon,\delta) \cap \mathcal{M}_2(\varepsilon)$, the largest cycle at time t_L necessarily carries some of the mass of the ISRW-distribution, since the initial size of the support is larger than the size of the subset not covered by the largest cycle (see Figure B.2 for a visual explanation).

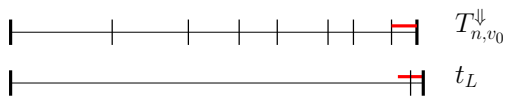


Figure B.2: The first line segment represents the cycle structure of the permutation restricted to the giant component of the associated graph process at time $T^{\Downarrow}_{n,v_0}$. The red line represents the size of the single cycle that contains the full mass of the ISRW-distribution. The second line segment represents the same at time t_L: a cycle of size $(1-\varepsilon^2)M$ has appeared, which necessarily carries some of the mass of the ISRW-distribution.

We proceed by computing the worst-case l^1-distance between the two distributions at time t_L, which will yield the desired inequality in (B.37). Put $\Delta = |\mathscr{C}_{\max}(A_{\Pi_n}(t_L))| - M$, which represents the growth of the largest component of the associated graph process compared to its size at time $T^{\Downarrow}_{n,v_0}$. Note that the uniform distribution in (B.37) gives the individual elements the following mass:

$$\mathsf{Unif}(\mathscr{C}_{\max}(A_{\Pi_n}(t_L)))(v) = \begin{cases} \frac{1}{M+\Delta}, & v \in \mathscr{C}_{\max}(A_{\Pi_n}(t_L)), \\ 0, & \text{otherwise.} \end{cases} \tag{B.39}$$

To compute the worst-case l^1-distance at time t_L, we maximise the individual summands in (B.35). To do so, we work with a distribution $\mu^\dagger$ supported on the giant component of the associated graph process that puts mass $1/(\varepsilon M)$ (see (B.36)) on the elements outside the largest cycle of size $(1-\varepsilon^2)M$, and spreads the remaining mass $\Sigma = 1 - \frac{\varepsilon^2 M + \Delta}{\varepsilon M}$ over the largest cycle. In symbols,

$$\mu^\dagger_v = \begin{cases} \frac{\Sigma}{(1-\varepsilon^2)M}, & v \in \mathfrak{X}_1^{(n)}(t_L), \\ 0, & v \notin \mathscr{C}_{\max}(A_{\Pi_n}(t_L)), \\ \frac{1}{\varepsilon M}, & \text{otherwise.} \end{cases} \tag{B.40}$$

The measure $\mu^\dagger$ constructed in this way gives the bound

$$d_{\mathrm{TV}}\Big(\mu^{n,v_0}(t_L), \mathsf{Unif}(\mathscr{C}_{\max}(A_{\Pi_n}(t_L)))\Big) \le d_{\mathrm{TV}}\Big(\mu^\dagger, \mathsf{Unif}(\mathscr{C}_{\max}(A_{\Pi_n}(t_L)))\Big). \tag{B.41}$$

Finally, we compute, using (B.35) and (B.39)–(B.40),

$$\begin{aligned} d_{\mathrm{TV}}\Big(\mu^\dagger, \mathsf{Unif}(\mathscr{C}_{\max}(A_{\Pi_n}(t_L)))\Big) &= \frac{1}{2}\left\|\mathsf{Unif}(\mathscr{C}_{\max}(A_{\Pi_n}(t_L))) - \mu^\dagger\right\|_{l^1} \\ &= \frac{1}{2}(\varepsilon^2 M + \Delta)\left[\frac{1}{\varepsilon M} - \frac{1}{M+\Delta}\right] + \frac{1}{2}(1-\varepsilon^2)M\left[\frac{1}{M+\Delta} - \frac{\Sigma}{(1-\varepsilon^2)M}\right] \\ &= \varepsilon + \frac{\Delta}{\varepsilon M} - \frac{\varepsilon^2 M}{M+\Delta} - \frac{\Delta}{M+\Delta}. \end{aligned} \tag{B.42}$$

On the event $\Omega_n^{(\mathrm{ER})}(T^{\Downarrow}_{n,v_0})$, the sizes of all but the largest component are uniformly bounded by $Cn^{2/3}$ for some $C > 0$. Therefore, for a_n growing slowly enough, more concretely $a_n = o(n^{1/3})$, it follows that $\Delta = o(M) = o(n)$. Therefore

$$d_{\mathrm{TV}}\Big(\mu^{n,v_0}(t_L), \mathsf{Unif}(\mathscr{C}_{\max}(A_{\Pi_n}(t_L)))\Big) \le \varepsilon - \varepsilon^2 + o(1) < \varepsilon, \tag{B.43}$$

where the last inequality is true for n large enough.

§B.5 Mixing in dynamic degree-two graphs

§B.5.1 Permutations and degree-two graphs

Let $\pi \in S_n$ be a permutation of $[n] = \{1, \dots, n\}$. Such a permutation admits a decomposition into distinct permutation *cycles*, and this decomposition can be used to create a mapping between permutations and graphs whose vertex degrees are all equal to two. Note that this mapping is *not* a bijection between the set of all degree-two graphs on n vertices $\mathfrak{G}_n^{\deg 2}$ and the set of all permutations of n elements S_n. Such a bijection is not possible in general: while the connected components of degree-two graphs indeed are cycles, permutation cycles also carry information about the direction of traversal within them.

Let us nonetheless construct two mappings between these sets. There is a natural bijection between $[n]$ seen as the set of permutation *elements* and the same set seen as the set of graph *vertices*. Thanks to this bijection, we will (abusing notation) make no distinction between the two sets. Moreover, given a permutation $\pi \in S_n$, we can construct a degree-two graph with n vertices as follows:

(a) Set e to be the least element of π, i.e., the element "1".

(b) Add an edge between the vertices e and $\pi(e)$.

(c) Set e to be $\pi(e)$, and repeat step 2 until the entire permutation cycle containing e is traversed.

(d) Once the entire permutation cycle is traversed, set e to be the least non-traversed element of π and repeat from step 2 onwards. If there is no such element, then the entire permutation is traversed and the algorithm terminates.

Note that the mapping induced by this algorithm is a surjection, and hence the relation "being represented by the same degree-two graph" is an equivalence relation on S_n. Equivalence classes are formed by permutations whose cycle decomposition differs only in the reversal of the cyclic order between some of the cycles. For example, take $\pi_1, \pi_2 \in S_3$ such that

$$\pi_1 = (1, 2, 3), \qquad \pi_2 = (3, 2, 1). \tag{B.44}$$

By following the algorithm described above, we see that these two permutations are represented by the same degree-two graph.

When constructing the mapping from $\mathfrak{G}_n^{\deg 2}$ to S_n, the problem of cycle orientation manifests itself again. Since all the connected components of the degree-two graph are cycles, it is trivial to represent them as permutation cycles, but we are free to choose the direction of traversal of these connected components. These discrepancies are not relevant, since none of the mathematical objects used in Chapter 3 depends on the direction of traversal of the permutation cycles. The most important property for our results is the size of the permutation cycles, or equivalently the size of the connected components.

§B.5.2 Infinite-speed random walk mixing on degree-two graphs

To connect the result of Chapter 3 to our previous work describing mixing of random walks on top of configuration models endowed with *rewiring* dynamics (see [9, 10, 11]), we note that rewiring of degree-two graphs and CFDP can easily be related to each other (recall Figure 3.7). Our techniques and results can be adapted to the setting where the underlying geometry is modelled by a graph process starting from the configuration where *all the vertices have a self-loop*, equipped with rewiring dynamics. This process has been previously studied in [120].

To state these results, let us first define the process and the underlying geometry. The following definition are an adaptation of Definitions 3.1.4 and 3.1.6:

Definition B.5.1 (ISRW on a graph). Take a sequence of graphs $(G_n(t))_{t\in\mathbb{N}_0}$ and a vertex $v_0 \in [n]$. Denote by $\boldsymbol{\mathcal{CC}}(G_n(t), v)$ the connected component of the graph $G_n(t)$ that contains the vertex v. Formally, the infinite-speed random walk (ISRW) starting from v_0 is a sequence of probability distributions $(\mu^{n,v_0}(t))_{t\in\mathbb{N}_0}$ supported on $[n]$, with initial distribution at time $t = 0$ given by

$$\mu^{n,v_0}(0) = \left(\mu_w^{n,v_0}(0)\right)_{w\in[n]}, \tag{B.45}$$

where $\mu_w^{n,v_0}(0)$, the mass at $w \in [n]$ at time $t = 0$, is given by

$$\mu_w^{n,v_0}(0) = \begin{cases} \frac{1}{|\mathcal{CC}(G_n(0),v)|}, & w \in \mathcal{CC}(G_n(0), v), \\ 0, & w \notin \mathcal{CC}(G_n(0), v), \end{cases} \tag{B.46}$$

and with distribution at later time $t \in \mathbb{N}$ given by

$$\mu^{n,v_0}(t) = \left(\mu_w^{n,v_0}(t)\right)_{w\in[n]}, \tag{B.47}$$

where

$$\mu_w^{n,v_0}(t) = \frac{1}{|\mathcal{CC}(G_n(t), w)|} \sum_{u\in\mathcal{CC}(G_n(t),w)} \mu_u^{n,v_0}(t-1). \tag{B.48}$$

Informally, ISRW *spreads infinitely fast over the connected component it resides on.* ■

Definition B.5.2 (Dynamic degree-two graph with rewiring dynamics). Fix the vertex set $\mathcal{V} = [n]$ and let $G_n(0)$ be the graph with all vertices having a single self-loop. At any later time $t \in \mathbb{N}$, $G_n(t)$ is obtained from $G_n(t-1)$ as follows:

(a) Pick two edges e_1, e_2 uniformly at random without replacement from the set of edges within $G_n(t-1)$.

(b) Break edge e_1 into half-edges h_1^A, h_1^B and edge e_2 into half-edges h_2^A, h_2^B (see [78, Section 7.2] for a definition of a half-edge).

(c) Graph $G_n(t)$ has all the unbroken edges of $G_n(t-1)$, while the broken edges e_1, e_2 are replaced by a uniform choice from 2 possible sets of new edges: $\{(h_1^A, h_2^A), (h_1^B, h_2^B)\}, \{(h_1^A, h_2^B), (h_1^B, h_2^A)\}$.

We call the sequence $(G_n(t))_{t\in\mathbb{N}_0}$ the *dynamic degree-two graph with rewiring dynamics.* ■

In this setting, we state the following theorem, analogous to Theorem 3.1.20:

Theorem B.5.3 (Mixing profile for ISRW on dynamic degree-two graphs with rewiring).

(1) *Uniformly in $v_0 \in [n]$,*

$$\frac{T_{n,v_0}^{\Downarrow}}{n} \xrightarrow{d} u^{\Downarrow}, \tag{B.49}$$

where $u^{\Downarrow}$ is the non-negative random variable with distribution (recall Definition 3.1.11(1))

$$\mathbb{P}(u^{\Downarrow} \le u) = \zeta(u), \qquad u \in [0, \infty). \tag{B.50}$$

(2) *Uniformly in $v_0 \in [n]$,*

$$\left(\mathcal{D}_n^{v_0}(un)\right)_{u\in[0,\infty)} \xrightarrow{d} \left(1 - \zeta(u)\mathbb{1}_{\{u>u^{\Downarrow}\}}\right)_{u\in[0,\infty)} \quad \textit{in the Skorokhod } M_1\textit{-topology.} \tag{B.51}$$

Proof. Since Schramm's coupling and related arguments are fully applicable in this setting (see Remark B.3.11), the arguments from Section 3.3 can be used again with only slight modifications, as outlined below:

(a) **Associated graph process and drop-down time:** The two constructions carry over without modifications. Even if a cycle-creating edge in the associated graph process produces a fragmentation only with probability 1/2 (since such an edge can with equal probability tear a segment of a cycle apart and create a new cycle or put the segment back into its original cycle in reversed order), this does not influence the arguments that underlie Lemma 3.3.2.

(b) **Fast mixing upon drop-down:** Since the proof of Proposition 3.3.10 is based on Proposition B.3.10, which can be adapted to the alternative setting (see Remark B.3.11), we conclude that Proposition 3.3.10 also carries over.

(c) **Drop-down in a single cycle:** Lemma 3.3.3 requires no adaptations.

(d) **Mixing profile:** Since all the ingredients used in the proof of Lemma 3.3.11 and Theorem 3.1.20 carry over, the proofs themselves do likewise.

□

Remark B.5.4 (ISRW on a dynamic degree-two configuration model). The setting of Theorem B.5.3 is *different* from a degree-two configuration model with rewiring, where the edges in the initial graph would be created by a random matching of half-edges. We conjecture that if the underlying geometry were modelled by a degree-two configuration model with rewiring, then ISRW would mix in $o_{\mathbb{P}}(n)$ steps and the mixing profile at scale n would be trivial. ♦

CHAPTER 4

Communication protocol for a satellite-swarm interferometer

This chapter is based on the following article:

O. Nagy, M. Pandey, G. Exarchakos, M. Bentum, and R. van der Hofstad. Communication protocol for a satellite-swarm interferometer. *arXiv preprint arXiv:2312.15814*, 2023.

Abstract

Orbiting low frequency antennas for radio astronomy (OLFAR) that capture cosmic signals in the frequency range below 30MHz could provide valuable insights on our Universe. These wireless swarms of satellites form a connectivity graph that allows data exchange between most pairs of satellites. Since this swarm acts as an interferometer, the aim is to compute the cross-correlations between most pairs of satellites. We propose a k-nearest-neighbour communication protocol, and investigate the minimum neighbourhood size of each satellite that ensures connectivity of at least 95% of the swarm. We describe the proportion of cross-correlations that can be computed in our method given an energy budget per satellite. Despite the method's apparent simplicity, it allows us to gain insight into the requirements for such satellite swarms. In particular, we give specific advice on the energy requirements to have sufficient coverage of the relevant baselines.

§4.1 Introduction

§4.1.1 Motivation and setting

Long-wavelength radio astronomy (also called Low-frequency radio astronomy) is targeting for an instrument for observing the Universe at frequencies below 30 MHz. For frequencies above 30 MHz several instruments have been implemented in the past two centuries, like LOFAR (LOw Frequency ARray) [72] in the Netherlands and its European extension ILT, the International LOFAR Telescope. However, at frequencies below 30 MHz, Earth-based observations are limited due to a combination of severe ionospheric distortions, almost full reflection of radio waves below 10 MHz, solar eruptions and the radio frequency interference (RFI) of human-made signals. Scientifically this frequency band is extremely interesting, for instance providing information from the faint signals from the Hydrogen in the Cosmological Dark Ages and Cosmic Dawn, the study of Solar activity and space weather at low frequencies, the measure of the auroral radio emission from the large planets in our Solar system, the determination of the radio background spectrum at the Earth-Moon L2 point, the creation of a new low-frequency map of the radio sky, the study of the Earth's ionosphere, and the detection of bright pulsars and other radio transient phenomena at very low frequencies [89].

Many paper studies have been performed in the last decade, to open up this last, virtually unexplored frequency domain in the electromagnetic spectrum [22, 44].

The basic idea is to form a swarm of satellites, each sampling the astronomical signals. Together, they work as an interferometer – in practice, this means that all the useful information is obtained only after *cross-correlation* of a pair of measurements from 2 different satellites.

In an ideal setting, one would carry out all the possible cross-correlations; in practice, there are limitations coming from a finite energy budget, data routing problems and others.

§4.1.2 Modelling assumptions

The problem under consideration is characterized by its inherent complexity, encompassing a multitude of complications stemming from various sources, including the intricate dynamics of the system, challenges associated with data transmission, equipment malfunctions, and other factors. It is important to emphasize that our primary objective is not to achieve a high degree of realism in the modelling process. Instead, our focus lies in conducting a feasibility study under simplified, yet realistic, assumptions. By delineating these assumptions, we aim to establish the extent of applicability of our investigation.

Number of satellites. Throughout this chapter, we denote the number of satellites in our swarm by n, and we think of n as being large. In our analysis, we thus make the explicit assumption that the number of satellites involved is large, yet remains fixed throughout the duration of the study and does not undergo any changes over time. We acknowledge, however, that allowing for dynamic changes in the satellite

constellation could introduce valuable insights into the effects of various scenarios, such as catastrophic malfunctions or the reinforcement of the satellite swarm through the deployment of additional satellites.

By considering the possibility of catastrophic malfunctions, we could model the impact of severe failures within the satellite network. These malfunctions could include critical subsystem failures, orbital anomalies, or unexpected events leading to the loss of functionality of one or more satellites. Incorporating such dynamic changes would allow us to assess the system's resilience and the overall robustness of the network in the face of unforeseen disruptions.

Furthermore, considering the deployment of additional satellites to reinforce the existing swarm introduces an important aspect of scalability and adaptability. By accounting for the potential deployment of extra satellites, we can evaluate the system's ability to respond to increasing demands, expand coverage, or mitigate the effects of satellite failures. This consideration becomes particularly relevant when exploring strategies for enhancing network reliability, improving signal coverage, or addressing future capacity requirements.

Position and movement of satellites. We adopt the assumption that the initial distribution of satellites is uniformly random within a unit cube. This means that each satellite's position is independently and uniformly chosen within the interval [0,1] for each of its three Cartesian coordinates.

The decision to employ this uniform random distribution can be interpreted as a deliberate choice not to pursue an optimization approach for satellite deployment in our study. Instead, we aim to investigate communication networks governed by *spatial* randomness. This approach allows us to capture the inherent unpredictability and diversity present in real-world deployments.

It is important to note that the assumption of static satellites imposes limitations on the applicability of our results. Our findings are most suitable for scenarios where the assumption of static satellite positions is a justifiable approximation. For instance, an illustrative example could be a satellite swarm deployed near one of the stable Lagrange points in the Earth-Sun system. In such cases, the gravitational forces and orbital dynamics may cause satellites to maintain relatively fixed positions over extended periods, rendering the static assumption reasonable. We do wish to stress that the uniformity assumption could arise as the *stationary distribution* of dynamical swarms of satellites, and thus presents a snapshot of the dynamical system. Further, our approach can be straightforwardly adapted should such a stationary distribution be non-uniform.

Structure of the communication network. In addition to the previously mentioned assumptions, another key assumption we make in our study is that the communication network between the satellites can be represented as a directed graph. Specifically, we construct this graph by connecting each vertex to its k nearest neighbours based on Euclidean distance, where k is chosen appropriately to strike a balance between network connectivity and energy consumption. This type of graph, where each vertex has directed edges connecting it to its k nearest neighbours, is commonly

referred to as a k-nearest-neighbour (k-NN) graph.

The choice to model the communication network as a k-NN graph may initially appear arbitrary. However, our research findings demonstrate that this network representation possesses numerous desirable properties when $k \geq 4$. By examining the network characteristics and performance metrics associated with k-NN graphs, we can gain insights into the behaviour and functionality of the communication system under study.

While there exist alternative network models that could be considered within this setting, such as the random geometric graph, we have intentionally limited our investigation to focus solely on k-NN graphs due to its simplicity combined with the valuable insights that it provides.

Energy expenditure. We introduce a significant simplification regarding the energy consumption of satellites. Specifically, we assume that there are only two primary ways in which a satellite can consume energy: communication-related activities and the computation of cross-correlations based on acquired measurements. This assumption, while simplifying the analysis, entails the exclusion of other potential energy expenditure factors that may exist in real-world scenarios. This simplified approach allows us to isolate and examine the energy requirements directly linked to these essential functions and evaluate their effects on system performance and efficiency.

We investigate a setting in which the total energy spent on communication is comparable to the total energy spent on computation, as this is the most interesting setting. Indeed, when either of the two energies is negligible compared to the other, we can ignore that aspect, making the problem significantly simpler. Let c denote the computation costs per cross-correlation, and $E_{\max}$ the total amount of energy at the disposal of each satellite. Furthermore, we desire that the total amount of computation power is, on average, a β-fraction of the total energy, i.e.,

$$c\binom{n}{2} = n\beta E_{\max}. \tag{4.1}$$

Without loss of generality, we may work in units such that $c = 1$. The parameters $E_{\max}$ and β will then serve as the key tuning parameters in the chapter.

Data flows in the network. We make the assumption that all data is readily available "on-demand" as long as a viable path exists between the data source and the destination. This assumption implies that once a valid communication path is established between two satellites, the necessary data can be efficiently transmitted and accessed without delay.

As our analysis progresses, we will demonstrate that, for appropriate values of k, there exists a path connecting nearly all pairs of satellites in the network. This observation underscores the connectivity and accessibility of the communication infrastructure we consider.

However, it is important to note that our assumptions disregard various factors related to data flow within the satellite network. Specifically, we neglect the effects of

routing, network capacity limitations, finite-speed propagation delays, noisy channels, and the need for retransmissions, among others.

§4.1.3 Previous works

Engineering applications of k-NN graphs

There is a plethora of challenges on wireless networks to which k-NN models have been applied. This is mostly because of its simplicity to implement, its decentralized nature, and its good performance. However, each challenge requires a slight modification of the algorithm. Here, we review k-NN based wireless topologies and transmission power control, as well as node placement, as they are the most relevant to the LOFAR connectivity problem we study.

Blough et al. [32], in one of the earliest studies on k-NN-based topology control, ensure that every node in a mesh wireless network is connected to at least k other neighbours. Each node adjusts the transmission range to maintain k bidirectional links. Supported by other studies (e.g., [31, 105]), the authors argue that the overhead generated on the routing layer to maintain a connected topology graph (a.k.a. large strongly connected component) using unidirectional links outweighs the benefits. The size of the routing tables that need to be maintained increases with unidirectional links, since any pair of nodes needs two paths to exchange information back and forth. Therefore, the vast majority of protocols in the L2 and L3 OSI layers assume bidirectional links. However, the OLFAR application does not require bidirectional exchange of interferometric observations between any pair of satellites. The cross-correlation of signals can happen at any aggregation point; hence, not all pairs of satellites need to exchange data.

Wireless coverage is another challenge to which k-NN algorithms have been applied. For instance, [90] and [84] devise algorithms to adjust the heading of mobile wireless sensors such that all sensors maintain a neighbourhood of size k indicating convergence of the group of sensors to the same direction. In a relevant problem, node placement and localization, the authors of [118, 152] use a fingerprinting method to localize a wireless device based on the received signal strength of the k nearest neighbours. Extending the same idea, but using the channel state information (CSI) instead of the signal strength, has been proposed by [136].

Finally, in resource allocation problems of 6G mobile telecommunication networks, k-NN has been used as a classification method to cluster data. In [148], k-NN is used to dynamically allocate the radiating elements of an antenna array to k users based on their spatial patterns. A set of classes of quality of service and channel state information are defined, and best resource allocation is devised. k-NN is used to assign the current channel and quality conditions of each user to one of those classes and, based on that, to allocate the radiating elements to form the necessary beams in a massive multiple-input-multiple-output (MIMO) setup. The expected response time of radio resource allocation decisions is in the range of < 1 ms [50]. In the OLFAR communications scenario, where satellites continuously change their position, the relevant resource allocation decisions need to also be performed sufficiently fast. Though we currently focus on static scenarios, the aforementioned study indicates the

benefits of k-NN even in more dynamic settings. The authors of [98] have used k-NN graphs to reduce a wireless node's complexity and calculation overhead of the best time schedule of the communication to their neighbours.

