
Online Matching in Geometric Random Graphs

Flore Sentenac
CREST, ENSAE
Palaiseau, France
flore.sentenac@ensae.fr

Nathan Noiry
S2A
Telecom Paris, France
nathan.noiry@telecom-paris.fr

Matthieu Lerasle
CREST, ENSAE
Palaiseau, France
matthieu.lerasle@ensae.fr

Laurent Ménard
Modal'X, UMR CNRS 9023, UPL, Univ. Paris Nanterre
F92000 Nanterre, France
laurent.menard@normalesup.org

Vianney Perchet
CREST, ENSAE Paris
Palaiseau, France
CRITEO AI Lab, Paris, France
vianney.perchet@normalesup.org

Abstract

We investigate online maximum cardinality matching, a central problem in ad allocation. In this problem, users are revealed sequentially, and each new user can be paired with any previously unmatched campaign that it is compatible with. Despite the limited theoretical guarantees, the greedy algorithm, which matches incoming users with any available campaign, exhibits outstanding performance in practice. Some theoretical support for this practical success has been established in specific classes of graphs, where the connections between different vertices lack strong correlations – an assumption not always valid in real-world situations. To bridge this gap, we focus on the following model: both users and campaigns are represented as points uniformly distributed in the interval $[0, 1]$, and a user is eligible to be paired with a campaign if they are "similar enough," meaning the distance between their respective points is less than c/N , where $c > 0$ is a model parameter. As a benchmark, we determine the size of the optimal offline matching in these bipartite random geometric graphs. We achieve this by introducing an algorithm that constructs the optimal matching and analyzing it. We then turn to the online setting and investigate the number of matches made by the online algorithm CLOSEST, which pairs incoming points with their nearest available neighbors in a greedy manner. We demonstrate that the algorithm's performance can be compared to its fluid limit, which is completely characterized as the solution to a specific partial differential equation (PDE). From this PDE solution, we can compute the competitive ratio of CLOSEST, and our computations reveal that it remains significantly better than its worst-case guarantee. This model turns out to be closely related to the online minimum cost matching problem, and we can extend the results obtained here to refine certain findings in that area of research. Specifically, we determine the exact asymptotic cost of CLOSEST in the ϵ -excess regime, providing a more accurate estimate than the previously known loose upper bound.

1 Introduction

Online matching is motivated, among others, by its application to ad allocation on the internet. Advertising platforms handle multiple companies paying for relevant ad space, and their goal is to maximize the number of valid ad-user allocations while adhering to budget constraints.

In this dynamic setting, users generate web pages sequentially, and ads must be allocated immediately to avoid forfeiting available ad slots. This dynamic allocation process can be represented as an online matching problem.

Formally, consider the bipartite graph $G(\mathcal{X}, \mathcal{Y}, \mathcal{E})$, with set of vertices $\mathcal{X} \cup \mathcal{Y}$ and set of edges $\mathcal{E} \in \mathcal{X} \times \mathcal{Y}$. In our context of online matching, the “offline” vertices $\mathcal{U}$ represent the side of the companies, they are present from the beginning, while the vertices $\mathcal{V}$, which represent the users, are revealed sequentially. At each arrival of a vertex $v \in \mathcal{V}$, the edges adjacent to vertex v are revealed. Based on those edges, the vertex can then be matched irrevocably to a previously unmatched neighbor. The excellent survey Mehta [2012] details in length applications, results, and techniques of online matching.

The first line of work studied online matching in the adversarial framework, where the algorithm is evaluated on the worst possible instance and vertex arrival order. It is folklore that greedy random algorithms, which match incoming vertices to any available neighbor have a competitive ratio of $1/2$ in the worst case. However, they achieve $1 - 1/e$ as soon as the incoming vertices arrive in Random Order, [Goel and Mehta, 2008]. The RANKING algorithm is the worst-case optimal, it achieves at least $1 - 1/e$ on any instance [Karp et al., 1990, Devanur et al., 2013, Birnbaum and Mathieu, 2008], and also has a higher competitive ratio in the Random Order setting [Mahdian and Yan, 2011]. Beyond this worst-case setting, the known i.i.d. model assumes there exists a probability distribution over types of vertices, from which the incoming vertex is drawn i.i.d. at every iteration. With the knowledge of that distribution, algorithms with much better competitive ratios than RANKING were designed [Manshadi et al., 2012, Jaillet and Lu, 2014, Brubach et al., 2019, Huang et al., 2022], the best one to date achieving a competitive ratio of 0.711.

This known i.i.d. model fits some situations and is certainly interesting, but still very general. The algorithms designed are tailored to the worst known i.i.d. model and fail to handle additional graph knowledge. Moreover, as the guarantee is given for the worst possible input distribution, it does not always reflect the average performance of those algorithms. It has been highlighted in Borodin et al. [2018] that on many average-case and practical input families, simple greedy strategies outperform or perform comparably to state-of-the-art algorithms designed for that known i.i.d. setting. They thus call for the formulation of new stochastic input models that better match practical inputs for certain application domains, e.g., online advertising.

As a consequence, another line of work considers standard online algorithms on some classes of random graphs, representing situations where some properties of the underlying graph are known. The seminal example would be online matching in Erdos-Renyi graphs [Mastin and Jaillet, 2013], or more generally in the configuration model that specifies a law on the degrees of the vertices [Noiry et al., 2021, Aamand et al., 2022]. The idea behind the latter is that one can “estimate” the typical popularity of a campaign (say, some of them target a large number of users while others are more selective). In those instances, greedy strategies can be precisely analyzed. For instance, in [Mastin and Jaillet, 2013], they show that the competitive ratio of greedy is larger than 0.837, which is much higher than the worst-case guarantee.

Unfortunately, these approaches fail to model correlations between edges. Campaigns that are “similar” tend to target the same users, and vice-versa (for instance, luxury products will target users with high incomes, while baby products obviously target families). A possible approach to model these correlations is through space embedding techniques: ads and users are represented by points in an Euclidian space, typically feature vectors, and an edge is present between two vertices if the points are close enough.

We shall in the following introduce and analyze the online matching problem for space-embedded graphs, called geometric graphs. As it is already challenging and interesting, we shall focus on the one-dimensional geometric graph.

This model turns out to be strongly related to the minimum cost metric matching one. Similarly to the features-embedded setting, vertices are points of some Euclidean (or any metric) space, and any vertices on opposite sides of the partition may be matched together. The cost of the matching is the total distance between the matched points [Gupta et al., 2019].

A major difference between the two models is that in the metric matching case, the incoming vertex is necessarily matched at every iteration. The cost function is also completely different: in the online matching setting, any valid match generates the same reward, which is not the case in the online metric matching setting. Thus no algorithm tailored for one of those problems generalizes easily to the other, which implies that the works on both topics are relatively independent.

The CLOSEST algorithm, which consists in matching greedily the incoming vertex to its closest available neighbor, has received some attention in that field. A first work, [Akbarpour et al., 2021], focuses on metric matching for a single set of N points drawn i.i.d. on the line. The authors show that the greedy algorithm that successively matches the two closest points together, has a competitive ratio of $O(\log(N))$. Later, back in the bipartite setting, some works focused on the so-called excess supply setting, where N offline points are drawn i.i.d. on the line and $N(1 - \epsilon)$ vertices arrive sequentially. In that setting, the cost of CLOSEST was first shown to be $O\left(\log(N)^3\right)$ [Akbarpour et al., 2021]. That result was later refined to $O(\frac{1}{\epsilon})$ in [Balkanski et al., 2022]. In that work, they also demonstrate a constant competitive ratio for CLOSEST when there is an equal number of vertices on both sides, as well as under a semi-random model, where the incoming vertices are picked adversarially.

Extending the techniques developed for the maximum cardinality matching case, we can precise the result in the ϵ -excess model and give the exact asymptotic cost of CLOSEST under that setting.