Previous studies confirm the wide adoption of k-NN graphs in various problems in wireless networks. Yet, these studies view wireless networks as undirected graphs; an unnecessary assumption for OLFAR. We extend those efforts with topology control of unidirectional orbiting swarms for OLFAR applications.

Connectivity properties of k-NN graphs

The connectivity properties of k-NN graph have received substantial attention, both in engineering and mathematical literature [1, 15, 16, 14, 52]. In [1], the behaviour of the shortest path distance in weighted k-NN graphs is studied. In [52], probabilistic properties, like the expected number of connected components of the k-NN graph for a random set of points are discussed.

In contrast to wired networks, wireless networks usually do not come with a fixed set of links between nodes; furthermore, these wireless connections need not be bidirectional. In a situation when one is given a network of wireless nodes distributed at random, e.g., in a way modelled by a Poisson process, it is natural to ask under what conditions is it reasonable to expect a "fully connected" network of nodes. Here we will consider nodes broadcasting with a fixed finite range of transmission, but we allow for these ranges to vary between the nodes. Naturally, the answer depends on the nature of links (unidirectional vs. bidirectional) and the precise meaning of the term fully connected.

The studies on connectivity properties of k-NN graphs carried out by the engineering community focus on a model where each of the links allows for bidirectional communication and these links are established between k-nearest neighbours. Originally, the research revolved around "magic numbers", that is, values of k which with high probability lead to a network connected in a sense that there exists a path between any two nodes (see e.g. [94, 117] and many others). It has been suggested that these "magic numbers" can be any of the integers between three and eight. A breakthrough came in the article [153], where the authors realized that to maintain connectivity as the area of the square in which we place the nodes (denoted by A) tends to infinity, while the intensity of the Poisson process governing the nodes remains constant, each node needs to be connected to a number of neighbours that is of the order $\log A$. This marks a departure away from the idea of a universal "magic number". Furthermore, the same article provides some bounds for the critical value, multiplying $\log A$, describing the zero-one law related to this notion of connectivity.

These results have stimulated a rigorous mathematical investigation. In a series of articles [15, 16], the authors formalize the idea that, given the setting of [153], connectivity is indeed obtained when k is of the order $\log A$, and first provide a much tighter bound on the critical multiplication constant ($0.3043 \leq c^{\star} \leq 0.5139$). Later, they invent a method to compute this critical constant precisely. Further results in these directions are for example [14].

All of these results concern the case with bidirectional links, which is *not* our setting. For a directed version of this problem, where connectivity is given by an existence of a *directed* path between any two nodes, the article [15] presents behaviour

similar to the undirected model, that is connectivity threshold at $k = c^\star \log A$, with $0.7209 \leq c^\star \leq 0.9967$. It should be noted that these results concern asymptotic behaviour when n grows very large, which limits their direct applicability.

§4.2 Communication network within the swarm

§4.2.1 Connectivity & strongly connected components

As outlined in Section 4.1.2, we assume that the communication network inside the swarm is modelled as a directed graph. Ideally, one would have a network that allows for communication between *any* two satellites. In mathematical terms, this would mean that the *largest strongly connected components* (LSCC for short) is the entire k-NN graph; i.e., there exists a (directed) path between any two vertices. On the other hand, such a strong guarantee in a spatially random network could be quite energetically demanding.

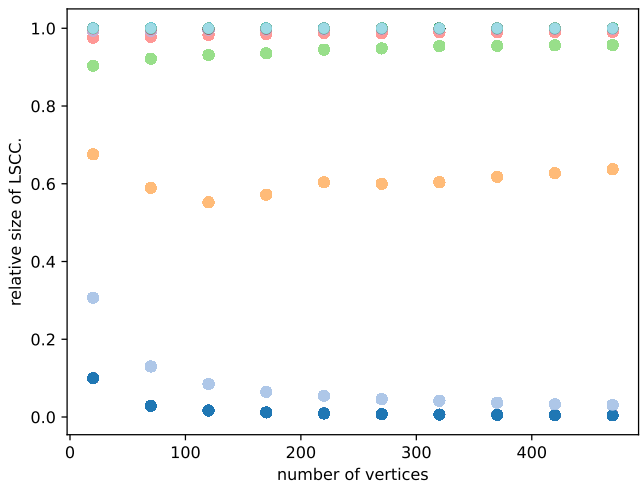


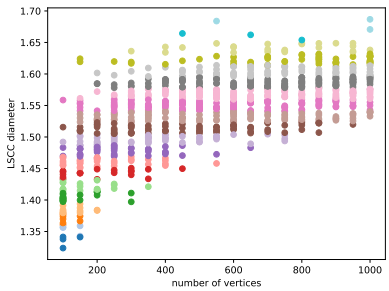
Figure 4.1: Mean fraction of vertices in the LSCC for different sizes of the graph and $k \in \{2, \dots 12\}$.

In this section, we show simulations that shed light on the effect of the choice of the parameter k on the size of the LSCC. Our results are plotted in Figure 4.1. This data shows that, in the setting of this study, it makes little sense to consider values of $k \leq 3$, since there the connectivity is simply too low. Starting with $k = 4$, we see that the LSCC typically spans at least $\sim 90\%$ of vertices in the graph. While this choice already provides a network that behaves in the required manner, it appears that the optimal choice of k would be $k = 5$, since, on the one hand, there is little to no improvement in the typical size of the LSCC for larger values of k, while larger values of k lead to higher energy requirements for maintaining communication.

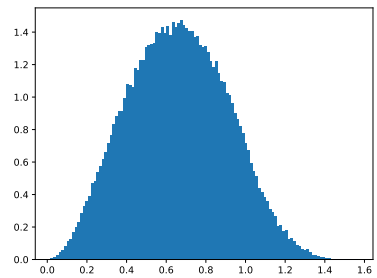
§4.2.2 Empirical distribution of baseline lengths

To obtain high-quality data from an interferometric observation, it is desirable to have target observations made with various baselines. In this section, we study the length of the longest baseline within the LSCC and the distribution of baseline lengths. Note that given our assumption about the geometry, the longest possible baseline has length $\sqrt{3} \approx 1.73$ – this corresponds to the diagonal across the unit cube. In our results, we confine ourselves to the parameter regime $n = 0 - 1,000$, and $k \in \{4, 5, 6\}$.

In Figure 4.2a, one can see the scatter plot of the lengths of longest baselines within the LSCC for $k = 5$. In the regime of small n's ($n \lesssim 200$), irrespective of the value of k, the length of the maximal baseline within the LSCC varies considerably. On the other hand, for swarms with at least ~ 200 satellites, we observe less variance around the mean. Scatter plots for $k = 4, 6$ can be found in Appendix C.



(a) Scatter plots of maximal baseline lengths for $k = 5$



(b) Histogram of baseline lengths for $n = 560$, $k = 5$.

Figure 4.2: Illustrative results related to baseline distribution.

Not only the maximal baseline is important for observation, but also their distribution. Our simulations show that the distribution of baseline lengths within the LSCC is well spread-out and spans all the way from very short to almost maximal possible lengths. A representative example of this distribution can be seen in Figure 4.2b. More histograms can be found in Appendix C.

§4.2.3 Distribution function for communication costs

Recall from Section 4.1.2 that we have n satellites distributed uniformly in the cube $[0, 1]^3$. Let U be a uniformly chosen satellite. The locations of the satellites close to U can be well approximated by those in a homogeneous Poisson Point Process (PPP) of intensity measure n times the Lebesgue measure. For such a PPP, the number of satellites in a region A of volume $\mathrm{Vol}(A)$ is Poisson with parameter $n\mathrm{Vol}(A)$, while the number of satellites in disjoint regions is independent. Indeed, for a PPP, the total number of points is Poisson with parameter n, rather than equal to n, while the point locations are independently uniformly distributed. Since the Poisson distribution is highly concentrated, a Poisson distribution with parameter n is quite close to n, so it makes little difference to work with a Poisson number of parameter n or with precisely

n. When rescaling distances by $n^{1/3}$, the PPP becomes of unit intensity, allowing to describe the local environment of the uniform satellite U. While this accurately describes the local environment of satellites in the interior of the cube, for satellites closer to the boundary, this obviously creates some inaccuracies that we aim to describe using finite-size corrections.

Let the cost of transmission be $P_{0,t}(k)$, defined as

$$P_{0,t}(k) := R_k^2, \tag{4.2}$$

where R_k is the distance to the kth nearest neighbour. Let $X_r = |B_r(U)|$ denote the number of points in the ball of radius r around the point U. By definition of a Poisson point process, X_r is distributed as a Poisson random variable with parameter equal to the volume of a ball times a parameter which depends on n as $4\pi r^3/3n$.

By this observation, we can compute the probability that $P_{0,t}(k) > c$, for $c > 0$, as

$$\begin{aligned}\mathbb{P}(P_{0,t}(k) > c) &= \mathbb{P}(X_{\sqrt{c}} \leq k-1) = \sum_{j=0}^{k-1} \mathbb{P}(X_{\sqrt{c}} = j) \\ &= \sum_{j=0}^{k-1} \mathrm{e}^{-\lambda(c)} \frac{\lambda(c)^j}{j!},\end{aligned} \tag{4.3}$$

where $\lambda(c) = 4\pi n c^{\frac{3}{2}}/3$.

It follows that the cumulative distribution function (CDF) of $P_{0,t}(k)$ is given by

$$\mathbb{P}(P_{0,t}(k) \leq c) = 1 - \mathrm{e}^{-\lambda(c)} \sum_{j=0}^{k-1} \frac{\lambda(c)^j}{j!}. \tag{4.4}$$

To obtain the probability density function, we differentiate (4.4) with respect to c, to obtain

$$\begin{aligned}f_{P_{0,t}(k)}(c) &= \mathrm{e}^{-\lambda(c)} \sum_{j=0}^{k-1} \frac{\lambda(c)^j}{j!} - \mathrm{e}^{-\lambda(c)} \sum_{j=0}^{k-2} \frac{\lambda(c)^j}{j!} \\ &= \frac{\lambda(c)^{k-1}}{(k-1)!} \mathrm{e}^{-\lambda(c)} \lambda'(c).\end{aligned} \tag{4.5}$$

Substituting $\lambda(c) = 4\pi n c^{\frac{3}{2}}/3$ gives

$$f_{P_{0,t}(k)}(c) = \frac{3}{2} \left(\frac{4\pi n}{3}\right)^k \frac{c^{\frac{3k}{2}-1}}{(k-1)!} \mathrm{e}^{-\frac{4\pi n c^{3/2}}{3}}. \tag{4.6}$$

Recall that the probability distribution function of the *generalized* gamma distribution is given by

$$f_{GG}(x, a, d, p) = \frac{\left(\frac{p}{a^d}\right) x^{d-1} \mathrm{e}^{-\left(\frac{x}{a}\right)^p}}{\Gamma\left(\frac{d}{p}\right)}, \tag{4.7}$$

where $d, p > 0$ are so-called shape parameters, and $a > 0$ is the scale parameter. Comparing (4.6) with (4.7), we see that the transmission costs are distributed as a generalized gamma distribution with parameters

$$f_{P_{0,t}(k)}(c) = f_{GG}\left(c, \left(\frac{4\pi n}{3}\right)^{-2/3}, \frac{3k}{2}, \frac{3}{2}\right). \tag{4.8}$$

This is the approximate transmission cost density for large numbers of satellites, to which we can compare our simulations that involve finite-size corrections, as discussed in more detail in the next section. In particular, (4.8) implies that the transmission cost of a satellite $P_{0,t}(k)$ scales as

$$n^{2/3} P_{0,t}(k) \xrightarrow{d} P_k, \tag{4.9}$$

where P_k is distributed according to the generalized gamma distribution in (4.8) with $n = 1$.

After having established the distribution of the powers of the swarm of satellites, observe that the maximal energy of the satellites is given by $E_{\max}$, which is a finite constant. The generalized gamma distribution, however, has an unbounded support, which means that not all connections present in the k-NN graph can actually be established by the swarm. In practice, the empirical distribution of transmission costs will therefore be supported only on the finite interval $[0, E_{\max}]$ and one could approximate it with a truncated distribution which has density

$$f^T_{P_{0,t}(k)}(c) = f_{P_{0,t}(k)}(c) \tag{4.10}$$

for $c \in [0, E_{\max}]$, and is equal to $E_{\max}$ with probability $\int_0^{E_{\max}} f_{P_{0,t}(k)}(s)\,\mathrm{d}s$.

In Section 4.3, where we study the performance of our communication and computation system, we investigate the LSCC of the realized connections of the swarm as a function of the key tuning parameters β and $E_{\max}$.

§4.2.4 Finite-size corrections

The distribution function for communication costs derived in the previous section was obtained in the limit of large numbers of satellites. Effectively, this ignores the fact that the cube has a *boundary*. This is in contrast with the setting at hand, where we assume that all the satellites are restricted to be within a unit cube, thus introducing *finite-size corrections*. Unfortunately, the error introduced by these finite-size corrections is significant when considering hundreds of deployed satellites.

To remedy this error, we argue that these finite-size corrections will predominantly appear in *one* of the parameters of the class of generalized gamma distributions. Furthermore, we estimate this parameter numerically and provide a simple empirical formula that captures these corrections in the studied regime.

As one can see in Figure 4.3, empirical distributions obtained via simulation are qualitatively similar to the ones obtained in the previous section, but there is a significant quantitative disagreement.

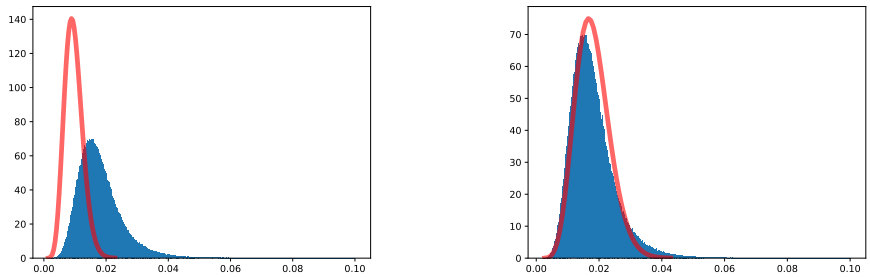


Figure 4.3: Comparison of the simulated transmission cost histograms with the generalized gamma distribution; (a) with and (b) without finite- size corrections, for $n = 600$ and $k = 5$.

By examining the parametrization of the distributions obtained in the previous section, it is intuitive that the finite-size corrections have to be applied predominantly to the parameter a, since it reflects the intensity of the considered PPP. In the case of configurations with a low number of satellites, one could argue that each of the satellites has a different *effective* intensity of the Poisson random variable describing its neighbourhood. While a direct symbolic computation seems infeasible, it is possible to *estimate* the parameter a (and hence also the effective intensity) by fitting the family of generalized gamma distribution with all parameters except for a fixed to values obtained in Section 4.2.3, and estimating the parameter a. This way, we obtain a correction in the form

$$a_{\text{corr}} = (0.685 \pm 0.002)n^{-0.73\pm0.01}, \tag{4.11}$$

which is valid in the regime of $n \in [100, 1,000]$, $k \in \{4, 5, 6\}$. By comparing generalized gamma distributions with a_{corr} as the scale parameter with empirical distributions obtained from simulations (see Figure 4.3), we see that this simple correction leads to much better approximation of empirical results.

§4.3 Applications to computation allocation

§4.3.1 Cross-correlations in the LSCC

In this section, we describe an algorithm to divide the computation of the cross-correlations of the satellites in the largest strongly connected component (LSCC). First, we investigate how the satellites can decide whether they are in the LSCC. After this, we determine how the satellites can divide the computations of the cross-correlations. Due to the choice of k in our k-NN graph, the LSCC contains at least 95% of the satellites.

By the definition of a strongly connected component, all the satellites in the LSCC have the same satellites in their in-component, and thus have the same in-component sizes. The satellites outside the LSCC will either have more satellites in their in-component compared to the LSCC in-component (for those in the out-component

of the LSCC) or much less (less than 5%). When they are in the out-component of the LSCC, their in-component is strictly larger than the LSCC. Otherwise, their in-component will be at most the complement of the LSCC. Thus, the mode of the in-component sizes (the one with the highest frequency) can be used to identify which satellites are in the LSCC. Based on their in-component sizes, each of the satellites can thus be classified into three types:

(a) **Satellites in the LSCC:** These are the satellites whose in-component size equals the mode of the in-component sizes.

(b) **Satellites in out-component of the LSCC minus the LSCC:** These are satellites whose in-component size is strictly larger than this mode.

(c) **Other satellites:** The size of in-component is less than 5%.

When the LSCC is large, the satellites of the type 1 and 2 are the most significant, as they have the most data. We perform the cross-correlations between all pairs of these satellites. The number of satellites of type 2 and 3 is at most 5% of the total satellites because they lie outside the LSCC.

After identifying the LSCC, we next investigate how to assign the cross-correlation computations optimally to the satellites, by proposing a strategy based on the energy distribution of the satellites. For this, we *label* the satellites.

The idea is that each satellite will communicate their energy surplus and their labels to the other satellites. Based on this information, the first labelled satellite gets the first few cross-correlations, the second labelled satellite gets the cross-correlations after the point where the first satellite finished, etc.

We denote the LSCC by $C_{(1)}$. We write the labels of the satellites in the LSCC as $\{1, 2, \ldots, m\}$, where $m = |C_{(1)}|$. We let the surplus energy of the ith satellite be $E_i = E_{\max} - P_i$, where $E_{\max}$ is the initial energy given to each satellite and P_i is the power needed by the ith satellite to communicate to its k nearest neighbours.

We denote the out-component and in-component of the LSCC by $C^+_{(1)}$ and $C^-_{(1)}$, respectively. We index the set of all $M = \binom{|C^-_{(1)}|}{2}$ cross-correlation pairs lexicographically as

$$C = \{c_1, c_2, \ldots, c_M\}.$$

Each element of C is an ordered pair denoting a cross-correlation. In the LSCC, we compute these cross-correlations in ascending order. The satellite with the smallest label in the LSCC does the initial cross-correlations with all its residual energy. The second smallest labelled satellite now performs the cross-correlations from where the first satellite left, and so on. Each satellite computes with all its residual energy.

In more detail, we assign the first $\lfloor E_1 \rfloor$ cross-correlations, that is, $\{c_1, c_2, \ldots, c_{\lfloor E_1 \rfloor}\}$ to the satellite with label 1. Satellite 2 performs the next $\lfloor E_2 \rfloor$ cross-correlations, that is, $\{c_{\lfloor E_1 \rfloor + 1}, \ldots, c_{\lfloor E_1 \rfloor + \lfloor E_2 \rfloor}\}$ and so on. This goes on until either all the cross-correlations are computed, or all satellites have done their computations.

When there is surplus energy remaining after the cross-correlations between the satellites in the in-component of the LSCC have all been computed, we use the surplus energy of the satellites in the out-component to compute their cross-correlations with the satellites in the LSCC's in-component. We write the labels for the satellites in the

out-component of LSCC minus LSCC as $\{m+1, m+2, \ldots, l\}$, where $l = |C^+_{(1)}|$. Then the total coverage of the cross-correlations $T(n)$ satisfies

$$T(n) = \sum_{i=1}^{|C^+_{(1)}|} \lfloor E_i \rfloor \wedge \binom{|C^-_{(1)}|}{2}, \tag{4.12}$$

and the proportion of cross-correlations α satisfies

$$\alpha = \frac{1}{\binom{n}{2}} \sum_{i=1}^{|C^+_{(1)}|} \lfloor E_i \rfloor \wedge \frac{\binom{|C^-_{(1)}|}{2}}{\binom{n}{2}}. \tag{4.13}$$

By Section 4.2.3,

$$P_i \sim GG\left(\left(\frac{4\pi n}{3}\right)^{-2/3}, \frac{3k}{2}, \frac{3}{2}\right), \tag{4.14}$$

where $GG(a, d, p)$ is the generalised gamma distribution with density given by (4.7). Since the empirical distribution of the transmission costs converges, for large n, we expect that

$$\alpha \approx \frac{2\mathbb{E}[E]}{n} \times \frac{|C^+_{(1)}|}{n} \wedge \left(\frac{|C^-_{(1)}|}{n}\right)^2, \tag{4.15}$$

where E has the same probability distribution as $E_i = \max\{E_{\max} - P_i, 0\}$, given by

$$E_i \stackrel{d}{=} \left(E_{\max} - GG\left(\left(\frac{4\pi n}{3}\right)^{-2/3}, \frac{3k}{2}, \frac{3}{2}\right)\right)_+, \tag{4.16}$$

where $x_+ = \max\{x, 0\}$.

We can rewrite (4.15) as

$$\alpha \approx \eta_+ \frac{2\mathbb{E}[E]}{n} \wedge \eta_-^2, \tag{4.17}$$

where η_+ and η_- are the proportions of satellites in the out- and in-component of the LSCC, respectively. This gives us our final formula for the total proportion of cross-correlations that can be computed in our communication and computation network, and is the main performance parameter of our system.

§4.3.2 Simulation study of cross-correlation coverage

To study the cross-correlation coverage in the k-NN network, and to demonstrate the utility of results derived in Sections 4.2, we simulated the following scenario:

(a) Initially, all the n satellites are assigned an energy budget $E_{\max} = q_{GG}(p)$, where $q_{GG}(p)$ is the quantile function of the density in (4.9) in Section 4.2.3 with finite-size correction described in Section 4.2.4, and p is a simulation parameter.

(b) We establish a *pruned* k-NN communication network, which is created from k-NN graph between the satellites, with edges removed when their existence causes the satellite's power consumption to exceed the energy budget $E_{\max}$. We iteratively remove the most energetically expensive edges, until the edges no longer deplete the power budget.

(c) We set the energy per one cross-correlation computation as in (4.1), and assign computation jobs as described in Section 4.3.1.

The choice of $E_{\max}$ and β strongly influences the outcome of the simulation. For large n, one needs to choose $E_{\max}$ unrealistically large to avoid all pruning. Thus, the quantile function $q_{GG}(p)$ allows us to pick $E_{\max}$ in a way that will very probably induce network pruning. The intensity of this effect can be effectively controlled by the parameter p (where we note that pruning is independent of β). The parameter β influences the intensity of *competition* for energy between the transmission and computational functions of satellites.

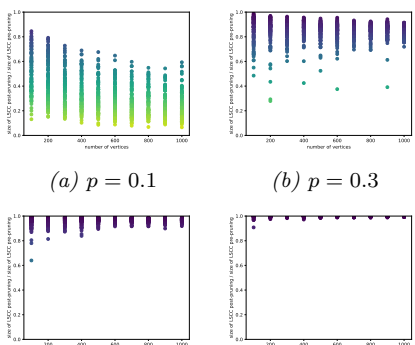
In our simulations, we focus on three quantities:

(a) LSCC reduction factor $\rho_L = \frac{|\text{LSCC post-pruning}|}{|\text{LSCC pre-pruning}|}$,

(b) LSCC coverage factor $\alpha_L = \frac{|\text{jobs assigned within LSCC}|}{|\text{jobs available within LSCC}|}$,

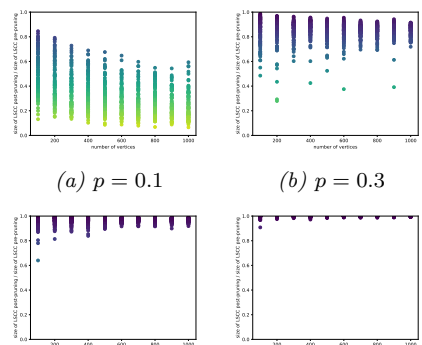
(c) Coverage factor $\alpha = \frac{|\text{jobs assigned within LSCC}|}{|\text{jobs available within swarm}|}$.

Our results indicate that these 3 quantities are highly sensitive to the choice of parameters, with some regimes exhibiting results concentrated in a small interval and others dominated by inherent randomness. Some of these results are shown below, see Figures 4.4, 4.5 and 4.6. For a more detailed discussion, see Appendix C.



(a) $p = 0.1$ *(b)* $p = 0.3$ *(c)* $p = 0.6$ *(d)* $p = 0.99$

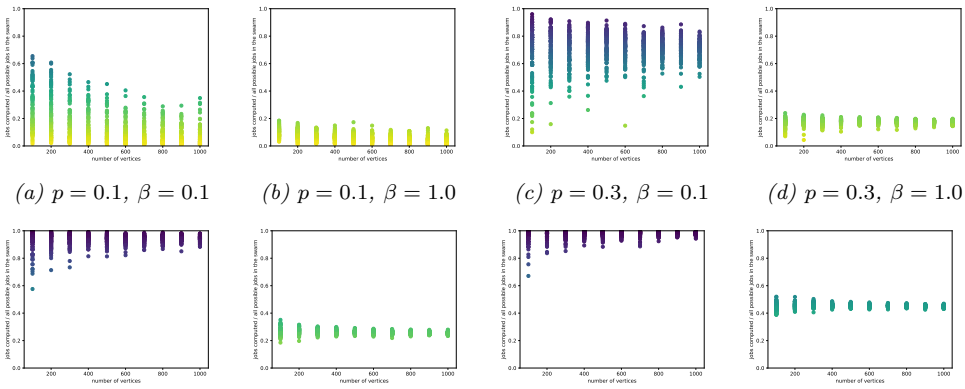
Figure 4.4: Examples of observed values of the LSCC reduction factor ρ_L for $k = 5$, $n \in (100, 1000)$, p varying.



(a) $p = 0.1$ *(b)* $p = 0.3$ *(c)* $p = 0.6$ *(d)* $p = 0.99$

Figure 4.5: Examples of observed values of the LSCC coverage factor α_L for $k = 5$, p, β varying.

In engineering terms, the results lead to an algorithm for the parameter choices. We assume that n, the size of the satellite swarm, is known. We pick k such that the LSCC is sufficiently large. For $k = 5$, the LSCC contains at least 95% of the vertices, while it is at least 98% for $k = 6$. With this k, we can then compute the power distributions over the swarm (including finite-size corrections). Further, the value of the computation cost c (in this study rescaled to be 1) fixes the product $\beta E_{\max}$ through (4.1). In turn, a value of $E_{\max}$ directly translates into a value of p through the relation $E_{\max} = q_{GG}(p)$, where $q_{GG}(p)$ is the quantile function associated to the density in (4.9). We then need to choose the pair β and p such that the desired coverage α is



(a) $p = 0.1$, $\beta = 0.1$ *(b)* $p = 0.1$, $\beta = 1.0$ *(c)* $p = 0.3$, $\beta = 0.1$ *(d)* $p = 0.3$, $\beta = 1.0$

(e) $p = 0.7$, $\beta = 0.1$ *(f)* $p = 0.7$, $\beta = 1.0$ *(g)* $p = 0.99$, $\beta = 0.4$ *(h)* $p = 0.99$, $\beta = 1.0$

Figure 4.6: Examples of observed values of the LSCC coverage factor α_L for $k = 5$, p, β varying.

reached. Our analysis provides insight into the relation between the performance α and these parameters.

§4.4 Conclusion and future work

We proposed a communication protocol for satellites, in which satellites choose their power so that they can communicate with their k nearest neighbours. Already for $k = 5$, the largest strongly connected component (LSCC) contains more than 95% of the satellites, thus ensuring sufficient coverage of the baselines needed. We computed the asymptotic distribution of the powers needed by the swarm of satellites, which is described by a generalized gamma distribution. This approximation follows by describing the local neighbourhoods of satellites in terms of a homogeneous Poisson process. Simulations confirm that this approximation already holds reasonably well, even when only a few hundred of satellites are considered after applying an appropriate finite-size corrections.

Based on the approximation of the communication costs, we proposed an algorithm to guarantee that a large proportion of cross-correlations is computed. This algorithm relied on the assumption that all satellites have access to all signals in the strongly connected component. We computed the *coverage* of the cross-correlations that can be computed under the assumption that the total computation costs are a proportion β of the total energy available to the satellites.

Some extensions of our work, for future research, are as follows:

(a) We compute the power that satellites need to communicate with their k nearest-neighbour satellites, Should the satellites be able to signal that they are not in the SCC, then they could increase their power so that they connect to more servers. While this increases their power level, this enhances connectivity. Since the satellites outside the (in-component of) the SCC are not heard at all, it may

be worth for them to raise their power. Unfortunately, this does not help the satellites in the out-component of the SCC (as they hear all the signals in the SCC), but it may help satellites that are even outside of that.

(b) What buffer allocation is needed in order for all $\binom{n}{2}$ cross-correlated pairs to be computed with high probability? Determining this threshold is probably mathematically challenging.

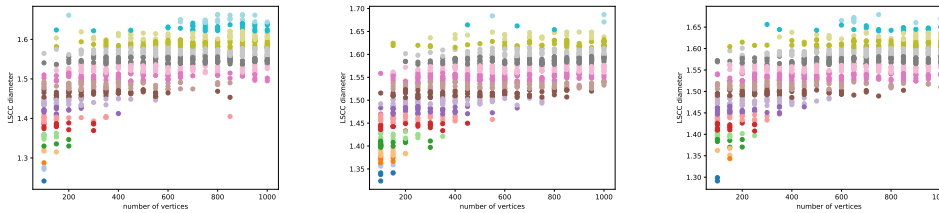
APPENDIX C

Appendices of Chapter 4

The following contains a more detailed discussion of some of our results and additional plots omitted from Chapter 4.

§C.1 Scatter plots of longest baselines

In Section 4.2.2 we have shown the scatter plots of the observed longest baselines for $k = 5$. Here we present scatter plots for all $k = 4, 5, 6$.



(a) $k = 4$ *(b)* $k = 5$ *(c)* $k = 6$

Figure C.1: Scatter plot of lengths of longest baselines for $n = 100 - 1000$*,* $k = \{4, 5, 6\}$*.*

§C.2 Histograms of empirical distributions of baseline lengths

In Figures C.2, C.3, C.4 we show histograms of empirical distributions of baseline lengths omitted from Section 4.2.2.

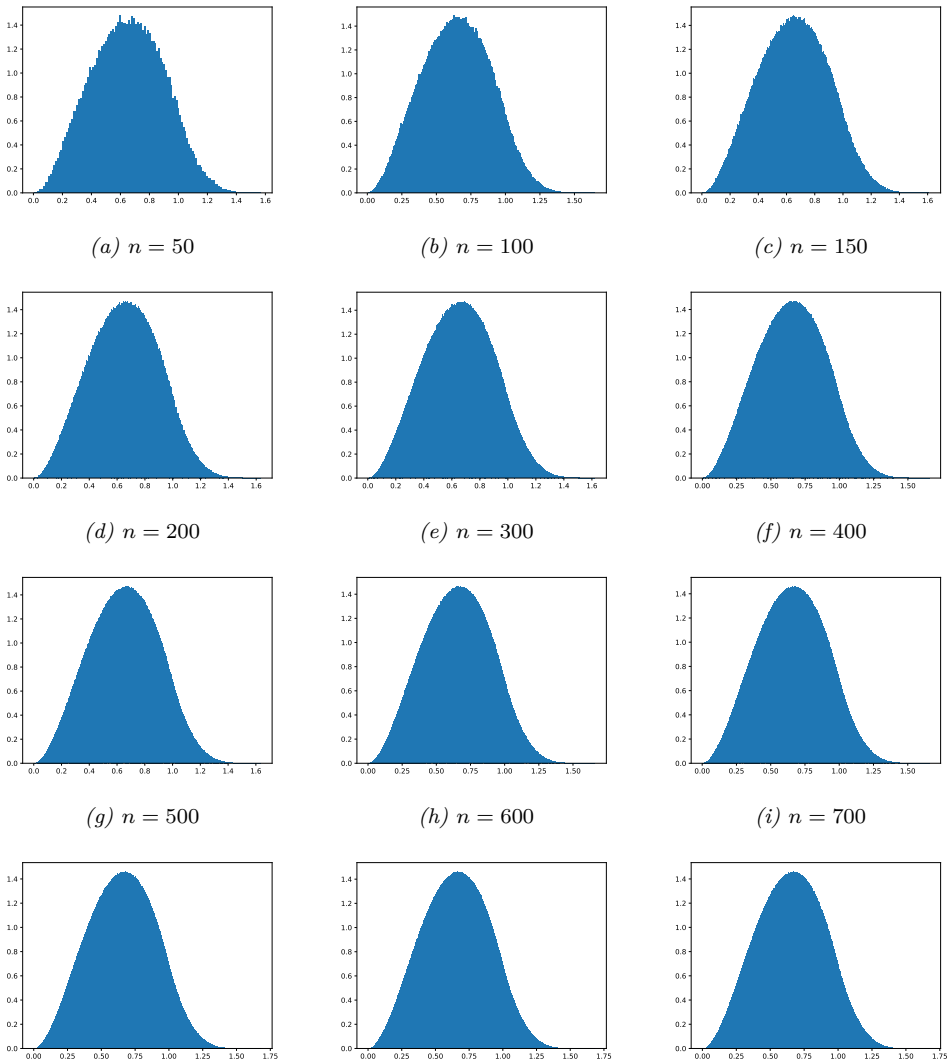
§C.3 Simulation study of cross-correlation coverage

§C.3.1 Introduction and setting

This section contains a more detailed discussion of the simulations presented in Section 4.2. For the reader's convenience, we first recall the setting of our simulation experiment. We are interested in the behaviour of an idealized model of the communication network within a swarm of satellites acting as a space-based interferometer. This network is generated based on a k-NN graph between the individual satellites. More concretely, the network is created as follows:

(a) Initially, all the n satellites are assigned an energy budget $E_{\max} = q_{GG}(p)$, where $q_{GG}(p)$ is the quantile function associated with the distribution function identified in Section 4.2.3 with finite-size correction described in Section 4.2.4 and p is a parameter of the simulation.

(b) We establish a *pruned* k-NN communication network, which is created from k-NN graph between the satellites, with edges removed when their existence causes the satellite's power consumption to exceed the energy budget $E_{\max}$. We iteratively



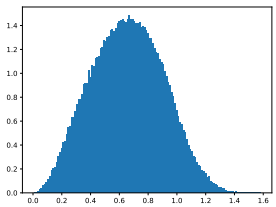
(a) $n = 50$ *(b)* $n = 100$ *(c)* $n = 150$

(d) $n = 200$ *(e)* $n = 300$ *(f)* $n = 400$

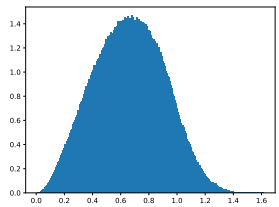
(g) $n = 500$ *(h)* $n = 600$ *(i)* $n = 700$

(j) $n = 800$ *(k)* $n = 900$ *(l)* $n = 1000$

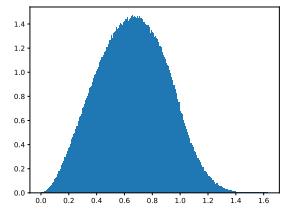
Figure C.2: Histograms of the empirical densities of baseline lengths for various n and $k = 4$.



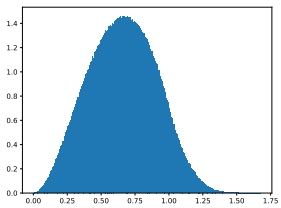
(a) $n = 50$



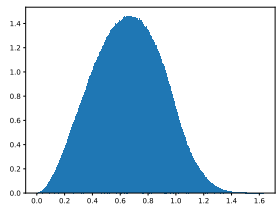
(b) $n = 100$



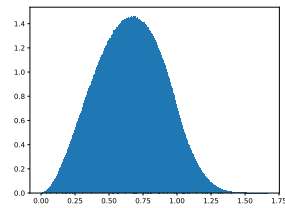
(c) $n = 150$



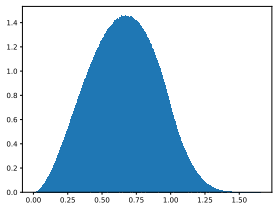
(d) $n = 200$



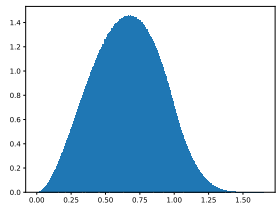
(e) $n = 300$



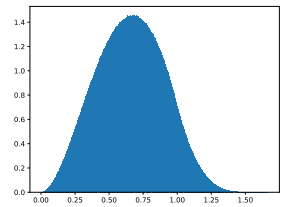
(f) $n = 400$



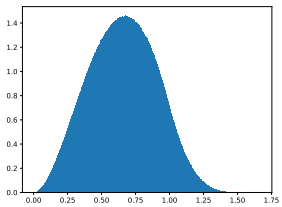
(g) $n = 500$



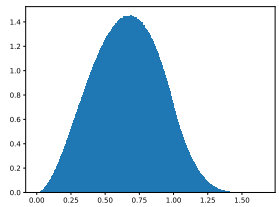
(h) $n = 600$



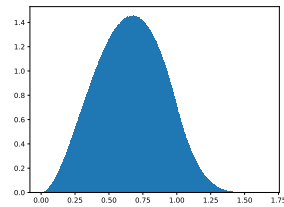
(i) $n = 700$



(j) $n = 800$

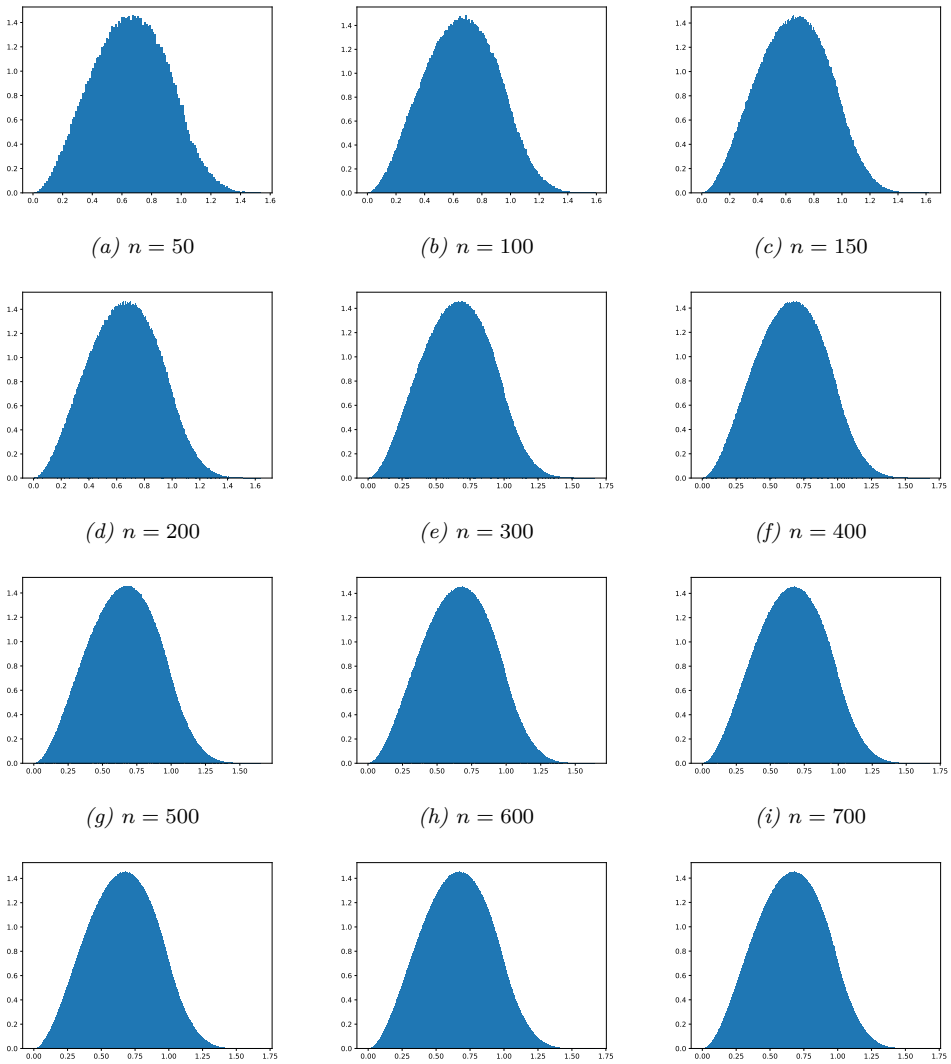


(k) $n = 900$



(l) $n = 1000$

Figure C.3: Histograms of the empirical densities of baseline lengths for various n and $k = 5$.



(a) $n = 50$ *(b)* $n = 100$ *(c)* $n = 150$

(d) $n = 200$ *(e)* $n = 300$ *(f)* $n = 400$

(g) $n = 500$ *(h)* $n = 600$ *(i)* $n = 700$

(j) $n = 800$ *(k)* $n = 900$ *(l)* $n = 1000$

Figure C.4: Histograms of the empirical densities of baseline lengths for various n and $k = 6$.

remove the most energetically expensive edge, until the edges no longer deplete the power budget.

(c) We set the energy per one cross-correlation computation as in Eq. 4.1 and assign computation jobs as described in Section 4.2.1.

For a more detailed discussion of the setting and motivation of our study, we refer the reader back to Chapter 4.

The choice of $E_{\max}$ and β strongly influences the outcome of the simulation. Since we aim to explore disconnected networks, the quantile function $q_{GG}(p)$ allows us to pick $E_{\max}$ in a way that will very probably induce network pruning and the intensity of this effect can be easily controlled by the parameter p (note that pruning is independent of β). The parameter β influences the intensity of *competition* for energy between transmission and computational functions of satellites.

In our simulations, we focus on three quantitative indicators:

(a) LSCC reduction factor $\rho_L = \frac{|\text{LSCC post-pruning}|}{|\text{LSCC pre-pruning}|}$,

(b) LSCC coverage factor $\alpha_L = \frac{|\text{jobs assigned within LSCC}|}{|\text{jobs available within LSCC}|}$,

(c) Coverage factor $\alpha = \frac{|\text{jobs assigned within LSCC}|}{|\text{jobs available within swarm}|}$.

In this document, we present a more detailed account of our results and discuss the influence of various parameters on the aforementioned quantities.

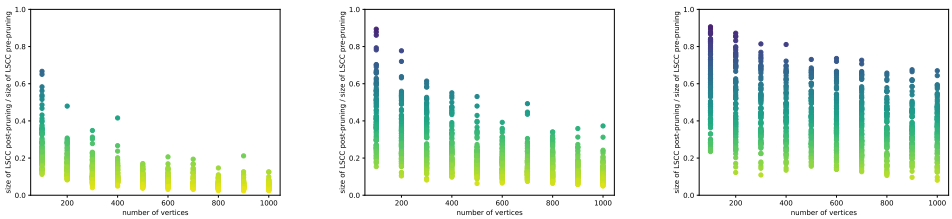
§C.3.2 Largest strongly connected component reduction factor

Firstly, note that the LSCC reduction factor ρ_L is *independent* of the parameter β, since the energy costs of initialisation and maintenance of the communication network are, in our model, given preference to those related to computation. In other words, the primary aim of every satellite is to establish its k-NN communication network, not to conserve power for possible computations.

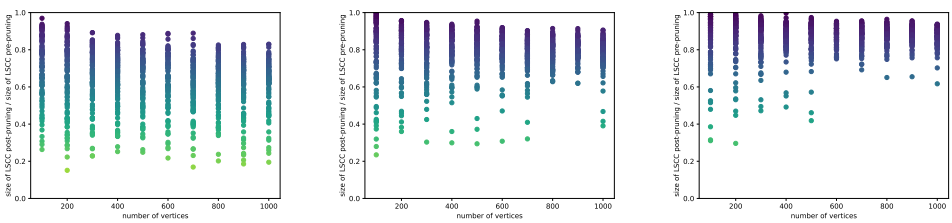
In Sections 4.2.3 and 4.2.4, we identified a closed-form expression for the probability distribution that is a good approximation of communication costs in the k-NN network. We use the quantile function $q_{\mathrm{GG}}(p) : (0, 1) \to \mathbf{R}^+$ associated to this distribution (that is, the inverse of the cumulative distribution function) to set the parameter $E_{\max}$ that corresponds to the power budget of a single satellite. As one would expect, the pruning has a more pronounced effect when p is small. In this setting, the results also show significant random fluctuations. As the parameter p increases, the LSCC reduction factor ρ_L is being pushed towards 1 (which would correspond to no reduction in the size of the LSCC) and results start to concentrate in the neighbourhood of 1.

We would also like to point out the effect of parameters n and k. As we have seen throughout the main paper, the randomness in the network is more significant in the regime of small n. Similarly, networks with higher values of the parameter k seem more robust, due to the presence of additional edges.

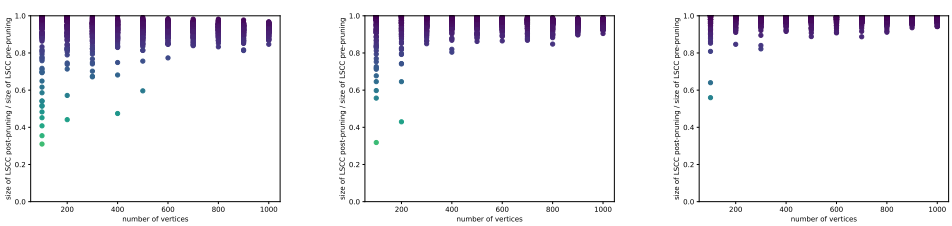
Some of our results are summarised in Figures C.5, C.6 and C.7.