1.1 Model and Contributions

We consider the class of bipartite geometric random graphs $\text{Geom}(c, N)$, for some $c > 0$, whose set of vertices $\mathcal{X}$ and $\mathcal{Y}$ are two sets of N points drawn independently and uniformly in $[0, 1]$ (representing the features of users/companies),

$$\mathcal{X} = (x_i)_{i \in [N]} \stackrel{\text{i.i.d.}}{\sim} \mathcal{U}[0, 1] \quad \text{and} \quad \mathcal{Y} = (y_i)_{i \in [N]} \stackrel{\text{i.i.d.}}{\sim} \mathcal{U}[0, 1],$$

and there is an edge between $x_i \in \mathcal{X}$ and $y_j \in \mathcal{Y}$ iff these features are "similar" enough, i.e.,

$$|x_i - y_j| < \frac{c}{N}.$$

This choice of parametrization ensures that the expected degree of a vertex remains bounded, of order c (neglecting the boundary effects). In particular, the graph remains sparse and the online matching problem is not trivial. In the following, we shall also sometimes refer to geometric graphs $\text{Geom}(\mathcal{U}, \mathcal{V}, c, N)$ with pre-specified sets of vertices $\mathcal{U}$ and $\mathcal{V}$ and edges generated by the same process as in $\text{Geom}(c, N)$.

In the context of metric matching, we denote the graph $\text{Geom}(N)$, whose set of vertices $\mathcal{X}$ and $\mathcal{Y}$ are two sets of N points drawn independently and uniformly in $[0, 1]$, and the length of the edges corresponds to the distance between the points. Similarly, we shall also sometimes refer to geometric graphs $\text{Geom}(\mathcal{U}, \mathcal{V}, N)$ with pre-specified sets of vertices $\mathcal{U}$ and $\mathcal{V}$. $\mathcal{Y}_{1-\epsilon}$ is a set of $N(1 - \epsilon)$ points drawn i.i.d. in $[0, 1]$.

First, we derive the size of the maximum matching in these geometric random graphs, as a function of the parameter c , along with the description of an algorithm constructing this matching. We provide a non-asymptotic convergence bound of the size of the maximum matching to $c/(c + 1/2)$. This bound is obtained through the study of the algorithm, via the construction of a potential function which is then treated as a random walk. More precisely, we shall prove the following theorem.

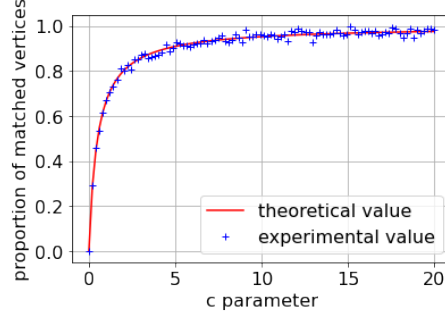


Figure 1: The asymptotic optimal offline matching size is displayed in red, as a function of the parameter c . Several simulations (blue crosses) for different values of c and for $N = 100$ vertices illustrate that this limit is reached rapidly.

THEOREM 1 (informal). Let $m^*(c, N)$ be the size of the maximum matching in an instance of $\text{Geom}(c, N)$. With probability at least $1 - O(\frac{1}{N})$,

$$m^*(c, N) = \frac{c}{c + \frac{1}{2}} N + O\left(\sqrt{N \ln(N)}\right).$$

This informal result is illustrated in Figure 1.1 which shows both the theoretical asymptotic value and the optimal value in several realizations of random graphs for a variety of parameters c .

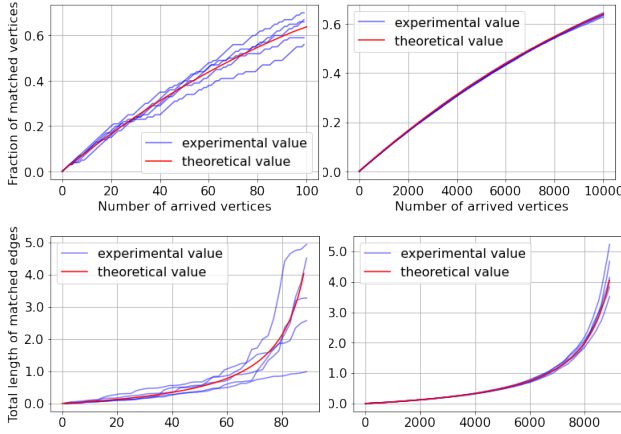


Figure 2: First row, from left to right: theoretical (red line) vs. experimental (blue lines) sizes of the online matching in the 1D uniform geometric graph ($c = 1$) as a function of the number of arrived vertices, for $N = 100$ and $N = 10,000$. Second row, from left to right: theoretical (red line) vs. experimental (blue lines) length of the online matching in the 1D uniform geometric graph as a function of the number of arrived vertices, for $N = 100$ and $N = 10,000$.

We then study the size of the matching constructed by the online algorithm CLOSEST, which matches any incoming vertex to its closest available neighbor. We show the convergence of this quantity to the solution of an explicit PDE. We do that by exhibiting tractable quantities that can be approximated via the Differential Equation Method [Wormald, 1995].

More precisely, we shall prove the following Theorem.

THEOREM 2 (informal). Let $\kappa(c, N)$ be the size of the matching obtained by the CLOSEST algorithm on an instance of $\text{Geom}(c, N)$, then

$$\frac{\kappa(c, N)}{N} \xrightarrow[N \rightarrow +\infty]{\mathbb{P}} 1 - \int_0^{+\infty} f(x, t) dx$$

where f is the solution of some explicit PDE, described later in Equation (2).

The difference between the theoretical and the actual sizes of the matchings (as a function of the number of online vertices observed) for different values of N is illustrated in Figure 2. Those two

results combined give the value of the competitive ratio of the CLOSEST algorithm in $\text{Geom}(c, N)$, illustrated in Figure 3.

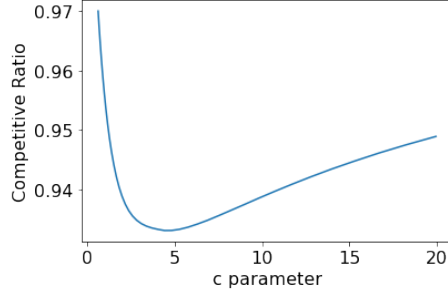


Figure 3: Asymptotic value of the competitive ratio of CLOSEST for several values of c .

We also extend the technique developed to the online metric matching problem, and prove the following theorem, illustrated in Figure 2.

THEOREM 3 (informal). *Let $\rho[N](t)$ be the length of the matching obtained by CLOSEST on an instance of $\text{Geom}(N)$ when $tN \leq N(1 - \epsilon)$ vertices have arrived, then*

$$\frac{\rho[N](t)}{N} \xrightarrow[N \rightarrow +\infty]{\mathbb{P}} \frac{1}{2} \left[\frac{1}{1-t} - 1 \right].$$

Organization of the paper Section 2 is dedicated to the offline case. We give the asymptotic formula of the size of the optimal offline matching. Our approach is algorithmic: we provide a way to construct this optimal matching, and we analyze the latter by carefully studying some random walk. Section 3 focuses on the online case. We focus on a simple greedy algorithm, CLOSEST, that matches any incoming vertex to the closest available one (as a greedy procedure would do) and we characterize the size/length of the matching it creates by studying its fluid limit. We shall prove it satisfies some PDE.

2 Maximum Matching in 1D Uniform Geometric Graphs

This section is dedicated to the offline case, where the whole underlying bipartite 1D geometric graph is known from the beginning. Consider algorithm SMALL-FIRST which iteratively matches the unmatched vertices with the smallest coordinates. The following proposition states this algorithm produces a maximum matching.

Proposition 1 *Algorithm SMALL-FIRST returns an optimal matching in 1D geometric graphs.*

Proof: First, there exists a perfect matching in which no two edges cross. Indeed, assume pairs (x, y') and (x', y) belong to the matching with $x < x'$ and $y < y'$. Since the graph is a geometric graph, there are also the edges (x, y) and (x', y') , so the matches can be uncrossed without modifying the size of the matching.

Among those perfect matches with no crossing edges, there is one in which each vertex is matched to its leftmost neighbor not already matched to another vertex. Again, if this is not the case, the matching can be modified without modifying its size.

This perfect matching with no crossing edges and where every vertex is matched to its leftmost available neighbor is the one returned by the algorithm SMALL-FIRST. $\square$

Using the optimality of SMALL-FIRST, we can now prove the first main result.

Theorem 1 *Let $M^*(c, N)$ be the expected size of the maximum matching in $\text{Geom}(c, N)$. The following holds:*

$$\lim_{N \rightarrow \infty} \frac{M^*(c, N)}{N} = \frac{c}{c + \frac{1}{2}}.$$

Let $m^*(c, N)$ be the size of the maximum matching in an instance of $\text{Geom}(c, N)$. With probability at least $1 - O\left(\frac{1}{N}\right)$:

$$m^*(c, N) = \frac{c}{c + \frac{1}{2}} N + O\left(\sqrt{N \ln(N)}\right).$$

Proof: The proof is decomposed into two steps. In the first step, we show that the size of the maximum matching in $\text{Geom}(c, N)$ is related to the size of the maximum matching in a geometric graph with a set of vertices generated by Poisson point processes. In the second step, the asymptotic size of the maximum matching in those graphs is derived through the study of a random walk.

Step 1, connection with Poisson point processes. Let $\Phi_{\mathcal{U}}^N$ be an independent homogeneous Poisson point process on the segment $[0, 1]$ of intensity N . The definition and some standard properties of Poisson Point Processes (PPP) are reported in Appendix A for the sake of completeness. Let $\mathcal{U}^N \sim \Phi^N$ and $\mathcal{V}^N \sim \Phi^N$ denote two independent PPP, and $\gamma^*(c, N)$ be the expected size of the maximum matching in $\text{Geom}(\mathcal{U}^N, \mathcal{V}^N, c, N)$.

Lemma 1 *Under the above notations, the following holds:*

$$|\gamma^*(c, N) - M^*(c, N)| \leq 4(1 + \sqrt{N \ln N}).$$

Proof: Let $N_{\mathcal{U}} = |\mathcal{U}|$ and $N_{\mathcal{V}} = |\mathcal{V}|$ denote the size of the vertex sets. By Chernoff bound:

$$P\left\{|N_{\mathcal{U}} - N| \geq 2\sqrt{N \ln N}\right\} \leq \frac{2}{N} \quad (1)$$

and the same holds for $N_{\mathcal{V}}$. We now explain how to recover the construction of $\text{Geom}(\mathcal{X}, \mathcal{Y}, c, N)$ from those two PPP (as they do not have the same vertex set sizes). From $\mathcal{U}$, define a new set $\tilde{\mathcal{U}}$ of N vertices as follows. If $N_{\mathcal{U}} > N$, delete uniformly at random $N_{\mathcal{U}} - N$ points from $\mathcal{U}$. If $N > N_{\mathcal{U}}$, add $N - N_{\mathcal{U}}$ points independently and uniformly distributed in $[0, 1]$ to $\mathcal{U}$. The set $\tilde{\mathcal{V}}$ is constructed from $\mathcal{V}$ similarly. This construction ensures that $\text{Geom}(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, c, N)$ and $\text{Geom}(\mathcal{X}, \mathcal{Y}, c, N)$ have the same law. Moreover, the transformation affects the size of the matching at most by the number of added and removed points. $\square$

Step 2, deriving $\gamma^*(c, N)$. A possible way to draw set $\mathcal{U}$ from ϕ^N is through a renewal process with exponential holding times. Precisely, let $\{F_k\}_{k \in \mathbb{N}}$ be a sequence of i.i.d. exponential random variables of parameter N . Ensemble $\mathcal{U}$ is defined as:

$$\mathcal{U} = \left\{ u_k = \sum_{i=1}^k F_i \text{ for } k \text{ s.t. } \sum_{i=1}^k F_i < 1 \right\}.$$

The same holds from $\mathcal{V}$.

To compute the size of the maximum matching in a given graph, we introduce a modified version of Algorithm SMALL-FIRST, that generates the graph together with the matching. Algorithm SMALL-FIRST-GENERATIVE (Algorithm 1) proceeds as follows. The positions of the first points in $\mathcal{U}$ and $\mathcal{V}$ are drawn independently from two exponential distributions of parameter N . At iteration t , note $u_{x(t)}$ and $v_{x(t)}$ the position of the last generated points in $\mathcal{U}$ and $\mathcal{V}$ respectively. Define the potential function

$$\psi(t) := u_{x(t)} - v_{x(t)}.$$

Until all the points on one side have been drawn, the following operations are iteratively performed:

- if $|\psi(t)| < \frac{c}{N}$, edge $(u_{x(t)}, v_{x(t)})$ is added to the matching and the next points on both sides of the graph are generated,
- if $\psi(t) > c$, the next point in $\mathcal{V}^N$ is generated,
- if $\psi(t) < -c$, the next point in $\mathcal{U}^N$ is generated.

Let τ_N be number of iterations in Algorithm SMALL-FIRST-GENERATIVE run with input parameter N . When $|\psi(t)| \leq c$, 2 vertices are matched, and when $|\psi(t)| > c$, 1 vertex is left unmatched, so

Algorithm 1: SMALL-FIRST-GENERATIVE

input: N, c ;
Draw $u_1 \sim \text{Exp}(N)$ and $v_1 \sim \text{Exp}(N)$;
Initialize $x(1) \leftarrow 1, y(1) \leftarrow 1$ and $m \leftarrow \emptyset$;
Define for $t = 1, \dots, \psi(t) \leftarrow u_{x(t)} - v_{x(t)}$;
while $u_{x(t)} < 1$ and $v_{x(t)} < 1$ **do**
 if $|\psi(t)| < c$ **then**
 $m \leftarrow m \cup (u_{x(t)}, v_{x(t)})$;
 $x(t+1) \leftarrow x(t) + 1$ and $y(t+1) \leftarrow y(t) + 1$;
 Draw $u_{x(t+1)} - u_{x(t)} \sim \text{Exp}(N)$ and $v_{y(t+1)} - v_{y(t)} \sim \text{Exp}(N)$;
 end
 if $\psi(t) > c$ **then**
 $x(t+1) \leftarrow x(t)$ and $y(t+1) \leftarrow y(t) + 1$;
 Draw $v_{y(t+1)} - v_{y(t)} \sim \text{Exp}(N)$;
 end
 if $\psi(t) < -c$ **then**
 $x(t+1) \leftarrow x(t) + 1$ and $y(t+1) \leftarrow y(t)$;
 Draw $u_{x(t+1)} - u_{x(t)} \sim \text{Exp}(N)$;
 end
end

the fraction of matched vertices is given by the following formula:

$$\frac{2 \sum_{t=1}^{\tau_N} \mathbb{1}_{\{|\psi(t)| \leq c\}}}{2 \sum_{t=1}^{\tau_N} \mathbb{1}_{\{|\psi(t)| \leq c\}} + \sum_{t=1}^{\tau_N} \mathbb{1}_{\{|\psi(t)| \geq c\}}}.$$

The potential function $\psi(t)$ is a Markov chain with the following transition probability:

$$\psi(t+1) - \psi(t) = \begin{cases} x_t - y_t & \text{if } |\psi(t)| \leq c \\ x_t & \text{if } \psi(t) \leq -c \\ -y_t & \text{if } \psi(t) \geq c \end{cases}$$

with x_t and y_t independent exponential random variables of parameter N . The rest of the proof consists in studying this random walk, its stationary distribution, and its convergence.

Lemma 2 *The Markov chain described $\psi(t)$ admits the following stationary distribution:*

$$\mu(t) = \begin{cases} \frac{1}{2c+2} & \text{if } |x| \leq c \\ \frac{e^{x+c}}{2c+2} & \text{if } x \leq -c \\ \frac{e^{-(x-c)}}{2c+2} & \text{if } x \geq c. \end{cases}$$

Proof: Let us denote by Π the transition kernel of random walk ψ ; then for any $x \in [-c, c]$:

$$\begin{aligned} (2c+2) \int_{-\infty}^{+\infty} \Pi(x, y) \mu(y) dy &= \int_{-\infty}^{-c} e^{-(x-y)} e^{y+c} dy + \int_c^{+\infty} e^{-(y-x)} e^{-y+c} dy \\ &\quad + \int_{-c}^x \frac{1}{2} e^{-(x-y)} dy + \int_x^c \frac{1}{2} e^{x-y} dy \\ &= 1. \end{aligned}$$

Similarly, for any $x \leq -c$:

$$\begin{aligned} (2c+2) \int_{-\infty}^{+\infty} \Pi(x, y) \mu(y) dy &= \int_{-\infty}^x e^{-(x-y)} e^{y+c} dy + \int_c^{+\infty} e^{-(y-x)} e^{-y+c} dy \\ &\quad + \int_{-c}^c \frac{1}{2} e^{x-y} dy \end{aligned}$$

$$=e^{x+c}.$$

By symmetry, the computation also holds $\forall x \geq c$. $\square$

By Chernoff bound, with probability at least $1 - e^{-\frac{N}{4}}$, $\tau_N \geq \frac{N}{2}$. Thus, by the pointwise ergodic theorem, the following convergence holds:

$$\gamma^*(c, N) = \mathbb{E} \left[\frac{2 \sum_{t=1}^{\tau_N} \mathbb{1}_{\{|\psi(t)| \leq c\}}}{2 \sum_{t=1}^{\tau_N} \mathbb{1}_{\{|\psi(t)| \leq c\}} + \sum_{t=1}^{\tau_N} \mathbb{1}_{\{|\psi(t)| > c\}}} \right] \xrightarrow{N \rightarrow +\infty} \frac{2\mu(|x| \leq c)}{2\mu(|x| \leq c) + \mu(|x| > c)} = \frac{c}{c + \frac{1}{2}}.$$

We now derive the high probability bound. Define $p_N := \frac{1}{\tau_N} \sum_{t=1}^{\tau_N} \mathbb{1}_{\{|\psi(t)| \leq c\}}$. Note that we have that the total number of vertices generated is equal to $2p_N\tau_N + (1 - p_N)\tau_N = (p_N + 1)\tau_N$.