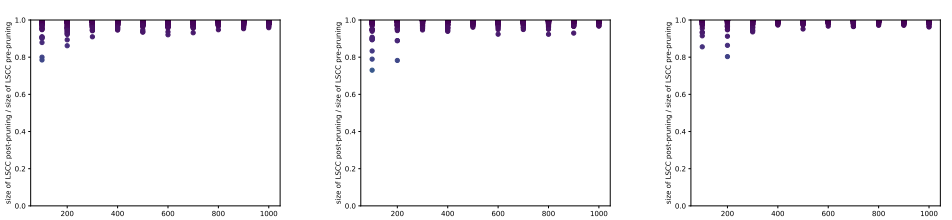
(a) $k = 4,\ p = 0.1$ *(b)* $k = 4,\ p = 0.2$ *(c)* $k = 4,\ p = 0.3$



(d) $k = 4,\ p = 0.4$ *(e)* $k = 4,\ p = 0.5$ *(f)* $k = 4,\ p = 0.6$

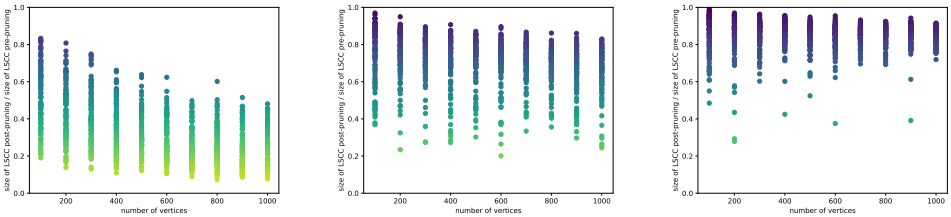


(g) $k = 4,\ p = 0.7$ *(h)* $k = 4,\ p = 0.8$ *(i)* $k = 4,\ p = 0.9$

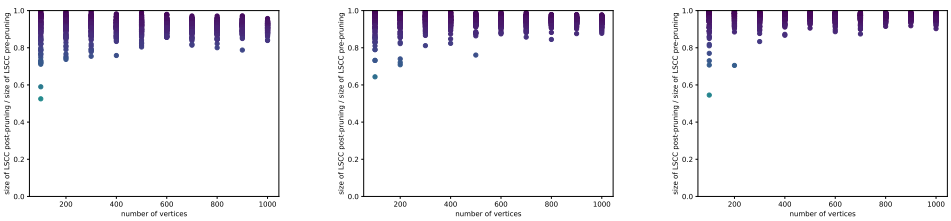


(j) $k = 4,\ p = 0.95$ *(k)* $k = 4,\ p = 0.97$ *(l)* $k = 4,\ p = 0.99$

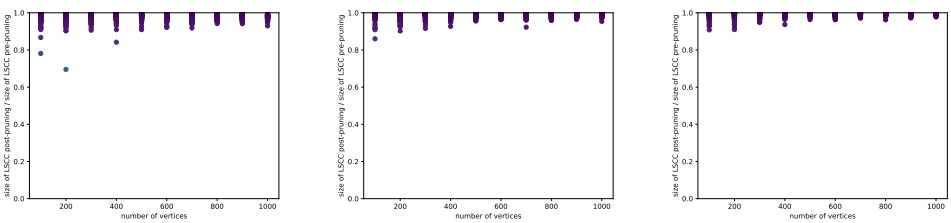
Figure C.5: Scatter plots of LSSC reduction factor ρ_L obtained from simulations in the parameter regime $n \in (100, 1000)$ for $k = 4$.



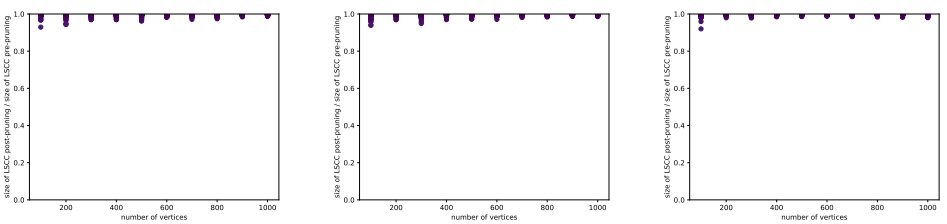
(a) $k = 5,\ p = 0.1$ *(b)* $k = 5,\ p = 0.2$ *(c)* $k = 5,\ p = 0.3$



(d) $k = 5,\ p = 0.4$ *(e)* $k = 5,\ p = 0.5$ *(f)* $k = 5,\ p = 0.6$

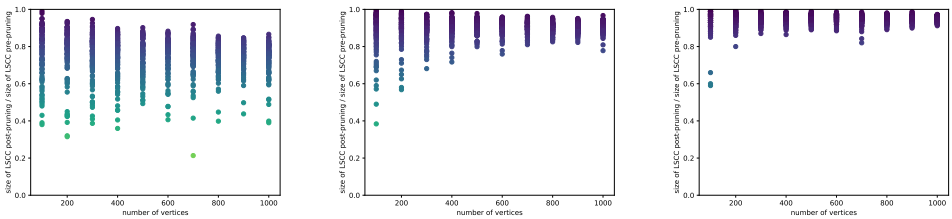


(g) $k = 5,\ p = 0.7$ *(h)* $k = 5,\ p = 0.8$ *(i)* $k = 5,\ p = 0.9$

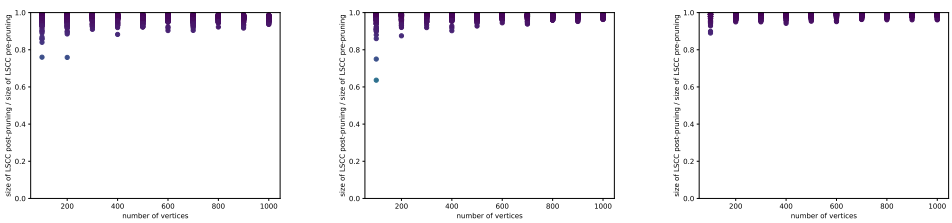


(j) $k = 5,\ p = 0.95$ *(k)* $k = 5,\ p = 0.97$ *(l)* $k = 5,\ p = 0.99$

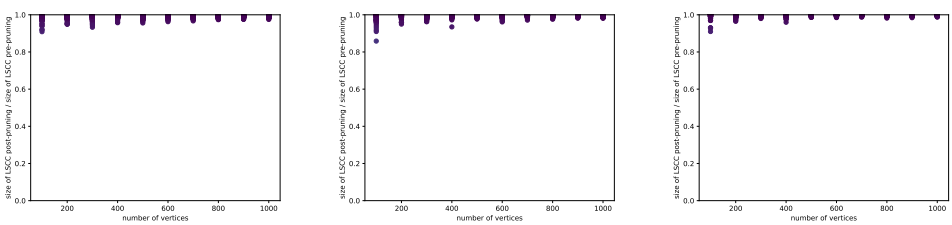
Figure C.6: Scatter plots of LSSC reduction factor ρ_L *obtained from simulations in the parameter regime* $n \in (100, 1000)$ *for* $k = 5$.



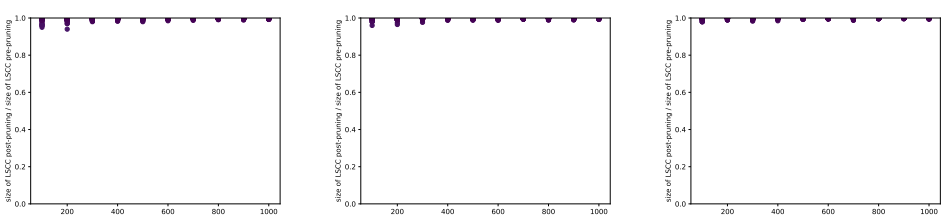
(a) $k = 6,\ p = 0.1$ *(b)* $k = 6,\ p = 0.2$ *(c)* $k = 6,\ p = 0.3$



(d) $k = 6,\ p = 0.4$ *(e)* $k = 6,\ p = 0.5$ *(f)* $k = 6,\ p = 0.6$



(g) $k = 6,\ p = 0.7$ *(h)* $k = 6,\ p = 0.8$ *(i)* $k = 6,\ p = 0.9$



(j) $k = 6,\ p = 0.95$ *(k)* $k = 6,\ p = 0.97$ *(l)* $k = 6,\ p = 0.99$

Figure C.7: Scatter plots of LSSC reduction factor ρ_L obtained from simulations in the parameter regime $n \in (100, 1000)$ for $k = 6$.

§C.3.3 Largest strongly connected component coverage factor

The LSCC coverage factor α_L, being a quantity related both to the size of the communication network and the allocation of computation jobs, depends on all the parameters $E_{\max}$ (or alternatively p), β, n and k.

The behaviour of α_L might be slightly misleading. For example, for $p = 0.1$, $k = 4$, and β low enough (see Figure C.8a), α_L is concentrated at the value 1. When interpreting these results, one must realise that this is because the LSCC itself is rather small, and therefore the assignment of all the jobs available within the LSCC, which are furthermore energetically cheap due to β being small, is almost trivial. If β is close to one, then even on a small LSCC there might not be enough energy to carry out all the computational jobs. As p becomes slightly larger, the results are quite spread out and the influence of β becomes significant. Finally, as p becomes closer to 1, β becomes the main controlling parameter.

Some of our results are plotted in Figures C.8, C.9 and C.10.

§C.3.4 Coverage factor

Unlike the LSCC coverage factor α_L, the (global) coverage factor α is easy to interpret. The coverage α increases with p and decreases with β, in line with intuitive expectations. Regimes of intermediate β and p can roughly be divided into 2 classes: regimes where the coverage α concentrates around some intermediate value and regimes dominated by inherent randomness, with values of α significantly spread out.

A somewhat counterintuitive phenomenon in the data is that for small values of p and n, the coverage is higher than naïvely expected. This is mostly due to the network being small, and corresponds to the similarly counterintuitive behaviour of α_L in this parameter regime explained in the previous section.

These results are illustrated in Figures C.11, C.12 and C.13.

§C.3.5 Additional plots

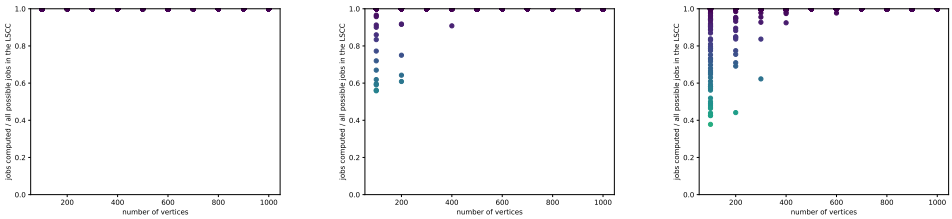
Additional plots are available via DOI:10.5281/zenodo.8433181.

The naming scheme of the files is as follows:

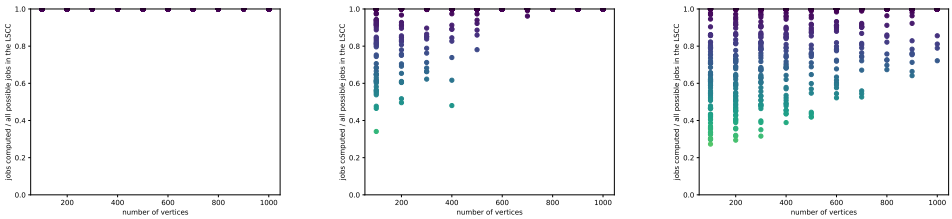
```
fig_{descriptor}_{k}_{p}_{beta}.pdf
```

where `{descriptor}` can be `fracLSCC` for plots related to LSCC reduction parameter ρ_L, `coverLSCC` for LSSC coverage α_L and `coverALL` for coverage parameter α. The placeholder `{k}` corresponds to the value of k chosen in the simulation, `p` denotes the chosen value of p and `{beta}` denotes the chosen value of β. For ρ_L, which is independent of β, the naming convention uses $\beta = 1$.

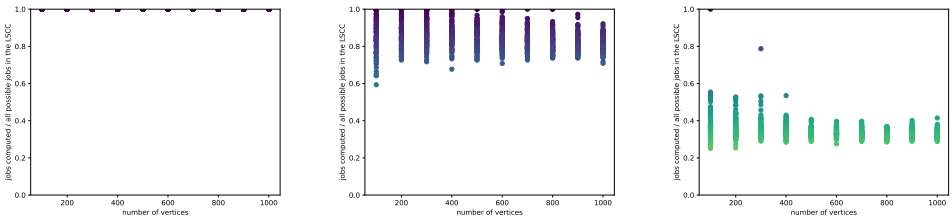
For example, `fig_coverALL_5-0.980-0.700.pdf` corresponds to the plot of the coverage parameter α for networks with $k = 5$, $p = 0.98$ and $\beta = 0.7$.



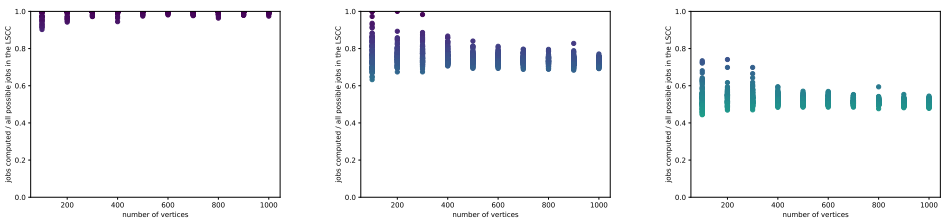
(a) $k = 4,\ p = 0.1,\ \beta = 0.1$ *(b)* $k = 4,\ p = 0.1,\ \beta = 0.7$ *(c)* $k = 4,\ p = 0.1,\ \beta = 1.0$



(d) $k = 4,\ p = 0.2,\ \beta = 0.1$ *(e)* $k = 4,\ p = 0.2,\ \beta = 0.6$ *(f)* $k = 4,\ p = 0.2,\ \beta = 1.0$

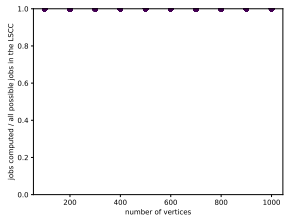


(g) $k = 4,\ p = 0.7,\ \beta = 0.1$ *(h)* $k = 4,\ p = 0.7,\ \beta = 0.4$ *(i)* $k = 4,\ p = 0.7,\ \beta = 1.0$

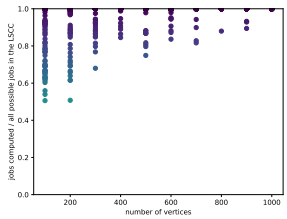


(j) $k = 4,\ p = 0.99,\ \beta = 0.5$ *(k)* $k = 4,\ p = 0.99,\ \beta = 0.7$ *(l)* $k = 4,\ p = 0.99,\ \beta = 1.0$

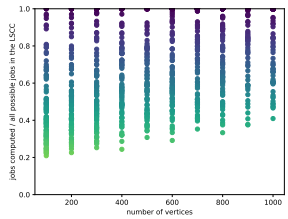
Figure C.8: Scatter plots of LSSC coverage factor α_L obtained from simulations in the parameter regime $n \in (100, 1000)$ for $k = 4$. Each row has p fixed and β is chosen to showcase different possible behaviour.



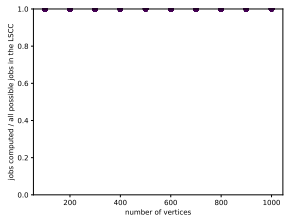
(a) $k = 5,\ p = 0.1,\ \beta = 0.1$



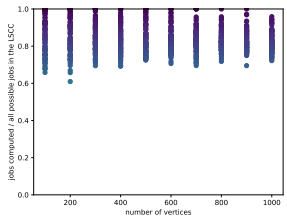
(b) $k = 5,\ p = 0.1,\ \beta = 0.4$



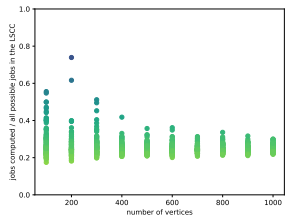
(c) $k = 5,\ p = 0.1,\ \beta = 1.0$



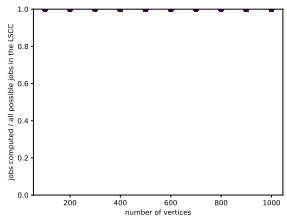
(d) $k = 5,\ p = 0.3,\ \beta = 0.1$



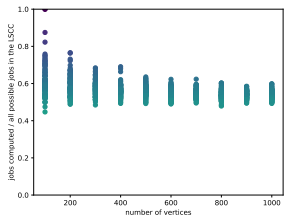
(e) $k = 5,\ p = 0.3,\ \beta = 0.3$



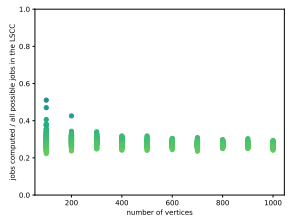
(f) $k = 5,\ p = 0.3,\ \beta = 1.0$



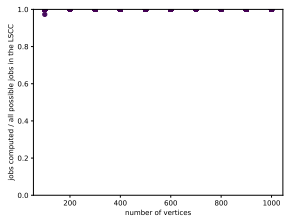
(g) $k = 5,\ p = 0.7,\ \beta = 0.1$



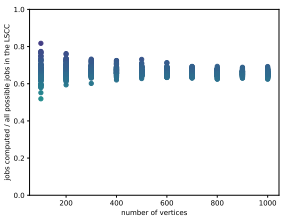
(h) $k = 5,\ p = 0.7,\ \beta = 0.5$



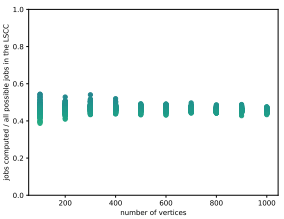
(i) $k = 5,\ p = 0.7,\ \beta = 1.0$



(j) $k = 5,\ p = 0.99,\ \beta = 0.4$

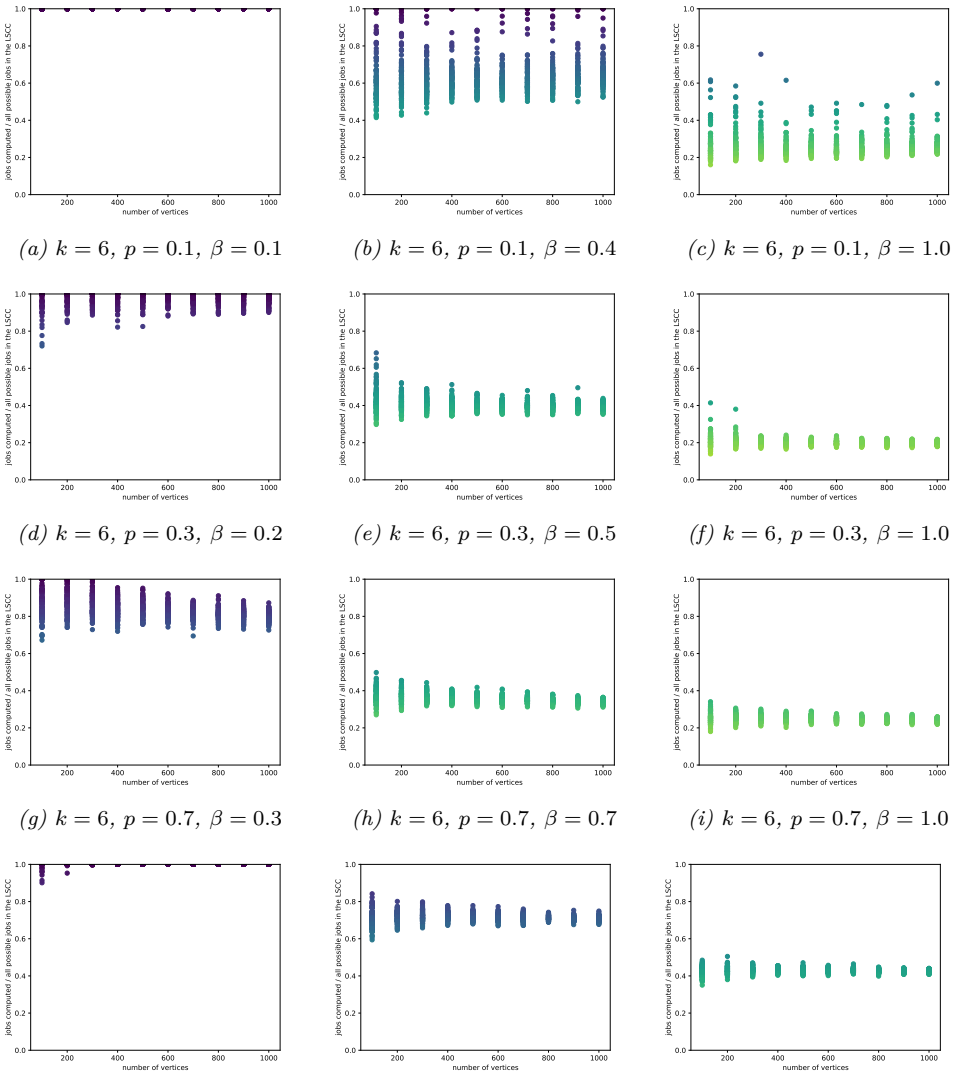


(k) $k = 5,\ p = 0.99,\ \beta = 0.7$



(l) $k = 5,\ p = 0.99,\ \beta = 1.0$

Figure C.9: Scatter plots of LSSC coverage factor α_L obtained from simulations in the parameter regime $n \in (100, 1000)$ for $k = 5$. Each row has p fixed and β is chosen to showcase different possible behaviour.



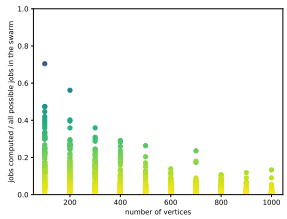
(a) $k = 6,\ p = 0.1,\ \beta = 0.1$ *(b)* $k = 6,\ p = 0.1,\ \beta = 0.4$ *(c)* $k = 6,\ p = 0.1,\ \beta = 1.0$

(d) $k = 6,\ p = 0.3,\ \beta = 0.2$ *(e)* $k = 6,\ p = 0.3,\ \beta = 0.5$ *(f)* $k = 6,\ p = 0.3,\ \beta = 1.0$

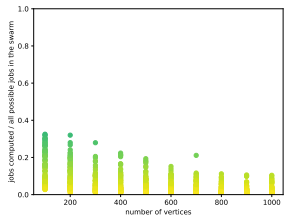
(g) $k = 6,\ p = 0.7,\ \beta = 0.3$ *(h)* $k = 6,\ p = 0.7,\ \beta = 0.7$ *(i)* $k = 6,\ p = 0.7,\ \beta = 1.0$

(j) $k = 6,\ p = 0.99,\ \beta = 0.4$ *(k)* $k = 6,\ p = 0.99,\ \beta = 0.6$ *(l)* $k = 6,\ p = 0.99,\ \beta = 1.0$

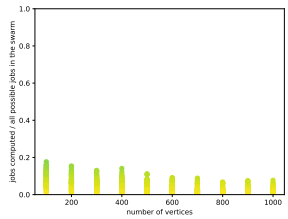
Figure C.10: Scatter plots of LSSC coverage factor α_L obtained from simulations in the parameter regime $n \in (100, 1000)$ for $k = 6$. Each row has p fixed and β is chosen to showcase different possible behaviour.