For any $y \in \mathbb{R}$, $\int_{-\infty}^{+\infty} \Pi(x, y) \mu(y) dy \leq \frac{1}{2}$ and for any $x \in \mathbb{R}$, $\int_{-\infty}^{+\infty} \Pi(x, y) \mu(x) dx \leq \frac{3}{4}$. Thus by Schur test's lemma, the operator norm of the kernel is bounded as $\|\Pi\|_\mu \leq \sqrt{3/8}$. By a version of Hoeffding's inequality adapted to Markov chains [Miasojedow, 2014], we get for any $\delta > 0$

$$\mathbb{P}(|p_N - \mu(|x| < c)| \geq \delta | \tau_N) \leq 4(c+1)e^{-\frac{\delta^2 \tau_N}{5}}$$

Setting $\delta = \sqrt{10 \frac{\ln(N)}{N}}$, we get:

$$\mathbb{P} \left(|p_N - \mu(|x| < c)| \geq \sqrt{10 \frac{\ln(N)}{N}} \mid \tau_N \geq \frac{N}{2} \right) \leq \frac{4(c+1)}{N}.$$

Moreover, by Chernoff bound, with probability at least $1 - e^{-\frac{N}{4}}$,

$$\tau_N \geq \frac{N}{2}.$$

Thus, the following holds:

$$\mathbb{P} \left(|p_N - \mu(|x| < c)| \geq \sqrt{10 \frac{\ln(N)}{N}} \right) \leq \frac{4(c+1)}{N} + e^{-\frac{N}{4}}.$$

Note $\frac{m^*(\mathcal{U}, \mathcal{V}, c, N)}{N}$ the fraction of matched vertices by the algorithm SMALL-FIRST-GENERATIVE. We have:

$$\begin{aligned} \frac{m^*(\mathcal{U}, \mathcal{V}, c, N)}{N} &= \frac{2p_N}{p_N + 1} \\ &= \frac{2\mu(|x| < c) + 2(p_N - \mu(|x| < c))}{\mu(|x| < c) + 1 + (p_N - \mu(|x| < c))}. \end{aligned}$$

Since $\frac{2\mu(|x| < c)}{\mu(|x| < c) + 1} = \frac{c}{c + \frac{1}{2}}$, this implies with probability at least $1 - O(\frac{1}{N})$:

$$\left| \frac{m^*(\mathcal{U}^N, \mathcal{V}^N, c, N)}{N} - \frac{c}{c + \frac{1}{2}} \right| \leq 4|p_N - \mu(|x| < c)| \leq 4\sqrt{\frac{10 \ln(N)}{N}}.$$

Combining this with Equation (1) implies that with probability at least $1 - O(\frac{1}{N})$,

$$\frac{m^*(\mathcal{X}, \mathcal{Y}^N, c, N)}{N} = \frac{c}{c + \frac{1}{2}} + O \left(\sqrt{\frac{\ln(N)}{N}} \right).$$

$\square$

3 Match to the closest point algorithm

In this section, we study the performances on the 1D Geometric Graph of the online matching algorithm CLOSEST, which matches the incoming vertex to its closest available neighbor if there is one. The following theorem states that $\kappa(c, N)$, the size of the matching obtained by CLOSEST algorithm on an instance of $\text{Geom}(c, N)$, is closely related to the solution of an explicit PDE.

Theorem 2 *Let $\kappa[c, N](t)$ be the size of the matching obtained by CLOSEST on an instance of $\text{Geom}(c, N)$ when tN vertices have arrived, then*

$$\frac{\kappa[c, N](t)}{N} \xrightarrow[N \rightarrow +\infty]{\mathbb{P}} 1 - \int_0^{+\infty} f(x, t) dx$$

with $f(x, t)$ the solution of the following differential equation

$$\begin{aligned} \frac{\partial f(x, t)}{\partial t} = & -\min(x, 2c)f(x, t) - \frac{1}{\int_0^{+\infty} f(x', t) dx'} \int_0^{+\infty} \min(x', 2c)f(x', t) dx' f(x, t) \\ & + \frac{1}{\int_0^{+\infty} f(x', t) dx'} \int_0^x \min(x', 2c)f(x', t) f(x - x', t) dx' \end{aligned} \quad (2)$$

with the following initial conditions

$$f(x, 0) = e^{-x}.$$

Remark: Function f depends on c , we omit that dependency in the notations.

Turning to metric matching, the following theorem gives the asymptotic cost of algorithm CLOSEST in the ϵ -excess regime.

Theorem 3 *Let $\rho[N](t)$ be the length of the matching obtained by CLOSEST on an instance of $\text{Geom}(N)$ when $tN \leq N(1 - \epsilon)$ vertices have arrived, then*

$$\frac{\rho[N](t)}{N} \xrightarrow[N \rightarrow +\infty]{\mathbb{P}} \frac{1}{2} \left[\frac{1}{1 - t} - 1 \right].$$

Proof structure: The proof of the two theorems is very similar and we prove them together, providing additional details for each subcase where it is needed. We first show that the score of CLOSEST in $\text{Geom}(\mathcal{X}, \mathcal{Y}, c, N)$ is closely related to its score in $G_{\text{glued}}(\tilde{\mathcal{U}}[k], \mathcal{Y}, c, N)$, where the vertices of the offline side are generated through a Poisson point process and have their coordinates rounded to a discrete grid (Section 3.1), with k the discretization parameter. We then show that the score of CLOSEST on the modified graph is closely related to the solution of a PDE through the differential equation method [Enriquez et al., 2019](Section 3.2). $\square$

3.1 Graph rounding

Let $\mathcal{X}$ be an ensemble of N points drawn i.i.d. uniformly in $[0, 1]$, and $\mathcal{Y}$ ($\mathcal{Y}_{1-\epsilon}$) be an independent ensemble of N ($N(1 - \epsilon)$) points also drawn i.i.d. uniformly in $[0, 1]$, with the vertices in the ensemble indexed according to the order in which they are drawn. We define $\kappa(\text{Geom}(\mathcal{X}, \mathcal{Y}, c, N))$ as the number of matched vertices by CLOSEST in the graph $\text{Geom}(\mathcal{X}, \mathcal{Y}, c, N)$ when the vertices in $\mathcal{Y}$ arrived in the order prescribed by their indexes. Similarly, define $\rho(\text{Geom}(\mathcal{X}, \mathcal{Y}_{1-\epsilon}))$ as the length of the matching produced by CLOSEST in graph $\text{Geom}(\mathcal{X}, \mathcal{Y}_{1-\epsilon})$ when the vertices in $\mathcal{Y}_{1-\epsilon}$ arrived in the order prescribed by their indexes.

The GRAPH-ROUNDING procedure, illustrated in Figure 4, associates to an ensemble $\mathcal{X}$ the rounded ensemble $\tilde{\mathcal{U}}[k]$ through the following steps:

- **Poissonization step:** Let $N_0 \sim \text{Poi}(N)$. If $N_0 > N$, then let $(u_i)_{i=1}^{N_0-N}$ be $N_0 - N$ points drawn uniformly and independently in $[0, 1]$ and define $\mathcal{U} = \mathcal{X} \cup \{u_i \mid i \in [N_0 - N]\}$. If $N_0 < N$, $\mathcal{U}$ is obtained by removing $N - N_0$ points selected uniformly at random from $\mathcal{X}$. This is the same procedure as the one used in Section 2.

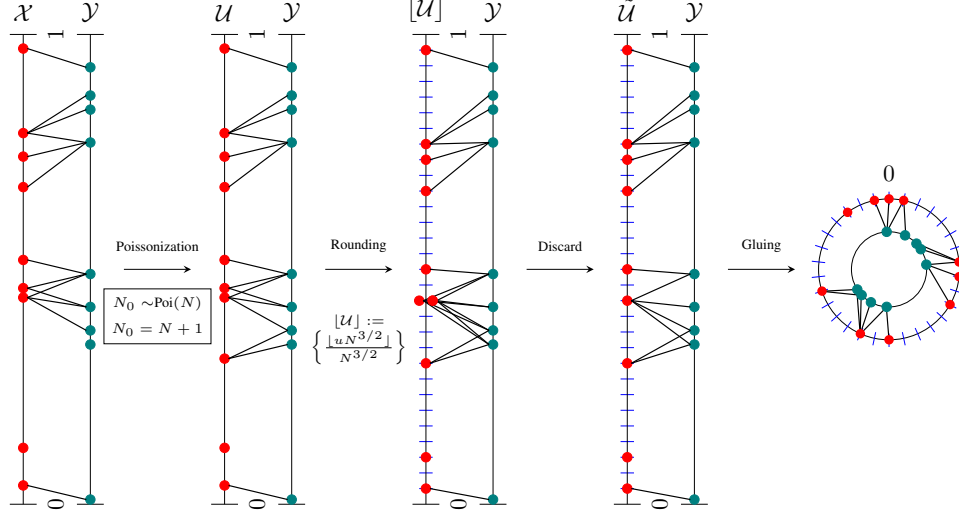


Figure 4: Graph Rounding

- **Rounding step:** Transform $\mathcal{U}$ in a new ensemble $\mathcal{U}[k]$ by rounding the coordinate of each point $u \in \mathcal{U}$ to $\frac{\lfloor uNk \rfloor}{Nk}$,

$$\mathcal{U}[k] := \left\{ \frac{\lfloor uNk \rfloor}{Nk} \mid u \in \mathcal{U} \right\}.$$
- **Discarding step:** For any $\ell \in [Nk]$, if multiple vertices have their coordinates rounded to $\frac{\ell}{Nk}$, a random vertex among those is selected, and all others are removed from the graph. This gives the final ensemble $\tilde{\mathcal{U}}[k]$.
- **Gluing step:** The interval $[0, 1]$ is mapped to the unit circle of circumference one. Formally, the distance is replaced with the following one: $d(x, y) = \min(|x - y|, |x + 1 - y|)$. We also add a vertex at coordinate 0 to $\tilde{\mathcal{U}}[k]$ if it is not already in it.

In the case of maximum cardinality matching, we denote the obtained graph $G_{glued}(\tilde{\mathcal{U}}[k], \mathcal{Y}, c, N)$, illustrated in Figure 4. In the case of metric matching we denote it $G_{glued}(\tilde{\mathcal{U}}[k], \mathcal{Y}_{1-\epsilon})$.

Note that after the poissonization step, the law of $\mathcal{U}$ is that of an ensemble of points drawn from a Poisson point process of intensity N in $[0, 1]$. Let $u_\ell = \ell/(Nk)$. For each $\ell \in [Nk]$,

$$\mathbb{P}(u_\ell \in \tilde{\mathcal{U}}[k]) = 1 - e^{-1/k} := p_k,$$

and the events $\left(\{u_\ell \in \tilde{\mathcal{U}}[k]\} \right)_{\ell \in [Nk]}$ are independent of each other.

The following proposition states that this construction does not impact too much the size of the matchings.

Proposition 2 *With probability at least $1 - \frac{4}{N} - \exp(-\sqrt{N}) - \exp(-\frac{2N}{k^2})$, we have:*

$$\left| \kappa(\text{Geom}(\mathcal{X}, \mathcal{Y}, c, N)) - \kappa(G_{glued}(\tilde{\mathcal{U}}[k], \mathcal{Y}, c, N)) \right| \leq 21 \frac{N}{k} + 8\sqrt{N \ln(N)}.$$

Let $\alpha = 3 \max(\frac{12}{\epsilon^2(1-\epsilon)}; \frac{32}{(\epsilon-\epsilon^2)^2})$. For any $N \geq 4$, $k \geq \frac{4}{\epsilon-\epsilon^2}$ it holds that:

$$\left| \rho(\text{Geom}(\mathcal{X}, \mathcal{Y}_{1-\epsilon})) - \rho(G_{glued}(\tilde{\mathcal{U}}[k], \mathcal{Y}_{1-\epsilon})) \right| \leq 84 \frac{(\alpha \ln(N))^2}{k} + 24\sqrt{\frac{\ln(N)}{N}} + \frac{\alpha^5 \ln(N)^3}{18N}.$$

with probability at least $1 - \frac{8\alpha+2}{N^2} - \frac{4}{N} - \exp(-\frac{2N}{k^2})$.

The proof of this technical proposition is postponed to Section 4.1.

3.2 Analysis of CLOSEST on the modified graph

This section is dedicated to the analysis of the score of CLOSEST on the rounded graphs, $G_{\text{glued}}(\tilde{\mathcal{U}}[k], \mathcal{Y}, c, N)$ in the case of maximum cardinality matching, and $G_{\text{glued}}(\tilde{\mathcal{U}}[k], \mathcal{Y})$ in the case of metric matching. This is done by tracking the number of gaps of a certain length between successive free vertices at every iteration.

Let N_t be the number of free vertices at iteration t . At time t , let $u_t(i)$ be the coordinate of the i^{th} free vertex, with the vertices enumerated according to their coordinates, and the convention $u_t(N_t + 1) = u_t(1)$. For $\ell \in [kN]$, define

$$F_{k,N}(\ell, t) := \left| \left\{ i \in [N_t] \text{ s.t. } N(u_t(i+1) - u_t(i)) = \frac{\ell}{k} \right\} \right|,$$

which is the number of gaps of size $\frac{\ell}{Nk}$ between two successive free vertices. Note that at iteration t , the number of free vertices is equal to $\sum_{\ell \in [kN]} F_N(\ell, t)$. Also define

$$M_{k,N}(\ell_-, \ell_+, t) := \left| \left\{ i \in [N_t] \text{ s.t. } (u_t(i) - u_t(i-1)) = \frac{\ell_-}{Nk} \text{ and } (u_t(i+1) - u_t(i)) = \frac{\ell_+}{Nk} \right\} \right|,$$

which is the number of times a gap of size $\frac{\ell_-}{kN}$ is followed by a gap of size $\frac{\ell_+}{kN}$ when enumerating the free vertices according to their coordinates.

Let $\mathcal{F}_t$ be the filtration associated with the values of the sizes of the gaps up to time t , $(F_{k,N}(\ell, t'))_{\ell, t' \leq t}$. The following lemma describes how many times two gaps of a certain size follow each other at iteration t , in expectation conditionally on $\mathcal{F}_t$.

Lemma 3 (Gaps repartition) *For all $t \in [N]$, for all $\ell_-, \ell_+ \in [kN]^2$,*

$$\mathbb{E} \left[M_{k,N}(\ell_-, \ell_+, t) \middle| \mathcal{F}_t \right] = F_{k,N}(\ell_-, t) \left(\mathbb{1}_{\{\ell_- \neq \ell_+\}} \frac{F_{k,N}(\ell_+, t)}{N_t - 1} + \mathbb{1}_{\{\ell_- = \ell_+\}} \frac{F_{k,N}(\ell_+, t) - 1}{N_t - 1} \right).$$

Sketch of Proof: This lemma is implied by a stronger result: conditionally on $\mathcal{F}_t$, the gaps are ordered uniformly random. This is proved by induction in Section 4.2. In the case of metric matching, this result can be found in Frieze et al. [1990], the proof here focuses on the case of maximum cardinality matching. $\square$

Note that the expression on the right-hand side is exactly the expectation obtained by drawing uniformly without replacements two successive gaps among all gaps at time t .

This lemma entails explicit computation of the expected evolution of the gaps, still conditionally on $\mathcal{F}_t$.