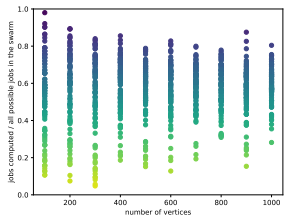
(a) $k = 4,\ p = 0.2,\ \beta = 0.1$



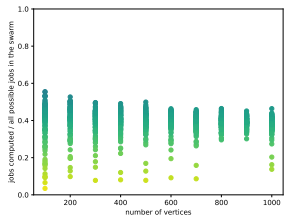
(b) $k = 4,\ p = 0.2,\ \beta = 0.5$



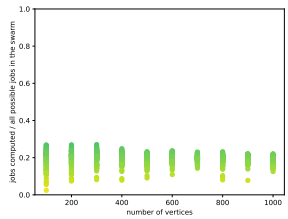
(c) $k = 4,\ p = 0.2,\ \beta = 1.0$



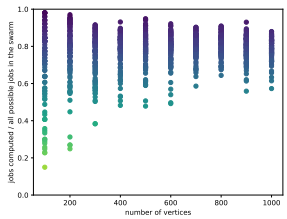
(d) $k = 4,\ p = 0.5,\ \beta = 0.1$



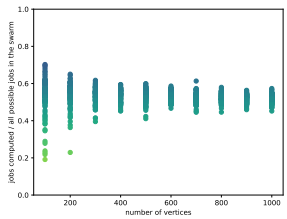
(e) $k = 4,\ p = 0.5,\ \beta = 0.5$



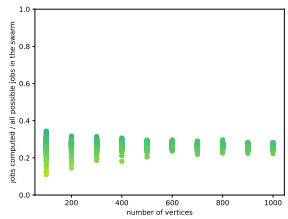
(f) $k = 4,\ p = 0.5,\ \beta = 1.0$



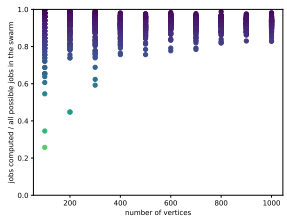
(g) $k = 4,\ p = 0.7,\ \beta = 0.1$



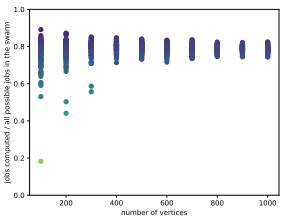
(h) $k = 4,\ p = 0.7,\ \beta = 0.5$



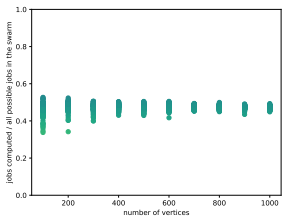
(i) $k = 4,\ p = 0.7,\ \beta = 1.0$



(j) $k = 4,\ p = 0.99,\ \beta = 0.1$

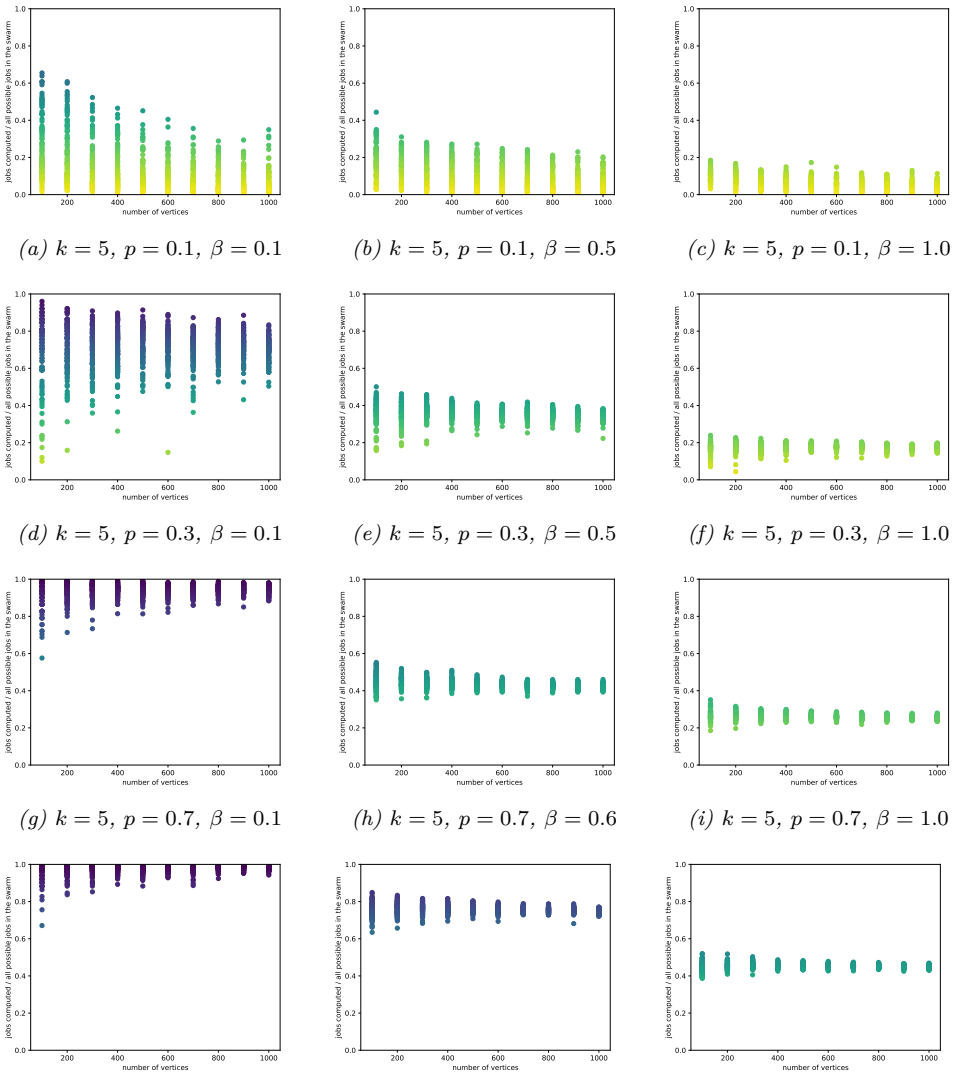


(k) $k = 4,\ p = 0.99,\ \beta = 0.1$



(l) $k = 4,\ p = 0.99,\ \beta = 1.0$

Figure C.11: Scatter plots of coverage factor α obtained from simulations in the parameter regime $n \in (100, 1000)$ for $k = 4$. Each row has p fixed and β is chosen to showcase different possible behaviour.



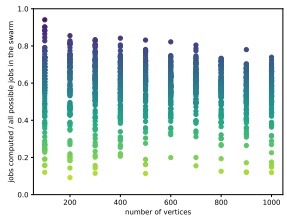
(a) $k = 5,\ p = 0.1,\ \beta = 0.1$ *(b)* $k = 5,\ p = 0.1,\ \beta = 0.5$ *(c)* $k = 5,\ p = 0.1,\ \beta = 1.0$

(d) $k = 5,\ p = 0.3,\ \beta = 0.1$ *(e)* $k = 5,\ p = 0.3,\ \beta = 0.5$ *(f)* $k = 5,\ p = 0.3,\ \beta = 1.0$

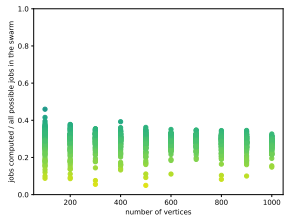
(g) $k = 5,\ p = 0.7,\ \beta = 0.1$ *(h)* $k = 5,\ p = 0.7,\ \beta = 0.6$ *(i)* $k = 5,\ p = 0.7,\ \beta = 1.0$

(j) $k = 5,\ p = 0.99,\ \beta = 0.1$ *(k)* $k = 5,\ p = 0.99,\ \beta = 0.6$ *(l)* $k = 5,\ p = 0.99,\ \beta = 1.0$

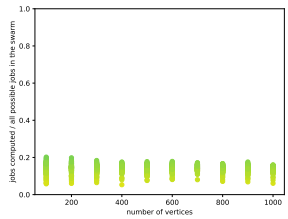
Figure C.12: Scatter plots of coverage factor α obtained from simulations in the parameter regime $n \in (100, 1000)$ for $k = 5$. Each row has p fixed and β is chosen to showcase different possible behaviour.



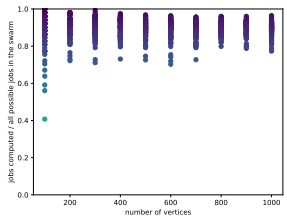
(a) $k = 6,\ p = 0.1,\ \beta = 0.1$



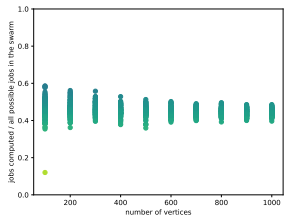
(b) $k = 6,\ p = 0.1,\ \beta = 0.5$



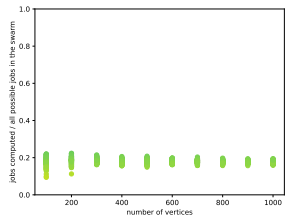
(c) $k = 6,\ p = 0.1,\ \beta = 0.1$



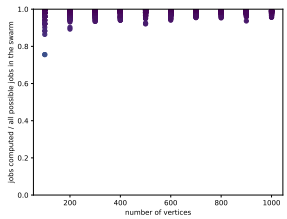
(d) $k = 6,\ p = 0.3,\ \beta = 0.1$



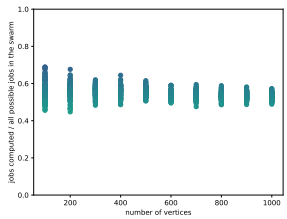
(e) $k = 6,\ p = 0.3,\ \beta = 0.4$



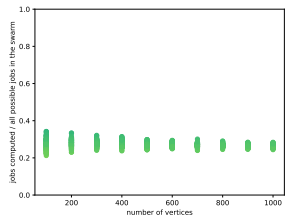
(f) $k = 6,\ p = 0.3,\ \beta = 1.0$



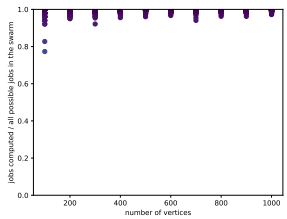
(g) $k = 6,\ p = 0.8,\ \beta = 0.1$



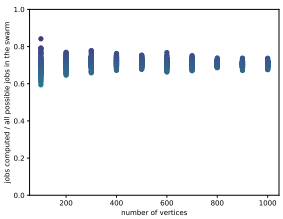
(h) $k = 6,\ p = 0.8,\ \beta = 0.5$



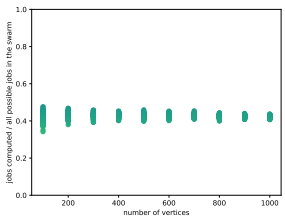
(i) $k = 6,\ p = 0.8,\ \beta = 1.0$



(j) $k = 6,\ p = 0.99,\ \beta = 0.1$



(k) $k = 6,\ p = 0.99,\ \beta = 0.6$



(l) $k = 6,\ p = 0.99,\ \beta = 1.0$

Figure C.13: Scatter plots of coverage factor α obtained from simulations in the parameter regime $n \in (100, 1000)$ for $k = 6$. Each row has p fixed and β is chosen to showcase different possible behaviour.

CHAPTER 5

Bringing order to network centrality measures

This chapter is based on the following article:

G. Exarchakos, R. van der Hofstad, O. Nagy, and M. Pandey. Bringing order to network centrality measures. Work in progress, 2024+.

Abstract

We introduce a quantitative method to compare arbitrary pairs of graph centrality measures, based on the ordering of vertices induced by them. The proposed method is conceptually simple, mathematically elegant, and allows for a quantitative restatement of many conjectures that were previously cumbersome to formalise. Moreover, it produces an approximation scheme useful for network scientists. We explore some of these use cases and formulate new conjectures that are of independent interest.

§5.1 Introduction

The importance of centrality measures in network science is similar to the importance of derivatives or integrals in calculus. While both concepts are *intuitively* clear to practitioners, they both defy a single *universally* applicable mathematical definition. The case of centrality measures is arguably less straightforward, since there are many reasonable centrality measures that capture different aspects of what "being central" can mean in a network. To illustrate this point, the database [85] currently lists more than 400 centrality measures, each motivated by a different use case. Unfortunately, there are usually no tractable relations between them. We refer to [114, Chapter 7] for a comprehensive overview of centrality measures.

While there are some more light-hearted uses of centrality measures (e.g., the Kevin Bacon game [60, 140] is related to closeness centrality, similarly the Erdős number [70] measure centrality in the collaboration network in mathematics), they have also been a subject of intense research across disciplines. The PageRank centrality measure, introduced in [122, 40], is said to follow the PageRank power-law hypothesis, which states that the in-degree and the PageRank follow the same power-law distribution in major real-world networks. Surprisingly, while often true, this hypothesis was recently disproved for directed Preferential Attachment model [17].

There is not much research on the *relations* between various centrality measures. In [54], an analytic derivation is given that relates closeness and degree centrality for undirected configuration models. Though in centrality measures, we are majorly concerned with the induced rankings rather than the exact formula, it suggests that the inverse of closeness centrality of a vertex is linearly dependent on the logarithm of its degree. The dependence of PageRank on the damping factor is studied in [33]. Comparisons of centrality measures appear in [108, 145, 65, 64, 39], often aiming to investigate which centrality measures outperform the others in a specific task. For example, [108] studies the effect of removal of top-ranked vertices according to various centrality measures on network connectivity, thus comparing centrality measures on their ability to identify connectivity hubs in real-world networks. Likewise, [39] compares the performance of various centrality measures in the specific task of keyword extraction from scientific articles. General comparisons of arbitrary centrality measures without any reference to a specific task are virtually non-existent, barring studies of correlations between centrality measures [65]. On a related note, there have been attempts to cluster centrality measures according to some similarities in their behaviour [134], but this clustering is mostly heuristic and does not reflect any quantifiable clustering criteria.

This chapter aims to bring order to centrality measures by proposing a *quantitative method* to compare arbitrary pairs of centrality measures defined on the same graph. Our method is based on the *vertex-ordering* induced by the centrality measures. Intuitively, it captures the disagreement between the collections of top-ranked vertices for two different centrality measures. This comparison scheme readily admits both practical and theoretical uses. Apart from the obvious approximation scheme, this method also allows us to put some previously stated informal and qualitative conjectures on solid footing, and to formulate a set of new conjectures, hinting at a more profound

relationship between various centrality measures. We achieve this by the centrality comparison curve, that we explain next.

§5.2 The centrality comparison curve

Let $G = (V_G, E_G)$ be a graph with vertex set V_G and edge set E_G. A *centrality measure* is a map $R\colon V_G \to \mathbb{R}$ such that $R(v) > R(u)$ should be interpreted as meaning that v is more central than u. For simplicity, from now on we will assume that $V_G = \{1, \dots, n\} = [n]$.

We can see R as a naive way to order the vertices in V_G. For our purposes, we need to refine this ordering so that it becomes *total* and it breaks ties consistently. To this end, we introduce the *induced ordering*.

Definition 5.2.1 (Induced vertex ordering). Fix a graph G, and let R, S be two centrality measures on G. Endow each vertex $v \in V_G$ with a number u_v sampled from the standard $\mathrm{Unif}(0,1)$ distribution. For any $a, b \in V_G$, define the *total* ordering $\prec_{(R,S)}$ via the following procedure:

(a) If $R(a) < R(b)$ then $a \prec_{(R,S)} b$.

(b) If $R(a) = R(b)$ and $S(a) < S(b)$, then $a \prec_{(R,S)} b$.

(c) If $R(a) = R(b)$ and $S(a) = S(b)$ and $u_a < u_b$, then $a \prec_{(R,S)} b$.

■

Note that $\prec_{(R,S)}$ is different from $\prec_{(S,R)}$. The tie-breaker rule in Definition 5.2.1 can be slightly modified, see below for the consequences of this.

The above total ordering allows us to define the main contribution of this chapter, the Centrality Comparison Curve:

Definition 5.2.2 (Centrality comparison curve). Fix a graph G with n vertices, take any two centrality measures R, S defined over G and consider the induced orderings $\prec_{(R,S)}$, $\prec_{(S,R)}$. For any total ordering $\prec$ of V_G, denote by $\mathsf{Top}_{\prec}(V_G, l)$ the first l elements in the ordered set $(V_G, \prec)$.

The centrality comparison curve (CCC) on a graph of size $|V_G| = n$ is a mapping $\mathsf{CCC}^G_{R,S}\colon (0,1] \to [n]/n$ defined by

$$\mathsf{CCC}^G_{R,S}(x) = \frac{\left|\mathsf{Top}_{\prec_{(R,S)}}([n], \lfloor xn \rfloor) \cap \mathsf{Top}_{\prec_{(S,R)}}([n], \lfloor xn \rfloor)\right|}{n} \tag{5.1}$$

In words, $\mathsf{CCC}^G_{R,S}(k/n)$ represents the overlap of the k most central vertices according to two centrality measures, normalised by the size of the graph. ■

We next remark on the elegant properties that CCC has.

Lemma 5.2.3 (Basic properties of CCC).

(a) *For any graph G and any two centrality measures R, S defined on G, and for any $x \in [0,1]$,*

$$\mathsf{CCC}^G_{R,S}(x) \leq x \tag{5.2}$$

(b) *Let $\phi : \mathbb{R} \to \mathbb{R}$ be a strictly monotone function. Then, for any graph G and any centrality measures R, S defined on G, and for any $x \in [0,1]$,*

$$\mathsf{CCC}^G_{R,S}(x) = \mathsf{CCC}^G_{\phi\circ R,\phi\circ S}(x), \tag{5.3}$$

where $\phi \circ R : V_G \to \mathbb{R}$ is the composition of maps ϕ and R, likewise for $\phi \circ S$.

Proof.

(a) Since

$$|\mathsf{Top}_{\prec_{(R,S)}}([n], \lfloor xn \rfloor)| \leq xn, \tag{5.4}$$

then also

$$|\mathsf{Top}_{\prec_{(R,S)}}([n], \lfloor xn \rfloor) \cap \mathsf{Top}_{\prec_{(S,R)}}([n], \lfloor xn \rfloor)| \leq xn, \tag{5.5}$$

which, combined with (5.1), yields the claim.

(b) Since strictly monotone functions are order-preserving, $\phi \circ R$ and $\phi \circ S$ produce the same induced vertex ordering (recall Definition 5.2.1) as R and S. The claim follows immediately.

□

The following lemma allows us to capture what it means for two centrality measures to be independent of each other.

Lemma 5.2.4 (Comparison with ordering chosen u.a.r.). *Let $(\prec^A_n, \prec^U_n)_{n=1}^{\infty}$ be a sequence such that, for every $n \in \mathbb{N}$:*

(a) *$\prec^A_n$ is a fixed, but arbitrary, total ordering of $[n]$.*

(b) *$\prec^U_n$ is a total ordering of $[n]$ chosen u.a.r., independently of anything else.*

Then for any $n \in \mathbb{N}$ and any $k \leq n, k \in N$:

$$\mathbb{E}\left[|\mathsf{Top}_{\prec^A_n}([n], k) \cap \mathsf{Top}_{\prec^U_n}([n], k)|\right] = \frac{k^2}{n}. \tag{5.6}$$

Proof. Note that $\mathsf{Top}_{\prec^A_n}([n], k)$ contains k elements by definition. For every $i \in \mathsf{Top}_{\prec^A_n}([n], k)$, it is true that

$$\mathbb{P}\left(i \in \mathsf{Top}_{\prec^U_n}([n], k)\right) = \frac{k}{n}, \tag{5.7}$$

since $\prec_n^U$ is an ordering chosen independently uniformly at random from all orderings of $[n]$. Therefore, it follows that

$$\mathbb{E}\left[\left|\mathsf{Top}_{\prec_n^A}([n],k) \cap \mathsf{Top}_{\prec_n^U}([n],k)\right|\right] = \sum_{i=1}^{k} \frac{k}{n} = \frac{k^2}{n}, \tag{5.8}$$

which is the desired claim. □

Remark 5.2.5. In the setting of Lemma 5.2.4, by taking $k = xn$, $x \in [0,1]$, it follows that

$$\mathbb{E}\left[\mathsf{CCC}^{G_n}_{\prec_n^A,\prec_n^U}(x)\right] \to x^2 \qquad \text{as } n \to \infty, \tag{5.9}$$

where $\prec_n^A$, $\prec_n^U$ are now seen as centrality measures implied by the ordering of the vertex label-set $[n]$ of the graph G_n. We believe that it is possible to strengthen this statement to process convergence of $(\mathsf{CCC}^{G_n}_{\prec_n^A,\prec_n^U}(x))_{x\in[0,1]}$ to $(x^2)_{x\in[0,1]}$, but we will not investigate these convergence properties further in this chapter.

♦

Motivated by Lemma 5.2.4, we introduce the following definition.

Definition 5.2.6 (ε-independent centrality measures). Fix $\varepsilon \in [0,1]$, let G be a graph and let R, S be two centrality measures defined on G. The pair of centrality measures R, S is ε-independent, if for every $x \in [0,1]$:

$$\mathsf{CCC}^G_{R,S}(x) \in (x^2 - \varepsilon, x^2 + \varepsilon). \tag{5.10}$$

■

Lemma 5.2.3 and Remark 5.2.5 give rise to a natural comparison: If $\mathsf{CCC}^G_{R,S}$ is close to the identity graph, R, S induce very similar rankings, while these rankings are fully independent if they are close to x^2. Finally, $\mathsf{CCC}^G_{R,S}$ makes no assumption about the edge density of the graph; hence it works for both sparse and dense graphs, as well as directed and undirected graphs.

In some applications, it is only relevant to what extent the two centrality measures agree on what are is the set of, say, top 10% of vertices. For these purposes, we introduce a simple *scalar* quantity derived off CCC.

$$\mathsf{CCCo}^G_{R,S}(p) = \frac{\mathsf{CCC}^G_{R,S}(p)}{p}, \tag{5.11}$$

which, for p small, can be seen as an approximation of the right-derivative at $x = 0$. Informally, one can expect roughly $n\mathsf{CCCo}^G_{R,S}(k/n)$ of the k most highly ranked vertices for R to also be in the k most highly ranked vertices for S. For example, $\mathsf{CCCo}^G_{R,S}(0.05)$ measures how well the centrality measures R, S agree w.r.t. the top 5% of most central vertices. The closer it is to 1, the better they agree.

§5.3 Putting CCC to practice

We use CCC on artificial and real-world directed networks to see which centrality measures are alike. For the artificial network, we use a directed configuration model [35, 110] of size $n = 1,000$, with in- and out-degree power-law exponent 3. We run this computation several times with different samples of the configuration model, and the red error bars indicate the standard deviation around the mean of the 50 runs. The real-world network is the directed "hep-ph" citation network [100]. Further details about these computations can be found in Appendix D.

We will explain the plots of CCCs on the example of PageRank and in-/out-degree.

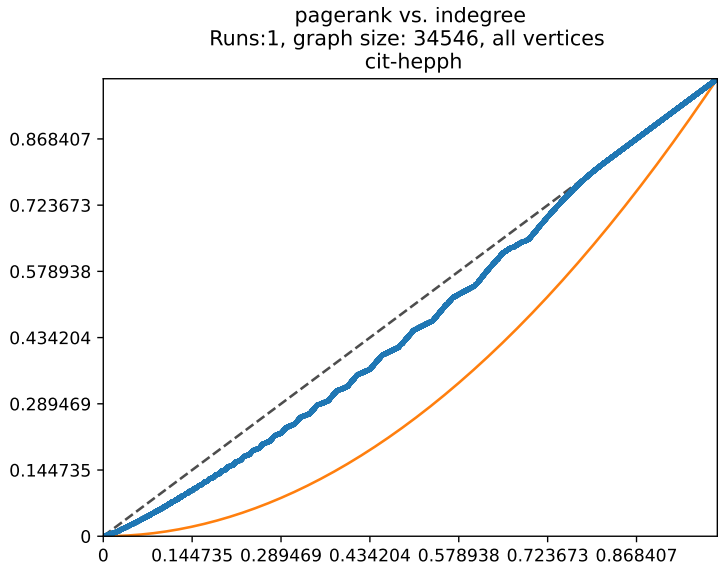


Figure 5.1: CCC for PageRank and in-degree for the citation network [99]

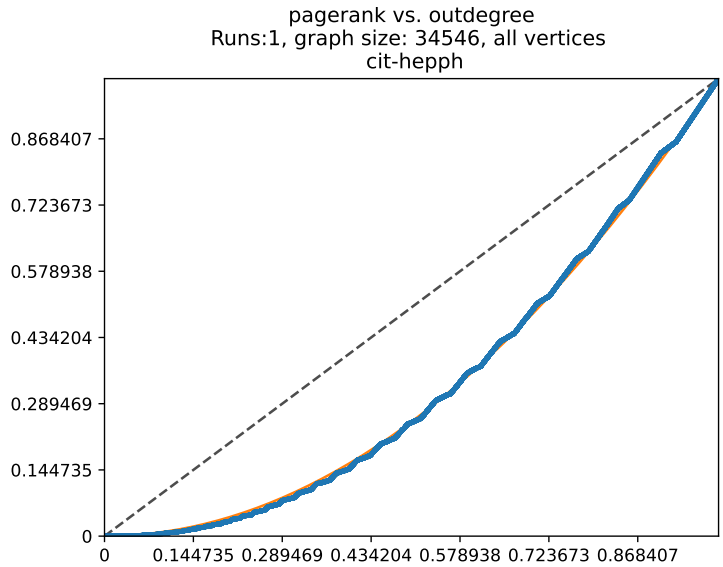


Figure 5.2: CCC for out-degree and PageRank for the citation network [99]

Illustrative example. Figures 5.1–5.2 show that in the real-world citation network, PageRank is close to in-degree, and essentially independent of out-degree. The first conclusion follows from the fact that the blue CCC is close to the dashed identity curve in Figure 5.1, while the other follows since the CCC in Figure 5.2 follows almost perfectly the yellow x^2 curve.

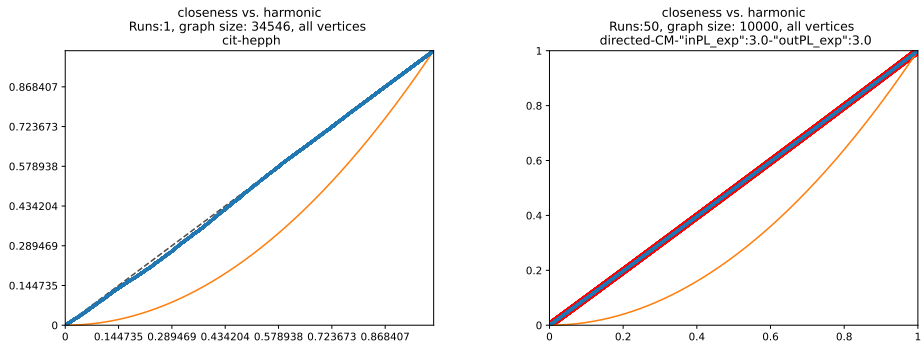


Figure 5.3: CCC for closeness and harmonic centrality, for the citation network and artificial network.

Closeness and harmonic centrality are similar. As can be expected, due to the close relation in their definition, closeness and harmonic centrality are often close (see Figure 5.3). A similar relation holds between betweenness and load centrality (see Figure 5.4).

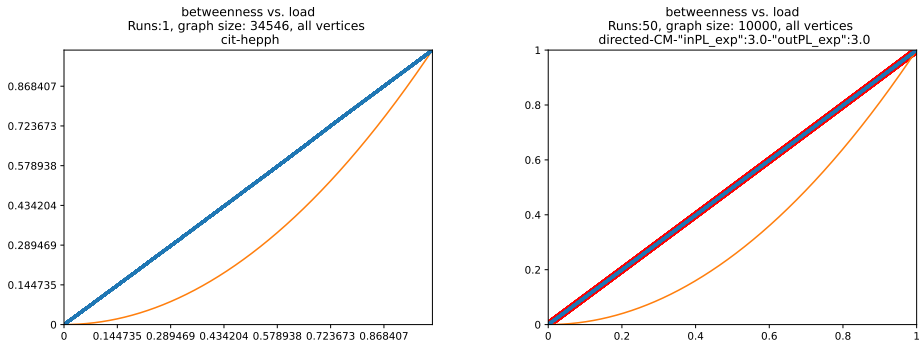


Figure 5.4: CCC for betweenness and load centrality, for the citation network and artificial network.

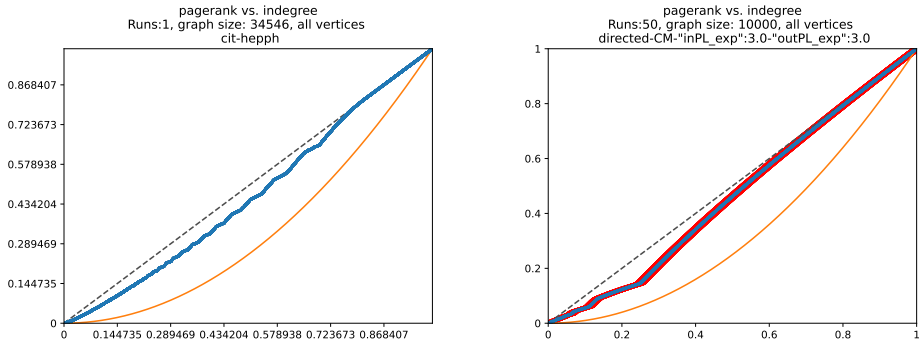


Figure 5.5: CCC for PageRank and degree, for the citation network and artificial network.

PageRank follows in-degree. While in-degree centrality is simply the in-degree of any given vertex, PageRank is a more involved stochastic process originally used for ranking pages in Google search results [122]. There is a large body of work showing that the PageRank distribution has the same power-law exponent as the in-degree distribution, see [63] and references therein. As a result, unsurprisingly, the CCC curve for these centralities is close to a straight line (see Figure 5.5).

PageRank hardly depends on damping factor. PageRank centrality has a parameter, called the *damping factor*, that is often taken as 0.85 [122]. However, the precise choice of the damping factor does not significantly matter, as can be seen in Figure 5.6.

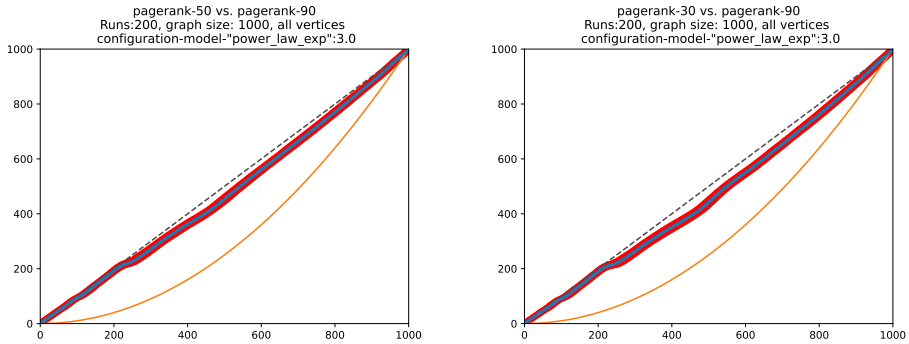


Figure 5.6: CCC for PageRank with damping factors 0.5 and 0.9, and 0.3 and 0.9, respectively, for artificial network.

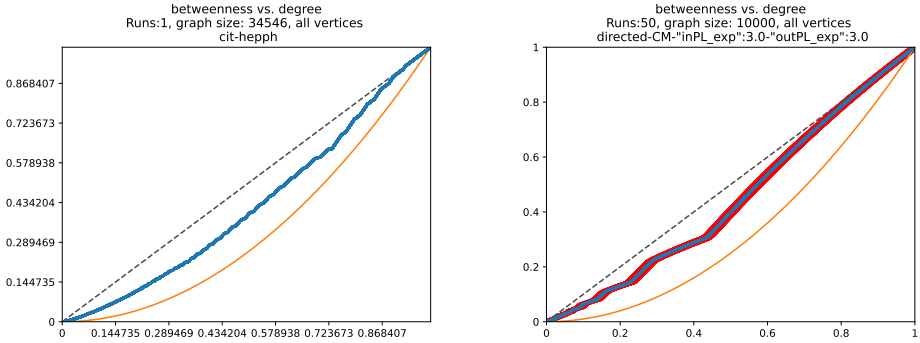


Figure 5.7: CCC for betweenness and degree, for the citation network and artificial network.

Betweenness and degree agree on top vertices. Betweenness and degree are obviously quite different. However, the top-ranked vertices have a high overlap for the configuration model, but not for the collaboration graph (see Figure 5.7).

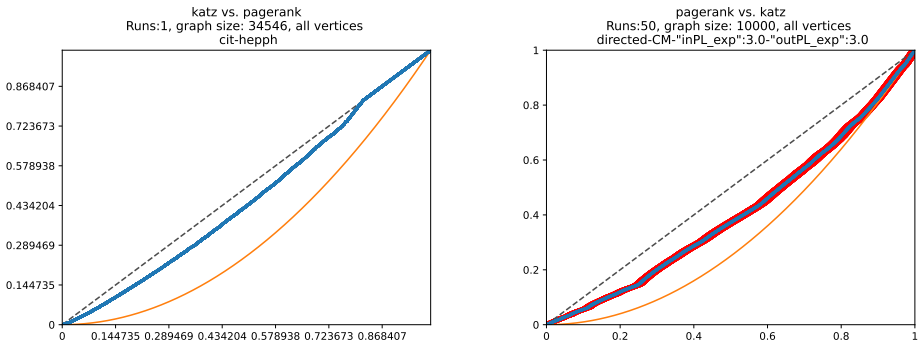


Figure 5.8: CCC for PageRank and Katz, for the citation network and artificial network.

PageRank is very different from Katz. Surprisingly, even though the definitions of PageRank centrality and Katz centrality are highly similar, their centralities behave quite differently (see Figure 5.8).

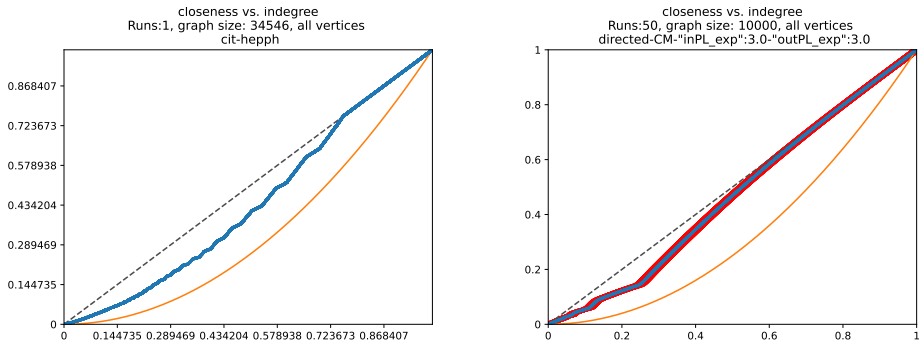


Figure 5.9: CCC for closeness and in-degree, for the citation network and artificial network.

Closeness and in-degree are weakly related. In the directed configuration model, closeness and in-degree are somewhat related (see Figure 5.9).

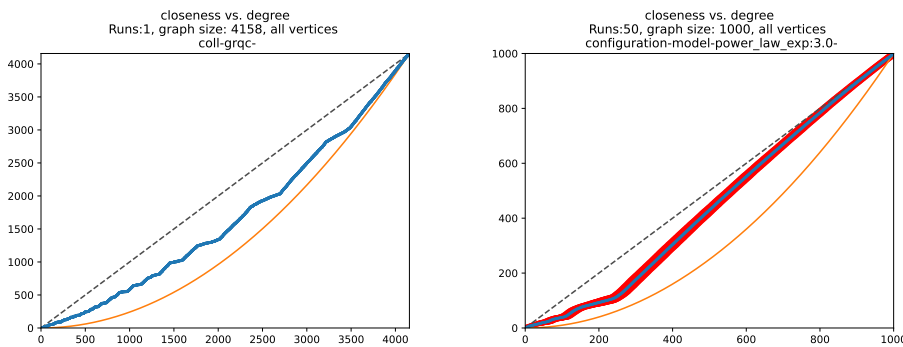


Figure 5.10: CCC for closeness and degree, for undirected collaboration network and artificial network.

Closeness and degree are different in undirected setting. In the configuration model, surprisingly, closeness and degree are quite different (see Figure 5.10), unlike what was reported in [54]. We refer to Appendix D for scatter plot of closeness and degree, which does show that there is some weak relation, which is, however, not obviously reflected in the CCC.

Local centrality measures. For PageRank, it is known that it is a *local* quantity, in that it can be well-approximated by its finite power iteration [63]. Obviously also in-degree, out-degree and degree centrality are local, and due to its definition, one can also expect Katz centrality to be a local measure. For local centrality measures, one can show that the CCC converges to a limiting continuous curve [56], that can be determined by the local limit of the graph. Local convergence, and locally tree-like behaviour,

have become central techniques in network science. However, it is *not* true that local measures are all alike (see, for example, Figure 5.8), even though their *relations* can be analysed through the local limit.

§5.4 Conclusions & conjectures

Centrality measures have been so far compared by computing the *correlation* between centralities. However, correlation captures only a *linear* dependence, and it can vary wildly when applying a monotone transformation. The advantage of the CCC is that it is *invariant* under monotone transformations, and thus more robust.

As such, the CCC gives a quantitative way of comparing how close centrality measures are, as well as the most central vertices in the network, or, alternatively, how independent they are. We see that the CCC confirms that load centrality is close to betweenness centrality, as well as closeness to harmonic centrality, as can be expected from their definitions. Further, PageRank is close to in-degree centrality, and is also hardly affected by the choice of damping factor. Remarkably, even though their definitions are close, PageRank and Katz centrality behave rather differently.

Another obvious use of CCC is to identify replacements for hard-to-compute centrality measures in cases when only the ranking of nodes is needed. For example, ranking by closeness and harmonic centrality, or by betweenness and load centrality, is identical in both artificial and real-world networks. Similarly, ranking by PageRank can be well-approximated by the ranking obtained from in-degrees.

CCC is an interesting mathematical object. The plots presented above make it plausible that the CCC, seen as a stochastic process, converges to a limiting process, which in the setting explored herein might be a deterministic curve. Ideas along these lines are formalised in [56], which provides an argument that CCC converges for *local* centrality measures for a certain class of random graphs.

APPENDIX D

Appendices of Chapter 5

§D.1 Definitions of centrality measures

As mentioned in the main text, centrality measures assign a number to every vertex in a given graph; this number represents a particular notion of what centrality can mean in different contexts. Formally, *centrality measure* R is a map $R\colon V_G \to \mathbb{R}$ such that $R(v) > R(u)$ should be interpreted as meaning that v is more central than u. There are many centrality measures that have been considered in the literature. In this paper, we consider in-degree, out-degree, total degree, PageRank, betweenness centrality, closeness centrality, load centrality, harmonic centrality, Katz centrality. Below, we present definitions of these centrality measures.

PageRank. Let $R : V_G \to \mathbb{R}_{\geq}0$ be the PageRank function, then it satisfies the following recursion

$$R(i) = c \sum_{j \in V_G} \frac{e_{j,i}}{d_j^+} R(j) + (1 - c), \tag{D.1}$$

where $e_{j,i}$ is the number of edges from j to i in G, d_j^+ is the out-degree of j, and $c \in (0, 1)$ is the damping factor.

Katz centrality. Katz centrality, introduced by Leo Katz [91], is a centrality measure defined as follows. Let $K_\alpha : V_G \to \mathbb{R}_{\geq}0$ be the Katz centrality function, then

$$K_\alpha(i) = \sum_{k \geq 1} \sum_{j \in V_G} \alpha^k A_{i,j}^k, \tag{D.2}$$

where A is the adjacency matrix of G and $\alpha \in (0, 1)$ is the attenuation factor. Notice that in order for the sum in (D.2) to converge, the value of the attenuation factor α has to be chosen such that it is smaller than the reciprocal of the largest eigenvalue of A (which is non-negative, due to the non-negativity of A).

Eigenvector centrality. Let $x\colon V_G \to \mathbb{R}_{\geq}0$ be the eigenvector centrality vector, then it satisfies

$$x^t A = \lambda x^t, \tag{D.3}$$

where A is the adjacency matrix of G and λ is its associated largest modulus eigenvalue. This eigenvector is unique up to a multiplicative factor. To define a unique score, one must normalise the eigenvector.

Betweenness centrality. Let $b\colon V_G \to \mathbb{R}_{\geq}0$ be the betweenness centrality function, then

$$b(v) = \sum_{i \in V_G} \sum_{j \in V_G} \frac{\sigma_{i,j}(v)}{\sigma_{i,j}}, \tag{D.4}$$

where $\sigma_{i,j}(v)$ counts the number of shortest paths from i to j containing v and $\sigma_{i,j}$ is the number of shortest paths from i to j.

Load centrality. Let $l\colon V_G \to \mathbb{R}_{\geq}0$, the load centrality function, then

$$l(v) = \frac{\sum_{i \in V_G} \sum_{j \in V_G} \sigma_{i,j}(v)}{\sum_{i \in V_G} \sum_{j \in V_G} \sigma_{i,j}}, \tag{D.5}$$

where $\sigma_{i,j}(v)$ counts the number of shortest paths from i to j containing v and $\sigma_{i,j}$ is the number of shortest paths from i to j.

Closeness centrality. Let $c\colon V_G \to \mathbb{R}_{\geq}0$ be the closeness centrality function, then

$$c(v) = \frac{|V_G| - 1}{\sum_{u \in V_G} d_G(u, v)}, \tag{D.6}$$

where $d_G(u, v)$ is the length of the shortest path between u and v in G.

Harmonic centrality. Let $h\colon V_G \to \mathbb{R}_{\geq 0}$ be the harmonic centrality function, then

$$h(v) = \sum_{u \neq v} \frac{1}{d_G(u, v)}, \tag{D.7}$$

where $d_G(u, v)$ is the length of the shortest path between u and v in G.

§D.2 Proof of properties of CCC and tie-breaking rule

Here we comment on some properties of the CCC. First, the CCC is symmetric w.r.t. the used centrality measures, since the definition (5.1) is symmetric.

We continue by discussing the tie-breaking rule. There are various possibilities to break ties. We choose to break ties in one centrality measure as a function of the other. That will make the CCC as close to the identity as possible. This is reflected in the CCC for (in-)degree centralities (since there are many vertices with the same (in-)degree, particularly for low (in-)degrees, so that the CCC is identical to a straight line in the top-right corner. The centrality values of the most central vertices are generally *unique*, so that the tie-breaking rule has no effect on the bottom-left part of the CCC. Alternatively, we could break ties using a uniform variable associated to each of the vertices, and then the top-right part of the CCCs will be much further from a straight line (see Figure D.1).

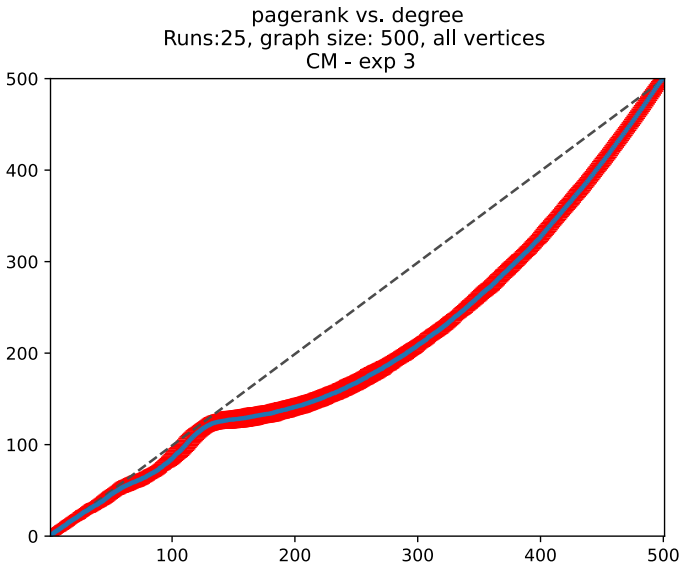


Figure D.1: PageRank vs. degree for an undirected configuration model with an alternative tie-breaking rule. Compare with the plot on the right of Figure D.6.

Generally, we prefer the tie-breaking rule discussed in the main article, since this makes the CCC easier to interpret.

§D.3 Details on the artificial networks used

The source code used to create plots in the main article is available on GitHub, via `https://github.com/nagyol/CCC-simulation`. The program makes extensive use of the NetworkX library [73].

In the case of (un-)directed configuration model with power-law degree sequence with exponent α, the individual degrees are chosen to be

$$\deg(v) = \lfloor (d_{\min} + Y) \rfloor,$$

where $d_{\min}$, the lower-bound on degrees, is chosen to be 1, and Y is a $\mathsf{Pareto}(\alpha - 1)$-distributed random variable, i.e., random variable with CDF

$$F_Y(x) = \begin{cases} 1 - \frac{1}{x^\alpha} & x \geq 1 \\ 0 & \text{otherwise.} \end{cases}$$

In the case of the directed configuration model, first the in-degree sequence is sampled via the procedure outlined above, and then the out-degree sequence is resampled until it has the same sum as the in-degree sequence.

§D.4 CCC for undirected graphs

In this section, we present the same figures as in the main text of the paper, but now for real-world and artificial undirected networks. Also in such networks, centrality measures are highly used and useful to identify important vertices in the network. The real-world network consists of the collaboration network studied in [99], while the artificial networks correspond to configuration models with power-law exponent 3.