Lemma 4 (Evolution law) *For all $t \in [N]$, for all $\ell \in [kN]$,*

$$\underbrace{\left| \sum_{\ell'=0}^{\ell} \frac{\min(2c, \frac{\ell'}{k})}{N} \frac{F_{k,N}(\ell', t) F_{k,N}(\ell - \ell', t)}{N_t} - \left(\frac{\min(2c, \frac{\ell}{k})}{N} + \sum_{\ell'} \frac{\min(2c, \frac{\ell'}{k})}{N} \frac{F_{k,N}(\ell', t)}{N_t} \right) F_{k,N}(\ell, t) \right|}_{A_{\ell,t}} - \mathbb{E} \left[F_{k,N}(\ell, t+1) - F_{k,N}(\ell, t) \middle| \mathcal{F}_t, N_t \geq w_1 N \right] \Big| \leq \frac{6}{w_1 N}.$$

In the case of metric matching, the expression simplifies to:

$$\left| \sum_{\ell'=0}^{\ell} \frac{\ell'}{Nk} \frac{F_{k,N}(\ell', t) F_{k,N}(\ell - \ell', t)}{(1-t)N} - \left(\frac{\ell}{kN} + \frac{1}{(1-t)N} \right) F_{k,N}(\ell, t) \right| - \mathbb{E} \left[F_{k,N}(\ell, t+1) - F_{k,N}(\ell, t) \middle| \mathcal{F}_t \right] \Big| \leq \frac{6}{N\epsilon}.$$

Remark: The following proof holds with $c = +\infty$ in the case of metric matching.

Proof: If the i^{th} free vertex at iteration t is s.t. $u_t(i) - u_t(i-1) = \frac{\ell_-}{k}$ and $u_t(i+1) - u_t(i) = \frac{\ell_+}{k}$, we use the shorthand

$$u_t(i) \in M_{k,N}(\ell_-, \ell_+, t).$$

At iteration t , if no vertex is matched, $F_{k,N}(\ell, t+1) = F_{k,N}(\ell, t)$, for all $\ell \in [Nk]$. Note i_t the index of the matched vertex if there is one. If $u_t(i_t) \in M_{k,N}(\ell_-, \ell_+, t)$, then

- $F_{k,N}(\ell_-, t+1) = F_{k,N}(\ell_-, t) - 1$ and $F_{k,N}(\ell_+, t+1) = F_{k,N}(\ell_+, t) - 1$,
- $F_{k,N}(\ell_+ + \ell_-, t+1) = F_{k,N}(\ell_+ + \ell_-, t) + 1$.

Moreover,

$$\mathbb{P}(v_t \text{ matched to } u_t(i) | u_t(i) \in M_{k,N}(\ell_-, \ell_+, t), \mathcal{F}_t) = \underbrace{\frac{\min(c, \frac{\ell_-}{2k}) + \min(c, \frac{\ell_+}{2k})}{N}}_{m(\ell_-, \ell_+)}.$$

Then,

$$\begin{aligned} \mathbb{P}(v_t \text{ matched in } M_{k,N}(\ell_-, \ell_+, t) | \mathcal{F}_t) &= \sum_{i \in [N_t]} \mathbb{E}[\mathbb{1}\{v_t \text{ matched to } u_t(i), u_t(i) \in M_{k,N}(\ell_-, \ell_+, t)\} | \mathcal{F}_t] \\ &= m(\ell_-, \ell_+) \mathbb{E}[\mathbb{1}\{u_t(i) \in M_{k,N}(\ell_-, \ell_+, t)\} | \mathcal{F}_t] \\ &= m(\ell_-, \ell_+) \mathbb{E}[M_{k,N}(\ell_-, \ell_+, t) | \mathcal{F}_t]. \end{aligned}$$

This implies the following chain of equality:

$$\begin{aligned} \mathbb{E}\left[F_{k,N}(\ell, t+1) - F_{k,N}(\ell, t) \middle| \mathcal{F}_t\right] &= -\left(\sum_{\ell'} m(\ell, \ell') \mathbb{E}[M_{k,N}(\ell, \ell', t) | \mathcal{F}_t] + m(\ell', \ell) \mathbb{E}[M_{k,N}(\ell', \ell, t) | \mathcal{F}_t]\right) \\ &\quad + \sum_{\ell' \leq \ell} m(\ell', \ell - \ell') \mathbb{E}[M_{k,N}(\ell', \ell - \ell', t) | \mathcal{F}_t] \\ &= -2 \sum_{\ell'} m(\ell, \ell') \mathbb{E}[M_{k,N}(\ell, \ell', t) | \mathcal{F}_t] + 2 \sum_{\ell' \leq \ell} \frac{\min(c, \frac{\ell'}{2k})}{N} \mathbb{E}[M_{k,N}(\ell', \ell - \ell', t) | \mathcal{F}_t], \end{aligned}$$

where the second line holds since by Lemma 3 $\mathbb{E}[M_{k,N}(\ell, \ell', t) | \mathcal{F}_t] = \mathbb{E}[M_{k,N}(\ell', \ell, t) | \mathcal{F}_t]$. We also have $\sum_{\ell'} M_{k,N}(\ell, \ell', t) = F_{k,N}(\ell, t)$, so the previous equation simplifies to:

$$\begin{aligned} \mathbb{E}\left[F_{k,N}(\ell, t+1) - F_{k,N}(\ell, t) \middle| \mathcal{F}_t\right] &= \sum_{\ell'=0}^{\ell} \frac{\min(2c, \ell'/k)}{N} \frac{F_{k,N}(\ell', t)(F_{k,N}(\ell - \ell', t) - \mathbb{1}_{\{\ell = \ell'\}})}{N_t - 1} \\ &\quad - \frac{\min(2c, \ell/k)}{N} F_{k,N}(\ell, t) - F_{k,N}(\ell, t) \sum_{\ell'} \frac{\min(2c, \ell'/k)}{N} \frac{F_{k,N}(\ell', t) - \mathbb{1}_{\ell = \ell'}}{N_t - 1}. \end{aligned}$$

We thus have:

$$\begin{aligned} \left| \sum_{\ell'=0}^{\ell} \frac{\min(2c, \ell'/k)}{N} \frac{F_{k,N}(\ell', t) F_{k,N}(\ell - \ell', t)}{N_t} - \left(\frac{\min(2c, \ell/k)}{N} + \sum_{\ell'} \frac{\min(2c, \ell'/k)}{N} \frac{F_{k,N}(\ell', t)}{N_t} \right) F_{k,N}(\ell, t) \right| \\ - \mathbb{E}\left[F_{k,N}(\ell, t+1) - F_{k,N}(\ell, t) \middle| \mathcal{F}_t\right] \leq \underbrace{\frac{\min(2c, \ell/2k) F_{k,N}(\ell/2, t)}{N N_t} + \frac{\min(2c, \ell/k) F_{k,N}(\ell, t)}{N N_t}}_{\leq \frac{2}{N_t}} \\ + \sum_{\ell'=0}^{\ell} \frac{\min(2c, \ell'/k)}{N} \frac{F_{k,N}(\ell', t)}{N_t} \underbrace{\frac{F_{k,N}(\ell - \ell', t)}{N_t - 1}}_{\leq 2} \end{aligned}$$

$$\begin{aligned}
& + \underbrace{\frac{F_{k,N}(\ell, t)}{N_t - 1}}_{\leq 2} \sum_{\ell'} \frac{\min(2c, \ell'/k)}{N} \frac{F_{k,N}(\ell', t)}{N_t}, \\
& \leq \frac{6}{N_t}.
\end{aligned}$$

To obtain the previous inequalities, we use that by construction, $\sum_{\ell} \frac{\ell}{kN} F_{k,N}(\ell, t) = 1$ always holds, which also implies $\frac{\ell}{kN} F_{k,N}(\ell, t) \leq 1$ for any ℓ . To simplify for the metric matching case, we use again that equality along with the identity $N_t = (1 - t)N$. $\square$

The following lemma gives a concentration bound on the initial values of the gaps.