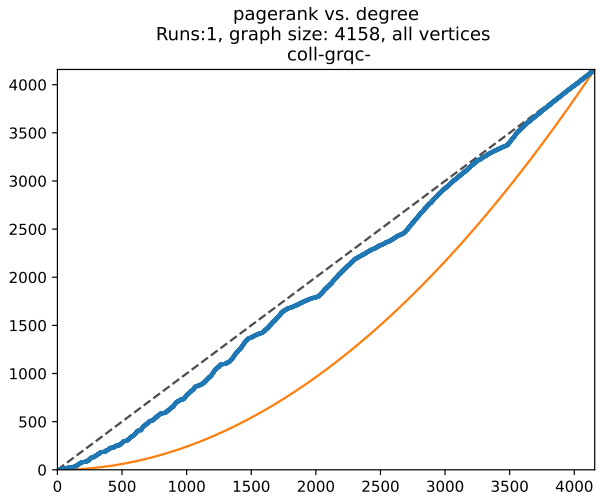


Figure D.2: CCC for PageRank and degree for collaboration network [99]

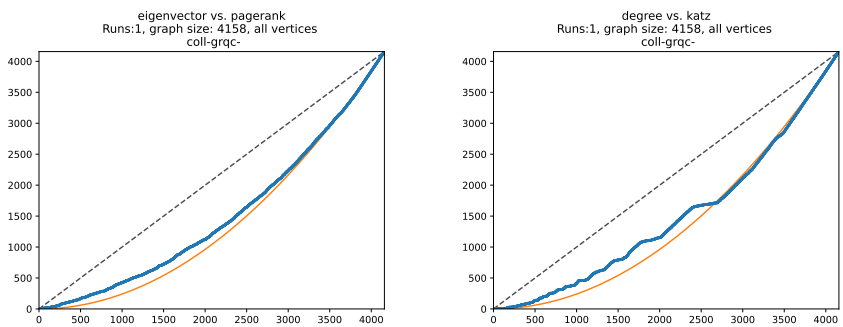


Figure D.3: CCC for eigenvector and PageRank, and degree and Katz centrality, respectively, for collaboration network [99]

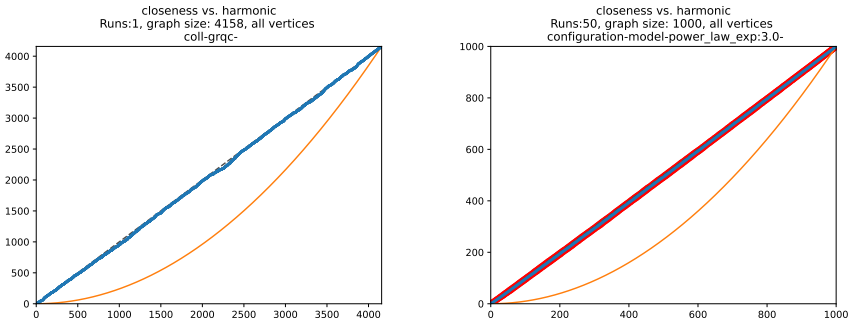


Figure D.4: CCC for closeness and harmonic centrality, for collaboration network and artificial network.

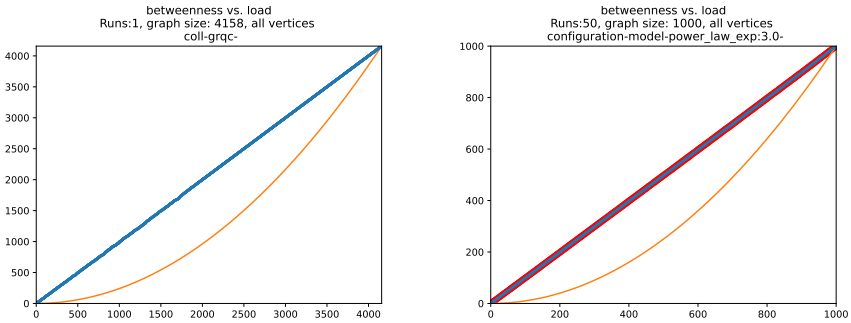


Figure D.5: CCC for betweenness and load centrality, for collaboration network and artificial network.

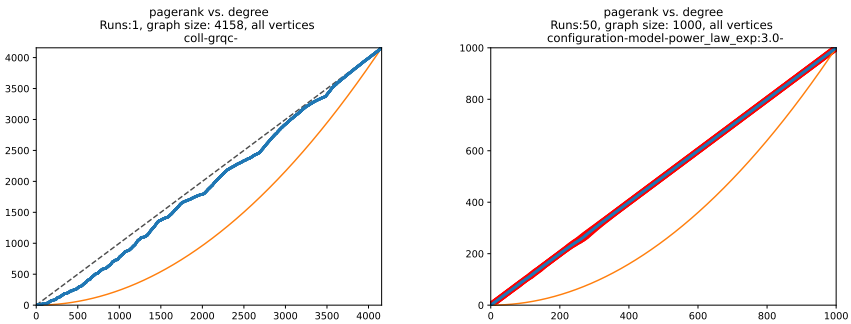


Figure D.6: CCC for PageRank and degree, for collaboration network and artificial network.

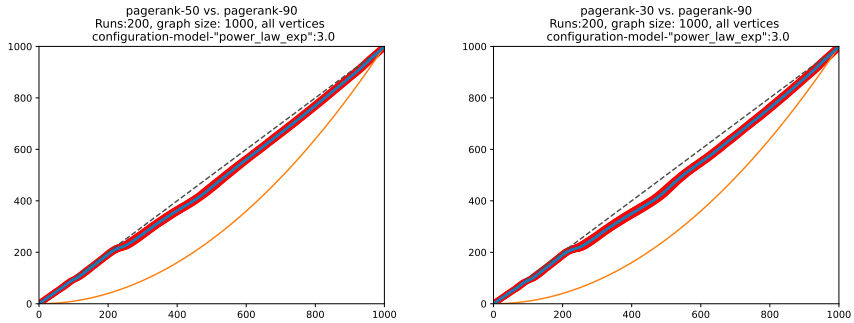


Figure D.7: CCC for PageRank with damping factors 0.5 and 0.9, and 0.3 and 0.9, respectively, for artificial network.

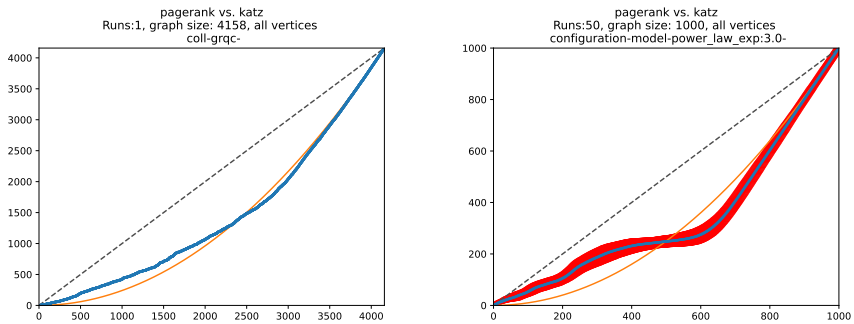


Figure D.8: CCC for PageRank and Katz, for collaboration network and artificial network.

CHAPTER 6

Open problems and suggestions for future research

This chapter gives of a summary of possible further research directions along the lines of the previous chapters.

§6.1 Ideas related to Chapter 2

Simple symmetric random walk. A natural extension of Chapter 2 would be to consider the *simple* symmetric random walk in place of the non-backtracking one. While the general idea of the framework presented therein is seemingly still useful, the stopping time that implies that mixing has occurred must change. The reason the first time that the random walk steps over a rewired edge cannot be the right candidate is simple: without the non-backtracking property, the random walk can make a step back and continue on what is essentially the initial graph. A possible candidate for an alternative stopping time, denoted by τ_{cand}, is the first time that all the following criteria are satisfied:

(a) The random walk has already stepped over a previously rewired edge. Let us call the time when this happens τ_1 and denote the edge where this crossing happens by e_1. The reason for this criterion is to ensure that the random walk has had a chance to observe a graph that is different from the initial graph.

(b) At least Δ steps have elapsed since τ_1, where the exact scaling of Δ is unclear, but it is plausibly $O(\log n)$. This is required to allow the random walk to move away from the edge e_1 that can send it back to the part of the graph that is unchanged from the initial graph.

(c) The edge e_1 has been traversed an odd number of times in the time interval $[\tau_1, \tau_{\text{cand}}]$. This is needed for the random walk to make the uniform jump used in the coupling construction. If the edge e_1 was traversed an even number of times, then the random walk would still observe the initial graph.

The above heuristic candidate stopping time is much more complicated than the stopping time used in Chapter 2 and computing its distribution might be cumbersome.

A different approach is to consider a simple symmetric random walk on a directed graph, which, under some technical assumptions, essentially behaves like a non-backtracking random walk. An example of this approach, albeit in a different context, can be found in [8].

More general graph dynamics. Chapter 2 presents a general framework that fits a large class of graph dynamics. Nevertheless, it is not difficult to think of graph dynamics where the coupling argument would fail. Since one of the major components in the argument is the uniform jump in the static geometry, it is possible to invalidate this part of the construction by considering a dynamics that gives preferential treatment to some small enough subset of "destination" half-edges. Perhaps coupling to a random walk started from these preferred half-edges might be fruitful in this setting.

§6.2 Ideas related to Chapter 3

Finite-speed random walk. It would be more natural to study a finite-speed random walk instead of the infinite-speed random walk considered in Chapter 3. Unfortunately, this setting appears to be *extremely* challenging. While it is possible to argue about regimes where either the rate of the random walk or the rate of the permutation dynamics dominates, the setting when these two rates are comparable is rather opaque. The main problem, which is not present in the infinite-speed setting, is that the random walk distribution need not mix on a permutation cycle before it gets fragmented. At this moment, we do not even have a heuristic idea about how to approach this problem.

Different stochastic processes. Dynamic permutations are an interesting example of a dynamic geometry for stochastic processes. They represent a geometry that is *disconnected*, but nevertheless allows for frequent merging and fragmentation of its constituent components. It would be interesting to study the infinite-speed version of the two-opinion voter model, which we explain next. Initially, assign every permutation element a uniform opinion. After each step of the permutation dynamics, all the elements on the same permutation cycle adopt the same opinion, namely, the majority opinion on a given cycle. In case of an equal split, pick the opinion that will prevail uniformly at random. It might be that some techniques used in Chapter 3 can be adapted to this setting, especially the ideas about the recurrence of large cycles.

§6.3 Ideas related to Chapter 4

Dynamics of the swarm. In Chapter 4, we make the simplifying assumption that the communication graph can be faithfully modelled by a *static* graph. While this assumption is well-justified in certain scenarios, it is easy to think of situations where it would simply be wrong. Therefore, a study of the dynamic situation would be of considerable value. To the best of our knowledge, dynamic k-NN graphs are an unexplored territory in the study of random graphs. Nevertheless, the results of Chapter 4 are still useful if the stationary distribution of the dynamic communication network coincides with the distribution of the static k-NN graph studied therein.

From a different perspective, it is questionable whether the k-NN graph is even the right choice for a satellite swarm in highly dynamic scenarios. It would be more natural to consider dynamic *random geometric graphs*, which would correspond to satellites with omnidirectional antennas and fixed transmission power.

Data flow modelling and dynamic optimisation. Another major simplification in Chapter 4 is that we chose not to model the flow of data in the network. In real-world satellite swarms, engineers have to consider data transmission problems, buffer allocation and optimal data routing. Incorporating these would make our results more valuable to the engineering community. Moreover, there might be further possibilities for optimisation of the swarm. For example, satellites could decide to switch themselves

off, or to reduce their transmission power, in cases when this would not influence the read-out from the interferometer.

§6.4 Ideas related to Chapter 5

Limits of CCCs. Centrality comparison curves, seen as stochastic processes, seem to often converge to some limiting process. While it makes little sense to discuss this topic in full generality, we are hopeful that some convergence results can be proven for specific *classes* of centrality measures on specific random graphs. Some work in this direction is already underway (see [56]).

Bibliography

[1] M. Alamgir and U. von Luxburg. Shortest path distance in random k-nearest neighbor graphs. In *Proceedings of the 29th International Conference on Machine Learning, ICML 2012*, volume 2, pages 1031–1038, 2012.

[2] R. Albert, H. Jeong, and A.-L. Barabási. Error and attack tolerance of complex networks. *Nature*, 406:378–382, 2000.

[3] D. Aldous. A random tree model associated with random graphs. *Random Struct. Algor.*, 1:383–402, 1990.

[4] D. Aldous. Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists. *Bernoulli*, 5:3–48, 1999.

[5] D. Aldous and J. A. Fill. Reversible Markov chains and random walks on graphs. Online version available at https://www.stat.berkeley.edu/∼aldous/RWG/book.pdf, 2002.

[6] G. Alon and G. Kozma. The probability of long cycles in interchange processes. *Duke Math. J.*, 162(9):1567–1585, 2013.

[7] J. Augustine, G. Pandurangan, and P. Robinson. Distributed algorithmic foundations of dynamic networks. *ACM SIGACT News*, 47(1):69–98, 2016.

[8] L. Avena, F. Capannoli, R. S. Hazra, and M. Quattropani. Meeting, coalescence and consensus time on random directed graphs. *arXiv preprint arXiv:2308.01832*, 2023.

[9] L. Avena, H. Güldaş, R. van der Hofstad, and F. den Hollander. Mixing times of random walks on dynamic configuration models. *Ann. Appl. Probab.*, 28(4):1977–2002, 2018.

[10] L. Avena, H. Güldaş, R. van der Hofstad, and F. den Hollander. Random walks on dynamic configuration models: A trichotomy. *Stoch. Proc. Appl.*, 129(9):3360–3375, 2019.

[11] L. Avena, H. Güldaş, R. van der Hofstad, F. den Hollander, and O. Nagy. Linking the mixing times of random walks on static and dynamic random graphs. *Stoch. Proc. Appl.*, 153:145–182, 2022.

[12] L. Avena, R. van der Hofstad, F. den Hollander, and O. Nagy. Mixing of fast random walks on dynamic random permutations. *arXiv preprint arXiv:2403.00094*, 2024.

[13] C. Avin, M. Koucký, and Z. Lotker. Cover time and mixing time of random walks on dynamic graphs. *Random Struct. Algor.*, 52:576–596, 2018.

[14] P. Balister and B. Bollobás. *Percolation in the k-nearest neighbor graph*, volume 28 of *Quaderni di matematica*, pages 83–100. Aracne Editrice, Roma, 2013.

[15] P. Balister, B. Bollobás, A. Sarkar, and M. Walters. Connectivity of random k-nearest-neighbour graphs. *Adv. Appl. Probab.*, 37(1):1–24, 2005.

[16] P. Balister, B. Bollobás, A. Sarkar, and M. Walters. A critical constant for the k nearest-neighbour model. *Adv. Appl. Probab.*, 41(1):1–12, 2009.

[17] S. Banerjee and M. Olvera-Cravioto. PageRank asymptotics on directed preferential attachment networks. *Ann. Appl. Probab.*, 32(4):3060–3084, 2022.

[18] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[19] A.-L. Barabási, R. Albert, and H. Jeong. Mean-field theory for scale-free random networks. *Phys. A*, 272(1-2):173–187, 1999.

[20] A. Ben-Hamou. A threshold for cutoff in two-community random graphs. *Ann. Appl. Probab.*, 30(4):1824–1846, 2020.

[21] A. Ben-Hamou and J. Salez. Cutoff for nonbacktracking random walks on sparse random graphs. *Ann. Probab.*, 45(3):1752–1770, 2017.

[22] M. J. Bentum, M. K. Verma, R. T. Rajan, A. J. Boonstra, C. J. M. Verhoeven, E. K. A. Gill, A. J. van der Veen, H. Falcke, M. K. Wolt, B. Monna, S. Engelen, J. Rotteveel, and L. I. Gurvits. A roadmap towards a space-based radio telescope for ultra-low frequency radio astronomy. *Adv. Space Res.*, 65(2):856–867, 2020.

[23] P. Berenbrink, G. Giakkoupis, A.-M. Kermarrec, and F. Mallmann-Trenn. Bounds on the voter model in dynamic networks. In *43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016)*, volume 55, pages 146:1–146:15, 2016.

[24] N. Berestycki and R. Durrett. A phase transition in the random transposition random walk. *Probab. Theory Relat. Fields*, 136:203–233, 2006.

[25] N. Berestycki and G. Kozma. Cycle structure of the interchange process and representation theory. *arXiv preprint arXiv:1205.4753*, 2012.

[26] N. Berestycki, E. Lubetzky, Y. Peres, and A. Sly. Random walks on the random graph. *Ann. Probab.*, 46(1):456–490, 2018.

[27] N. Berestycki, O. Schramm, and O. Zeitouni. Mixing times for random k-cycles and coalescence-fragmentation chains. *Ann. Probab.*, 39(5):1815–1843, 2011.

[28] N. Berestycki, R. van der Hofstad, and J. Salez. In preparation.

[29] J. E. Björnberg, M. Kotowski, B. Lees, and P. Miłoś. The interchange process with reversals on the complete graph. *Electron. J. Prob.*, 24:1–43, 2019.

[30] F. Bloch, M. O. Jackson, and P. Tebaldi. Centrality measures in networks. *Soc. Choice Welf.*, 61(2):413–453, 2023.

[31] D. M. Blough, M. Leoncini, G. Resta, and P. Santi. On the symmetric range assignment problem in wireless ad hoc networks. In R. Baeza-Yates, U. Montanari, and N. Santoro, editors, *Foundations of Information Technology in the Era of Network and Mobile Computing*, pages 71–82, Boston, MA, 2002. Springer US.

[32] D. M. Blough, M. Leoncini, G. Resta, and P. Santi. The k-neigh protocol for symmetric topology control in ad hoc networks. In *Proceedings of the 4th ACM International Symposium on Mobile Ad Hoc Networking & Computing*, MobiHoc03, pages 141–152, New York, NY, USA, 2003. ACM.

[33] P. Boldi, M. Santini, and S. Vigna. PageRank as a function of the damping factor. In *Proceedings of the 14th International Conference on World Wide Web*, WWW '05, pages 557–566, New York, NY, USA, 2005. ACM Press.

[34] P. Boldi and S. Vigna. Axioms for centrality. *Internet Math.*, 10(3–4):222–262, 2014.

[35] B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *Eur. J. Combin.*, 1(4):311–316, 1980.

[36] B. Bollobás. *Random Graphs*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2nd edition, 2001.

[37] C. Bordenave, P. Caputo, and J. Salez. Random walk on sparse random digraphs. *Probab. Theory Relat. Fields*, 170:933–960, 2018.

[38] C. Bordenave, P. Caputo, and J. Salez. Cutoff at the "entropic time" for sparse Markov chains. *Probab. Theory Relat. Fields*, 173(1):261–292, 2019.

[39] F. Boudin. A comparison of centrality measures for graph-based keyphrase extraction. In R. Mitkov and J. C. Park, editors, *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 834–838, Nagoya, Japan, 2013. Asian Federation of Natural Language Processing.

[40] S. Brin and L. Page. The anatomy of a large-scale hypertextual Web search engine. *Comp. Netw. ISDN Syst.*, 30(1):107–117, 1998.

[41] R. Burton and Y. Kovchegov. Mixing times via super-fast coupling. *Technical Report ORST-MATH 2011*, 2011.

[42] P. Caputo and M. Quattropani. Mixing time trichotomy in regenerating dynamic digraphs. *Stoch. Proc. Appl.*, 137:222–251, 2019.

[43] F. Caravenna, A. Garavaglia, and R. van der Hofstad. Diameter in ultra-small scale-free random graphs. *Random Struct. Algor.*, 54(3):444–498, 2019.

[44] B. Cecconi, M. Dekkali, C. Briand, B. Segret, J. N. Girard, A. Laurens, A. Lamy, D. Valat, M. Delpech, M. Bruno, P. Gélard, M. Bucher, Q. Nenon, J.-M. Grießmeier, A.-J. Boonstra, and M. Bentum. NOIRE study report: towards a low frequency radio interferometer in space. In *2018 IEEE Aerospace Conference Proceedings*, pages 1–19, 2018.

[45] S. Chatterjee and P. Diaconis. Speeding up Markov chains with deterministic jumps. *Probab. Theory Relat. Fields*, 178(3–4):1193–1214, 2020.

[46] A. E. F. Clementi, R. Silvestri, and L. Tervisan. Information spreading in dynamic graphs. In *PODC '12: Proceedings of the 2012 ACM symposium on Principles of distributed computing*, pages 37–46, 2012.

[47] C. Cooper. *Random walks, interacting particles, dynamic networks: randomness can be helpful*, volume 6796 of *Lecture Notes in Computer Science*, pages 1–14. Springer Berlin Heidelberg, 2011.

[48] P. Diaconis, E. Mayer-Wolf, O. Zeitouni, and M. P. W. Zerner. The Poisson-Dirichlet law is the unique invariant distribution for uniform split-merge transformations. *Ann. Probab.*, 32(1B):915–938, 2004.

[49] P. Diaconis and M. Shahshahani. Generating a random permutation with random transpositions. *Probab. Theory Relat. Fields*, 57(2):159–179, 1981.

[50] S. D'Oro, M. Polese, L. Bonati, H. Cheng, and T. Melodia. dApps: Distributed applications for real-time inference and control in O-RAN. *IEEE Commun. Mag.*, 60(11):52–58, 2022.

[51] R. B. Eggleton and D. A. Holton. The graph of type $(0, \infty, \infty)$ realizations of a graphic sequence. In A. F. Horadam and W. D. Wallis, editors, *Combinatorial Mathematics VI*, pages 41–54. Springer, 1979.

[52] D. Eppstein, M. S. Paterson, and F. F. Yao. On nearest-neighbor graphs. *Discrete Comput. Geom.*, 17(3):263–282, 1997.

[53] P. Erdős and A. Rényi. On random graphs I. *Publ. Math. Debrecen*, 6:290–297, 1959.

[54] T. S. Evans and B. Chen. Linking the network centrality measures closeness and degree. *Commun. Phys.*, 5(1):172, 2022.

[55] G. Exarchakos, R. van der Hofstad, O. Nagy, and M. Pandey. Bringing order to network centrality measures. Work in progress, 2024+.

[56] G. Exarchakos, R. van der Hofstad, and M. Pandey. Work in progress, 2024+.

[57] L. Federico. Almost-2-regular random graphs. *Australas. J. Comb.*, 86(1):76–96, 2023.

[58] W. Feller. *An Introduction to Probability Theory and Its Applications, Volume 1*. Wiley, 1968.

[59] D. Figueiredo, P. Nain, B. Ribeiro, E. de Souza e Silva, and D. Towsley. Characterizing continuous time random walks on time varying graphs. *ACM SIGMETRICS Perform. Eval. Rev.*, 40(1):307–318, 2012.

[60] B. Fowler. The exact history of 'Six Degrees of Kevin Bacon'. *NZ herald*, 2019.

[61] L. C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, pages 35–41, 1977.

[62] A. Frieze and N. Karoński. *Introduction to Random Graphs*. Cambridge University Press, 2016.

[63] A. Garavaglia, R. van der Hofstad, and N. Litvak. Local weak convergence for PageRank. *Ann. Appl. Probab.*, 30(1):40–79, 2020.

[64] V. Gaur, G. Soni, O. P. Yadav, and A. P. S. Rathore. Comparison between centrality measures for a network based on cascading nature of nodes. In *Lecture Notes in Networks and Systems*, pages 181–187. Springer Singapore, 2021.

[65] N. Ghazzali and A. Ouellet. Comparative study of centrality measures on social networks. In *Information Systems for Crisis Response and Management in Mediterranean Countries*, pages 3–16. Springer International Publishing, 2017.