Lemma 5 (Initial Conditions) *For any $\ell < kN$, with probability at least $1 - \frac{4}{N^4}$,*

$$\left| F_{N,k}(\ell, 0) - Nkp_k^2 e^{-\frac{\ell}{k}} \right| \leq 3\sqrt{12N \ln(N)}.$$

Proof: Let $\mathcal{G}(p_k)$ be the geometric distribution with parameter p_k . Consider the process of placing a sequence of vertices $(v_i)_{i=1}^{+\infty}$ in $\mathbb{R}$, placing v_0 at zero, and having

$$Nk(v_{i+1} - v_i) \stackrel{\text{i.i.d.}}{\sim} \mathcal{G}(p_k).$$

With that procedure, the law of the vertices placed before 1 is exactly the law of the vertices in $\tilde{\mathcal{U}}[k]$. We have

$$\mathbb{E} \left[\mathbb{1}_{\{v_{i+1} - v_i = \frac{\ell}{Nk}\}} \right] = p_k e^{-\frac{\ell-1}{k}}.$$

Thus, by concentration of sums of Bernoulli random variable and since $k p_k \leq 1$:

$$\begin{aligned}
& \mathbb{P} \left(\left| \sum_{i=1}^{Nkp_k} \mathbb{1}_{\{v_{i+1} - v_i = \frac{\ell}{Nk}\}} - Nk e^{-\frac{\ell-1}{k}} p_k^2 \right| > \sqrt{12N \ln(N)} \right) \\
& \leq \mathbb{P} \left(\left| \sum_{i=1}^{Nkp_k} \mathbb{1}_{\{v_{i+1} - v_i = \frac{\ell}{Nk}\}} - Nk e^{-\frac{\ell-1}{k}} p_k^2 \right| > \sqrt{12k e^{-\frac{\ell-1}{k}} p_k^2 N \ln(N)} \right), \\
& \leq \frac{2}{N^4}.
\end{aligned} \tag{3}$$

We have:

$$\begin{aligned}
\mathbb{P} \left(v_{Nkp_k + \sqrt{12N \ln(N)}} < 1 \right) &= \mathbb{P} \left(|\tilde{U}[k]| \geq Nkp_k + \sqrt{12N \ln(N)} \right) \\
&\leq \mathbb{P} \left(|\tilde{U}[k]| \geq Nkp_k \left(1 + \sqrt{\frac{12N \ln(N)}{Nkp_k}} \right) \right) \\
&\leq \frac{1}{N^4}.
\end{aligned}$$

Where the last inequality holds by Chernoff bound for sums of Bernoulli random variables. Similarly, $\mathbb{P}(v_{Nkp_k - \sqrt{12N \ln(N)}} > 1) \leq \frac{1}{N^4}$.

Thus, with probability at least $1 - \frac{2}{N^4}$, there are between $Nkp_k - \sqrt{12N \ln(N)}$ and $Nkp_k + \sqrt{12N \ln(N)}$ vertices placed before 1. The count of the gaps for $\tilde{\mathcal{U}}[k]$ is thus within $2\sqrt{12N \ln(N)} + 1$ of the counts of the gaps up to v_{Nkp_k} . Combining with Equation 3 gives the Lemma. $\square$

3.3 Application of the differential equation method

The methods employed to relate the discrete process to the differential equation, albeit similar in spirit, now differ for maximal cardinality matching and metric matching. The proofs are separated into two different subsections.

3.3.1 Maximum cardinality matching

The following proposition is obtained by applying the differential equation method (DEM, see Appendix B) on the discretized graphs.

Proposition 3 (DEM, maximum cardinality matching) *For any $t \in [0, 1]$,*

$$\frac{1}{N} \sum_{\ell=0}^{kN} F_{k,N} \left(\ell, \frac{\lfloor tN \rfloor}{N} \right) \xrightarrow[N \rightarrow +\infty]{\mathbb{P}} \sum_{\ell=0}^{+\infty} f_k(\ell, t).$$

where the functions $(f_k(\ell, t))_\ell$ are the solutions of the following system of differential equations:

$$\begin{aligned} \frac{\partial f_k(\ell, t)}{\partial t} = & -\min\left(\frac{\ell}{k}, 2c\right) f_k(\ell, t) - \frac{1}{\sum_{\ell'=0}^{+\infty} f_k(\ell', t)} \left[\sum_{\ell'=0}^{+\infty} \min\left(\frac{\ell'}{k}, 2c\right) f_k(\ell', t) \right] f_k(\ell, t) \\ & + \frac{1}{\sum_{\ell'=0}^{+\infty} f_k(\ell', t)} \sum_{\ell'=0}^{+\infty} \min\left(\frac{\ell'}{k}, 2c\right) f_k(\ell', t) f_k(\ell - \ell', t), \end{aligned}$$

with the initial conditions:

$$f_k(\ell, 0) = kp_k^2 e^{-\frac{\ell}{k}}.$$

Remark: The functions f_k depends on c , we omit that dependency in the notations.

Proof: We first look at the initial condition. Define the event

$$\mathcal{A}_1 := \left\{ \sum_{\ell < kN^{1/8}} \left| F_{N,k}(\ell, 0) - Nke^{-\frac{\ell}{k}} \left(1 - e^{-\frac{1}{k}}\right)^2 \right| \leq 3kN^{-3/8} \sqrt{12 \ln(N)} \times N \right\}.$$

By union bound on the result of Lemma 5, it holds with probability at least $1 - \frac{4kN^{1/8}}{N^4}$.

We now work on the trend condition. Denote

$$r_{N,k}(t) = \sum_{\ell=kN^{1/8}}^{+\infty} f_k(\ell, t) \text{ and } \tilde{r}_{N,k}(t) = \sum_{\ell=kN^{1/8}}^{+\infty} \min\left(\frac{\ell}{k}, 2c\right) f_k(\ell, t).$$

Define the function $\Phi_{k,N} : \mathbb{R}^{kN^{1/8}+1} \rightarrow \mathbb{R}^{kN^{1/8}}$ as:

$$\begin{aligned} \Phi_{k,N}^\ell \left(t, (y_\ell)_{\ell=1}^{kN^{1/8}} \right) = & - \frac{\sum_{\ell'=1}^{kN^{1/8}} \min(\frac{\ell'}{k}, 2c) y_{\ell'} + \tilde{r}_{N,k}(t)}{\sum_{\ell'=1}^{kN^{1/8}} y_{\ell'} + r_{N,k}(t)} y_\ell - \min\left(\frac{\ell}{k}, 2c\right) y_\ell \\ & + \frac{\sum_{\ell'=0}^{\ell} \min(\frac{\ell'}{k}, 2c) y_{\ell'} y_{\ell-\ell'}}{\sum_{\ell'=1}^{kN^{1/8}} y_{\ell'} + r_{N,k}(t)}, \end{aligned}$$

and consider the domain:

$$\mathcal{D} := \left\{ (a_\ell)_{\ell=1}^{kN^{1/8}} \text{ s.t. } \frac{1}{2} e^{-4c} \leq \sum_{\ell} a_\ell \leq 2 \right\}.$$

Lemma 6 (Trend, maximum cardinality matching) *It holds that for any $t \leq N$:*

$$\sum_{\ell=1}^{kN^{1/8}} \left| \mathbb{E} \left[F_{k,N}(\ell, t+1) - F_{k,N}(\ell, t) \middle| \mathcal{F}_t, \left(\frac{F_{k,N}(\ell, t)}{N} \right)_{\ell=1}^{kN^{1/8}} \in \mathcal{D} \right] - \Phi_{k,N}^\ell \left(t, \left(\frac{F_{k,N}(1, t)}{N} \right)_{\ell=1}^{kN^{1/8}} \right) \right| \leq w_5 N^{-1/8},$$

with w_5 a constant independent of N . Moreover, for any sequence $(a_\ell)_{\ell=1}^{kN^{1/8}} \in \mathcal{D}$, $\forall \ell' \leq kN^{-1/8}$, $\forall t \in [0, 1]$,

$$\Phi_{k,N}^{\ell'} \left(t, (a_\ell)_{\ell=1}^{kN^{1/8}} \right) \leq 16c.$$

This proof is purely computational and delayed to section 4.4.

We now prove $\Phi_{k,N}$ is Lipschitz on $\mathcal{D}$.

Lemma 7 (Lipschitz, maximum cardinality matching) *On domain $\mathcal{D}$, the $\Phi_{k,N}$ is L_c -lipschitz with respect to the L_1 norm, with L_c a constant independent of N .*

This proof is computational and delayed to section 4.4.