[66] G. Giakkoupis, T. Sauerwald, and A. Stauffer. *Randomized rumor spreading in dynamic graphs*, pages 495–507. Springer Berlin Heidelberg, 2014.

[67] E. N. Gilbert. Random graphs. *Ann. Math. Stat.*, 30(4):1141–1144, 1959.

[68] K.-I. Goh, B. Kahng, and D. Kim. Universal behavior of load distribution in scale-free networks. *Phys. Rev. Lett.*, 87(27):278701, 2001.

[69] C. Goldschmidt, D. Ueltschi, and P. Windridge. Quantum Heisenberg models and their probabilistic representations. *Contemp. Math*, 552:177–224, 2011.

[70] J. Grossman. The Erdős Number Project, 2022. Available online at https://sites.google.com/oakland.edu/grossman/home/the-erdoes-number-project.

[71] H. Güldaş. Exploration on and of networks. PhD thesis, Leiden University, 2019.

[72] M. P. van Haarlem, M. W. Wise, A. W. Gunst, G. Heald, J. P. McKean, J. W. T. Hessels, et al. LOFAR: the low-frequency array. *Astron. Astrophys.*, 556, 2013.

[73] A. Hagberg, P. Swart, and D. S Chult. Exploring network structure, dynamics, and function using NetworkX. Technical report, 2008.

[74] J. Hermon and J. Salez. The interchange process on high-dimensional products. *Ann. Appl. Probab.*, 31(1):84–98, 2021.

[75] J. Hermon, A. Šarković, and P. Sousi. Cutoff for random walk on random graphs with a community structure. *arXiv preprint arXiv:2212.04469*, 2022.

[76] J. Hermon, A. Sly, and P. Sousi. Universality of cutoff for graphs with an added random matching. *Ann. Probab.*, 50(1):203–240, 2022.

[77] J. Hermon and P. Sousi. A comparison principle for random walk on dynamical percolation. *Ann. Probab.*, 48:2952–2987, 2020.

[78] R. van der Hofstad. *Random Graphs and Complex Networks*, volume 1 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, 2016.

[79] R. van der Hofstad. *Random Graphs and Complex Networks*, volume 2 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, 2024.

[80] R. van der Hofstad, G. Hooghiemstra, and P. Van Mieghem. Distances in random graphs with finite variance degrees. *Random Struct. Algor.*, 27(1):76–123, 2005.

[81] R. van der Hofstad, G. Hooghiemstra, and D. Znamenski. Distances in random graphs with finite mean and infinite variance degrees. *Electron. J. Probab.*, 12:703–766, 2007.

[82] R. van der Hofstad and J. Komjáthy. When is a scale-free graph ultra-small? *J. Stat. Phys.*, 169(2):223–264, 2017.

[83] L. Holst. The Poisson-Dirichlet distribution and its relatives revisited. *Preprint of the Royal Institute of Technology, Stockholm, Sweden*, 2001.

[84] A. Jadbabaie, J. Lin, and A. S. Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Trans. Autom. Control*, 48(6):988–1001, 2003.

[85] M. Jalili, A. Salehzadeh-Yazdi, Y. Asgari, S. S. Arab, M. Yaghmaie, A. Ghavamzadeh, and K. Alimoghaddam. Centiserver: A comprehensive resource, web-based application and R package for centrality analysis. *PLOS ONE*, 10(11):1–8, 2015.

[86] S. Janson, D. E. Knuth, T. Łuczak, and B. Pittel. The birth of the giant component. *Random Struct. Algor.*, 4(3):233–358, 1993.

[87] S. Janson, T. Luczak, and A. Rucinski. *Random Graphs*. Wiley Series in Discrete Mathematics and Optimization. Wiley, 2011.

[88] S. Janson and G. Sorkin. Successive minimum spanning trees. *Random Struct. Algor.*, 61:126–172, 2019.

[89] S. Jester and H. Falcke. Science with a lunar low-frequency array: from the dark ages of the universe to nearby exoplanets. *New Astron. Rev.*, 53(1-2):1–26, 2009.

[90] Y. Jiang, S. Pan, Y. He, and D. Zhai. Coverage algorithm of k-nearest neighbor based on communication beacon in wireless mobile sensor network. In *2019 IEEE 20th International Conference on High Performance Switching and Routing (HPSR)*, pages 1–5. IEEE, 2019.

[91] L. Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, 1953.

[92] M. Kempton. Non-backtracking random walks and a weighted Ihara's theorem. *Open J. Discrete Math.*, 06(04):207–226, 2016.

[93] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse. Identification of influential spreaders in complex networks. *Nat. Phys.*, 6(11):888–893, 2010.

[94] L. Kleinrock and J. Silvester. Optimum transmission radii for packet radio networks or why six is a magic number. In *Proceedings of the IEEE national telecommunications conference*, volume 4, pages 1–4, 1978.

[95] F. Kuhn and R. Oshman. Dynamic networks: models and algorithms. *ACM SIGACT News*, 42(1):82–96, 2011.

[96] E. Landau. Zur relativen Wertbemessung der Turnierresultate. *Deutsches Wochenschach*, 11(366-369):3, 1895.

[97] G. F. Lawler and V. Limic. *Random Walk: A Modern Introduction*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2010.

[98] M. Lee, G. Yu, and G. Y. Li. Graph embedding-based wireless link scheduling with few training samples. *IEEE Trans. Wireless Commun.*, 20(4):2282–2294, 2021.

[99] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Trans. Knowl. Discov. Data*, 1(1):2, 2007.

[100] J. Leskovec and A. Krevl. SNAP Datasets: Stanford large network dataset collection. `http://snap.stanford.edu/data`, 2014.

[101] D. Levin and Y. Peres. *Markov Chains and Mixing Times*. American Mathematical Society, 2017.

[102] C. Li, Q. Li, P. Van Mieghem, H. E. Stanley, and H. Wang. Correlation between centrality metrics and their application to the opinion model. *Eur. Phys. J. B*, 88(3):1–13, 2015.

[103] E. Lubetzky and A. Sly. Cutoff phenomena for random walks on random regular graphs. *Duke Math. J.*, 153(3):475–510, 2010.

[104] T. Łuczak and B. Pittel. Components of random forests. *Comb. Probab. Comput.*, 1:35–52, 1992.

[105] M. K. Marina and S. R. Das. Routing performance in the presence of unidirectional links in multihop wireless networks. In *Proceedings of the 3rd ACM International Symposium on Mobile Ad Hoc Networking & Computing*, MobiHoc '02, pages 12–23, New York, NY, USA, 2002. Association for Computing Machinery.

[106] P. Matthews. A strong uniform time for random transpositions. *J. Theor. Probab.*, 1:411–423, 1988.

[107] P. Van Mieghem, K. Devriendt, and H. Cetinay. Pseudoinverse of the laplacian and best spreader node in a network. *Phys. Rev. E*, 96(3), 2017.

[108] D. C. Mocanu, G. Exarchakos, and A. Liotta. Decentralized dynamic understanding of hidden relations in complex networks. *Sci. Rep.*, 8(1):1571, 2018.

[109] S. A. Molchanov and V. A. Panov. The Dickman–Goncharov distribution. *Russ. Math. Surv.*, 75(6):1089, 2020.

[110] M. Molloy and B. Reed. The size of the giant component of a random graph with a given degree sequence. *Combin. Probab. Comput.*, 7(3):295–305, 1998.

[111] J. L. Moreno and H. H. Jennings. Statistics of social configurations. *Sociometry*, 1(3/4):342–374, 1938.

[112] O. Nagy, M. Pandey, G. Exarchakos, M. Bentum, and R. van der Hofstad. Communication protocol for a satellite-swarm interferometer. *arXiv preprint arXiv:2312.15814*, 2023.

[113] M. E. J. Newman. The structure and function of complex networks. *SIAM Rev.*, 45(2):167–256, 2003.

[114] M. E. J. Newman. *Networks.* Oxford University Press, 2018.

[115] M. E. J. Newman and D. J. Watts. Renormalization group analysis of the small-world network model. *Phys. Lett. A*, 263(4):341–346, 1999.

[116] M. E. J. Newman and D. J. Watts. Scaling and percolation in the small-world network model. *Phys. Rev. E*, 60(6):7332–7342, 1999.

[117] J. Ni and S. A. G. Chandler. Connectivity properties of a random radio network. *IEE Proceedings-Communications*, 141(4):289–296, 1994.

[118] J. Oh and J. Kim. Adaptive k-nearest neighbour algorithm for WiFi fingerprint positioning. *ICT Express*, 4(2):91–94, 2018.

[119] S. Oldham, B. Fulcher, L. Parkes, A. Arnatkevičiūtė, C. Suo, and A. Fornito. Consistency and differences between centrality measures across distinct classes of networks. *PLOS ONE*, 14(7):e0220061, 2019.

[120] S. Olesker-Taylor. Cutoff for rewiring dynamics on perfect matchings. *Ann. Appl. Probab.*, 33(1):641–676, 2023.

[121] M. Othon and P. G. Spirakis. Elements of the theory of dynamic networks. *Commun. ACM*, 61(2):72–81, 2018.

[122] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank citation ranking: bringing order to the web. Stanford InfoLab Technical Report 1999-66, 1999.

[123] Y. Peres, P. Sousi, and J. E. Steif. Mixing time for random walk on supercritical dynamical percolation. *Probab. Theory Relat. Fields*, 176(3):809–849, 2020.

[124] Y. Peres, A. Stauffer, and J. E. Steif. Random walks on dynamical percolation: mixing times, mean squared displacement and hitting times. *Probab. Theory Relat. Fields*, 162(3):487–530, 2015.

[125] K. E. Petersen. *Ergodic Theory*. Cambridge University Press, 1989.

[126] R. Pung, J. A. Firth, L. G. Spurgin, A. Chang, J. Kong, J. Wong, O. J. Jin, et al. Using high-resolution contact networks to evaluate SARS-CoV-2 transmission and control in large-scale multi-day events. *Nat. Commun.*, 13(1):1956, 2022.

[127] O. Riordan and L. Warnke. The phase transition in bounded-size Achlioptas processes, 2017.

[128] J. Salez. A sharp log-Sobolev inequality for the multislice. *Ann. H. Lebesgue*, 4:1143–1161, 2021.

[129] J. Salez. Cutoff for non-negatively curved Markov chains. *J. Eur. Math. Soc.*, 2023.

[130] J. Salez. The varentropy criterion is sharp on expanders. *arXiv preprint arXiv:2307.10066*, 2023.

[131] L. Saloff-Coste. Random walks on finite groups. In *Encyclopaedia of Mathematical Sciences*, pages 263–346. Springer Berlin Heidelberg, 2004.

[132] A. D. Sarma, A. R. Molla, and G. Pandurangan. Distributed computation in dynamic networks via random walks. *Theor. Comput. Sci.*, 581:45–66, 2015.

[133] T. Sauerwald and L. Zanetti. Random walks on dynamic graphs: mixing times, hitting times, and return probabilities. In *46th International Colloquium on Automata, Languages,and Programming (ICALP 2019)*, volume 132, pages 93:1–93:15, 2019.

[134] A. Saxena and S. Iyengar. Centrality measures in complex networks: a survey. *arXiv preprint arXiv:2011.07190*, 2020.

[135] O. Schramm. Compositions of random transpositions. *Israel J. Math.*, 147:221–243, 2005.

[136] A. Sobehy, E. Renault, and P. Mühlethaler. CSI-MIMO: k-nearest neighbor applied to Indoor Localization. In *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, pages 1–6, 2020.

[137] R. Solomonoff and A. Rapoport. Connectivity of random nets. *Bull. Math. Biophys.*, 13(2):107–117, 1951.

[138] P. Sousi and S. Olesker-Taylor. Cutoff for random walk on dynamical Erdős–Rényi graph. *Ann. Inst. Henri Poincaré Probab. Stat.*, 56:2745–2773, 2020.

[139] J. Spencer and N. Wormald. Birth control for giants. *Combinatorica*, 27:587–628, 2007.

[140] I. Teotonio. Google adds six degrees of Kevin Bacon to search engine. *Toronto Star*, 2012.

[141] A. R. Thompson, J. M. Moran, and G. W. Swenson. *Interferometry and Synthesis in Radio Astronomy*. Springer International Publishing, 2017.

[142] B. Tóth. Improved lower bound on the thermodynamic pressure of the spin 1/2 Heisenberg ferromagnet. *Lett. Math. Phys.*, 28(1):75–84, 1993.

[143] N. V. Tsilevich. Stationary random partitions of positive integers. *Theory Probab.*, 44(1):60–74, 2000.

[144] D. Vial and V. Subramanian. Restart perturbations for lazy, reversible Markov chains: trichotomy and pre-cutoff equivalence, 2019.

[145] K. Vignery and W. Laurier. A methodology and theoretical taxonomy for centrality measures: what are the best centrality indicators for student networks? *PLOS ONE*, 15(12), 2020.

[146] Z. Wan, Y. Mahajan, B. W. Kang, T. J. Moore, and J.-H. Cho. A survey on centrality metrics and their network resilience analysis. *IEEE Access*, 9, 2021.

[147] A. Q. Wang, M. Pollock, G. O. Roberts, and D. Steinsaltz. Regeneration-enriched Markov processes with application to Monte Carlo. *Ann. Appl. Probab.*, 31(2), 2021.

[148] J.-B. Wang, J. Wang, Y. Wu, J.-Y. Wang, H. Zhu, M. Lin, and J. Wang. A machine learning framework for resource allocation assisted by cloud computing. *IEEE Netw.*, 32(2):144–151, 2018.

[149] Z. Wang, C. T. Bauch, S. Bhattacharyya, A. d'Onofrio, P. Manfredi, M. Perc, N. Perra, M. Salathé, and D. Zhao. Statistical physics of vaccination. *Phys. Rep.*, 664, 2016.

[150] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.

[151] W. Whitt. *Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues*. Springer Series in Operations Research and Financial Engineering. Springer New York, 2002.

[152] Y. Xie, Y. Wang, A. Nallanathan, and L. Wang. An improved k-nearest-neighbor indoor localization method based on spearman distance. *IEEE Signal Process. Lett.*, 23(3):351–355, 2016.

[153] F. Xue and P. R. Kumar. The number of neighbors needed for connectivity of wireless networks. *Wirel. Netw.*, 10(2):169–181, 2004.

List of publications

The results presented in this dissertation can be found in the following publications:

- L. Avena, H. Güldaş, R. van der Hofstad, F. den Hollander, and O. Nagy. Linking the mixing times of random walks on static and dynamic random graphs. *Stoch. Proc. Appl.*, 153:145–182, 2022
- L. Avena, R. van der Hofstad, F. den Hollander, and O. Nagy. Mixing of fast random walks on dynamic random permutations. *arXiv preprint arXiv:2403.00094*, 2024.
  Under consideration in *Probability Theory and Related Fields.*
- O. Nagy, M. Pandey, G. Exarchakos, M. Bentum, and R. van der Hofstad. Communication protocol for a satellite-swarm interferometer. *arXiv preprint arXiv:2312.15814*, 2023.
  Under review in *IEEE Transactions on Aerospace and Electronic Systems.*

This list is up-to-date as of October 7, 2024.

Samenvatting

Stel je een kopje warme thee voor, waar je wat melk in giet in. Aanvankelijk ziet de vloeistof er misschien wat chaotisch uit, maar uiteindelijk wordt deze egaal van kleur. Als we in het kopje zouden roeren, dan zou de uniforme bruine kleur zich nog sneller ontwikkelen.

De concepten in de bovenstaande paragraaf reiken veel verder dan warme dranken. Veel wanordelijke systemen vertonen ordelijk gedrag na verloop van tijd. Met andere woorden, ze naderen een evenwicht. Net zoals de kleur van de drank kan worden gebruikt om het mengen van thee en melk kwantitatief te beschrijven, zo zijn er abstracte kwantitatieve methoden om te meten hoe ver een systeem van zijn evenwicht verwijderd is. Een prominent voorbeeld voor zo'n methode is de totale variatieafstand. Bovendien zullen we doorgaans een snellere vermenging waarnemen wanneer er een externe dynamica wordt toegepast. Dit geldt zowel voor een kopje thee als voor toevalswandelingen op dynamische toevallige grafen.

Het eerste wetenschappelijke hoofdstuk in dit proefschrift laat zien dat, voor niet-teruggaande toevalswandelingen op toevallige grafen gegenereerd door het configuratiemodel met een herverbindingsdynamica, vermenging over het algemeen sneller optreedt dan op de statische versie van de toevallige graaf. Daarnaast wordt een gedetailleerde analyse van dit scenario gegeven en wordt aangetoond dat de eerder waargenomen driedeling in het gedrag van de totale variatieafstand eigenlijk een ontaarde vorm van een zesdeling is.

Het tweede wetenschappelijke hoofdstuk presenteert een nog extremer voorbeeld van versnelde vermenging. De oneindig snelle toevalswandeling op een dynamische toevallige permutatie zou nooit mengen zonder de invloed van de dynamica. In feite zou deze altijd in de begintoestand blijven, ongeacht de verstreken tijd. Het samenvoegen en splitsen van permutatie-cycli, veroorzaakt door de dynamica, stelt de toevalswandeling in staat om zichzelf gelijkmatig te verdelen over alle elementen van de onderliggende permutatie. Bovendien laat de totale variatieafstand een opvallend patroon zien: deze blijft aanvankelijk zo ver mogelijk van het evenwicht, daalt na een toevallige tijd plotseling, en volgt een deterministisch pad naar volledige vermenging.

Soms kan een zeer dynamische situatie worden gereduceerd tot zijn statische tegenhanger. Dit is het centrale thema van het derde wetenschappelijke hoofdstuk, waarin het communicatienetwerk binnen een satellietzwerm wordt onderzocht. Hoewel de satellieten in werkelijkheid voortdurend op een complexe manier bewegen, is het mogelijk om in plaats daarvan het statische netwerk te bestuderen dat ontstaat als een langetermijngemiddelde. Dit hoofdstuk onderzoekt de prestaties van een dergelijk netwerk en de energiekosten die nodig zijn om het te onderhouden.

Onderzoeksprojecten kunnen ook onverwachte bijproducten opleveren. Een dergelijk

bijproduct van het satellietproject is een kwantitatieve methode voor het vergelijken van willekeurige paren van centraliteitsmaten, gedefinieerd op dezelfde graaf. Dit nieuwe wiskundige object wordt geïntroduceerd in het vierde hoofdstuk, waarin ook vele aantrekkelijke eigenschappen worden belicht en talrijke numerieke voorbeelden worden gegeven.

Summary

Imagine a cup of warm tea, and pour some milk into it. Initially, the liquid might appear chaotic, but eventually it will become uniformly coloured. And, if we were to stir the cup, then the uniform brown colour would develop even faster.

The concepts in the paragraph above extend far beyond the realm of hot beverages. Many random systems exhibit orderly behaviour after a sufficiently large time. In other words, they approach equilibrium. Just as the colour of the beverage can be used to quantitatively describe the mixing of tea and milk, there are abstract quantitative methods to measure how far a system is from its equilibrium. A prominent example of such a method is the total variation distance. Moreover, if external dynamics are applied, then we typically observe faster mixing. This is true both for a cup of tea and for random walks on dynamic random graphs.

The first scientific chapter in this thesis shows that, for non-backtracking random walks on the random graph generated by the configuration model with rewiring dynamics, mixing generally occurs faster than it would on the static version of the random graph. Additionally, this chapter provides a detailed analysis of this scenario and reveals that the previously observed three-way split in the behaviour of the total variation distance is actually a degenerate form of a six-way split.

The second scientific chapter presents an even more extreme example of accelerated mixing. The infinite-speed random walk on a dynamic random permutation would never mix without the influence of the dynamics. In fact, it would always remain in the initial state, regardless of the time that has elapsed. The merging and splitting of permutation cycles caused by the dynamics allow the random walk to evenly distribute itself across all elements of the underlying permutation. Moreover, the total variation distance illustrates a striking pattern: it initially stays as far from equilibrium as possible, then after a random time suddenly drops, and subsequently follows a deterministic path towards full mixing.

Occasionally, a highly dynamic situation can be reduced to its static counterpart. This is the central theme of the third scientific chapter, which examines the communication network within a satellite swarm. While in reality the satellites are constantly moving in a complex manner, it is possible instead to study the static network that emerges as a long-term average. This chapter explores the performance of such a network and the energy costs required to maintain it.

Research projects can also produce unexpected by-products. One such by-product of the satellite project is a quantitative method for comparing arbitrary pairs of centrality measures defined on the same graph. This new mathematical object is introduced in the fourth chapter, which also highlights its many appealing properties and provides numerous numerical examples.

Acknowledgements

Regulations command both brevity and restraint in these acknowledgments, which therefore cannot capture my full gratitude to all those who left their mark on the time of my doctoral studies.

Nevertheless, I want to use this opportunity to thank:

- **Frank den Hollander, Remco van der Hofstad, and Luca Avena**, for being my supervisors and co-authors, for all our discussions, and for letting me glimpse the life of a working mathematician at three different career stages.

  I would like to especially thank Remco for nominating me for the MSRI Graduate Summer School, which was one of the best experiences of my graduate studies.

- **Dutch Research Council**, for financing my doctoral research via the Gravitation grant NETWORKS.

- **Walter Jaffe**, for all his advice that helped me navigate the peculiarities of the academic life, for discussions about astronomy and interferometry (which found an unexpected use in Chapter 4), for extending his professional network to me at the end of my studies, and for providing me with a roof over my head after I came to Leiden.

- **Mark Bentum, George Exarchakos and Manish Pandey**, for the work we did together in the projects about satellite swarms and centrality measures.

- **Camiel Koopmans, Daoyi Wang, Twan Koperberg, Federico Capannoli, and Pierfrancesco Dionigi**, for all the time spent together fighting problems in probability theory and (un-)related fields. I especially thank Federico for co-organising the iPOD seminar with me.

- **My family**, for their unwavering support during my studies and for their encouragement when it was needed.

- And to all the other people who enabled me to carry out the research in this dissertation.

There was more to the last couple of years than just work and research, and I was fortunate to experience wonderful moments that had little to nothing to do with either. While this is not an appropriate place to recount these, I am grateful to all the people with whom I could share these moments.

Curriculum vitæ

Oliver Nagy was born in the summer of 1994 in Bratislava, Slovakia.

After graduating from Gymnázium M. R. Štefánika in Šamorín, Slovakia, he started his university studies at the Faculty of Mathematics and Physics of Charles University in Prague, Czech Republic. There he was admitted to the degrees of bachelor (February 2017, General Physics) and master (February 2020, Theoretical Physics).

During his studies in Prague, he worked in the computer laboratories of his former faculty and did an internship with an IT-focused start-up.

In February 2020, he started his PhD studies at the Mathematical Institute of Leiden University, within the framework of the NETWORKS consortium. His employment at the Mathematical Institute came to an end in April 2024, and this dissertation is one of the fruits of this labour.

Oliver's career continues outside of academia. After a brief stint as a quantitative risk analyst at ABN AMRO, he began his training to become a patent attorney at the Dutch branch of Kilburn & Strode. He maintains a connection to the mathematical community through his role as a junior editor for *Snapshots of Modern Mathematics from Oberwolfach*.