The last step missing is to prove that $(f_k(\ell, t))_{\ell=1}^{kN^{1/8}}$ remains far enough from the boundaries of $\mathcal{D}$ for $t \in [0, 1]$.

Note that we have:

$$0 \geq \frac{\partial \sum_\ell f_k(\ell, t)}{\partial t} \geq -4c \sum_\ell f_k(\ell, t).$$

As $\sum_\ell f_k(0, t) = 1$, this implies that for any $t \leq 1$, we have $e^{-4c} \leq \sum_\ell f_k(\ell, t) \leq 1$. We prove in 4.4 (Equation 14) that $r_{N,k}(t) \leq N^{-1/8}$ for any $t \in [0, 1]$. Thus, for any N s.t. $N^{-1/8} \leq \frac{1}{4}e^{-4c}$, for any $t \in [0, 1]$, $(f_k(\ell, t))_{\ell=1}^{kN^{1/8}}$ remains within $\frac{1}{4}e^{-4c}$ of the boundaries of $\mathcal{D}$.

We are now ready to apply, the Differential Equation Method, Theorem 4.

Fix $a(N) = kN^{1/8}$. The trend hypothesis holds by Lemma 6 with $\delta(N) = w_5 N^{-1/8}$. The boundedness hypothesis holds with $\beta = 3$ as at most a vertex is matched by iteration. Event $\mathcal{A}_1$ holds with probability at least $1 - \frac{4kN^{1/8}}{N^4}$, under event $\mathcal{A}_1$, the initial condition hypothesis holds with $\lambda(N) = \max \left(3kN^{-1/4} \sqrt{12 \ln(N)}, (w_5 + 16cL_c kN^{-3/4}) \min(16c, L_c^{-1}) \right) N^{-1/8}$, and by Lemma 7, $\Phi_{k,N}$ is L_c -Lipschitz with respect to the L_1 norm on $\mathcal{D}$. For any N large enough s.t. $N^{-1/8} \leq \frac{1}{4}e^{-4c}$ and $3e^L \lambda(N) \leq \frac{1}{4}e^{-4c}$, $(f_{k,N}(1, \ell))_{\ell=1}^{kN^{1/8}}$ has L_1 -distance at least $3e^L \lambda(N)$ from the boundary of $\mathcal{D}$ for any $t \in [0, 1]$.

Theorem 4 implies that, for any large enough N s.t. $N^{-1/8} \leq \frac{1}{4}e^{-4c}$ and $3e^L \lambda(N) \leq \frac{1}{4}e^{-4c}$, with probability at least $1 - 2Ne^2 e^{-N\lambda(N)^2/81} - \frac{3kN^{1/8}}{N^4}$ we have

$$\underbrace{\max_{0 \leq t \leq 1} \sum_{1 \leq \ell \leq kN^{1/8}} \left| \frac{1}{N} F_{k,N}(\ell, tN) - f_k(\ell, t) \right|}_{\xrightarrow{N \rightarrow +\infty} 1} < \underbrace{3e^L \lambda(N)}_{\xrightarrow{N \rightarrow +\infty} 0}.$$

The proposition follows. $\square$

The last step of the proof is to link the functions $f_k(\ell, t)$ to a single function f , that does not depend on k .

Lemma 8 *For any $t \in [0, 1]$, it holds:*

$$\| f(x, t) - kf_k(\lfloor kx \rfloor, t) \|_1 \leq \frac{\omega}{k}.$$

with ω a constant depending only on c and $f(x, t)$ the solution of the following PDE

$$\begin{aligned} \frac{\partial f(x, t)}{\partial t} = & -\min(x, 2c)f(x, t) - \frac{1}{\int_0^{+\infty} f(x', t)dx'} \int_0^{+\infty} \min(x', 2c)f(x', t)dx' f(x, t) \\ & + \frac{1}{\int_0^{+\infty} f(x', t)dx'} \int_0^x \min(x', 2c)f(x', t)f(x - x', t)dx'. \end{aligned} \quad (4)$$

with initial conditions $f(x, 0) = e^{-x}$.

The proof of this lemma is postponed to Section 4.4. Combining it with Propositions 3 and 2 gives the theorem.

3.3.2 Metric matching

In the case of metric matching, more work is still required before the application of the Differential Equation Method. We start by showing that the length concentrates around its expectation. We then show that the contribution of long edges is small, and finally control the contribution of short edges via the Differential Equation Method.

Define $c_t[k]$ as the length of the added edge at time t in $G_{glued}(\tilde{\mathcal{U}}[k], \mathcal{Y}_{1-\epsilon})$.

Lemma 9 *The total length concentrates around its expectation.*

$$\mathbb{P} \left(\left| \sum_{t=1}^{N(1-\epsilon)} c_t[k] - \mathbb{E} \left[\sum_{t=1}^{N(1-\epsilon)} c_t[k] \right] \right| \geq \delta + \frac{4\alpha + 1}{N} \right) \leq \frac{4\alpha + 3}{N^2},$$

$$\text{with } \delta = \left[\sqrt{\frac{2 \ln(N)(k+4)}{N}} + \frac{(4\alpha+1)(k+4)}{N^2} \right] \alpha^2 \ln(N)^2.$$

Proof: Define $a_i = \mathbb{1}\{\frac{i}{Nk} \in \tilde{U}\}$. The total length of the matching obtained by CLOSEST is a function of $(a_i)_{i=1}^{Nk}$ and $(y_i)_{i=1}^{N(1-\epsilon)}$, which we denote $h((a_i)_{i=1}^{Nk}, (y_i)_{i=1}^{N(1-\epsilon)})$.

Consider a sequences $(y_i)_{i=1}^{N(1-\epsilon)}$ and two sequences $(a_i)_{i=1}^{Nk}$ and $(a'_i)_{i=1}^{Nk}$ s.t. they differ by at most one element and $\mathcal{E}_{\mathcal{I}}$ holds with both sequences. Then, by the same reasoning as in the proof of Lemma 16,

$$|h((a_i)_\ell, (y_i)_i) - h((a'_i)_\ell, (y_i)_i)| \leq \frac{4\alpha^2 \ln(N)^2}{N}.$$

The same result holds with the roles of y and a reversed. Thus, by McDiarmid's inequality:

$$\begin{aligned} \mathbb{P} \left(\left| h((a_i)_\ell, (y_i)_i) - \mathbb{E} \left[\sum_t c_t[k] \middle| \mathcal{E}_{\mathcal{I}} \right] \right| \geq \delta \right) & \leq \mathbb{P}(\overline{\mathcal{E}_{\mathcal{I}}}) + 2 \exp \left\{ - \frac{N \left(\delta - \mathbb{P}(\overline{\mathcal{E}_{\mathcal{I}}})(k+4)\alpha^2 \ln(N)^2 \right)^2}{(k+4)\alpha^4 \ln(N)^4} \right\} \\ & \leq \frac{4\alpha + 3}{N^2}. \end{aligned}$$

On the other hand:

$$\left| \mathbb{E} \left[\sum_t c_t[k] \middle| \mathcal{E}_{\mathcal{I}} \right] - \mathbb{E} \left[\sum_t c_t[k] \right] \right| \leq N \mathbb{P}(\overline{\mathcal{E}_{\mathcal{I}}}) \leq \frac{4\alpha + 1}{N}.$$

□

The following lemma bounds the total expected contribution of long edges to the final length.

Lemma 10 [Adaptation of Lemma 12-13 from Balkanski et al. [2022]] For any $\eta \in [\frac{4\epsilon-\epsilon^2}{16N}, 1]$, $k \geq \frac{1}{\epsilon}$ and $N \geq \frac{25}{1-\epsilon}$, we have:

$$\mathbb{E} \left[\sum_{t \leq N(1-\epsilon)} c_t[k] \mathbb{1}\{c_t[k] \geq \frac{\eta}{N}\} \right] \leq C_\epsilon e^{-\eta C'_\epsilon}.$$

with C_ϵ and C'_ϵ two constants depending only on ϵ .

The point generation process on this offline side in article [3] is that of N points drawn i.i.d. in $[0, 1]$, whereas we focus on points generated on a discrete grid, and the statement we aim for differs slightly. For those reasons and for completeness, we provide a brief proof of the statement in Appendix C, which essentially follows the steps of [3].

We now apply the differential equation method to control the contribution of the short edges.

Lemma 11 For any $N \geq \frac{1}{\epsilon^2}$ have:

$$\left| \mathbb{E} \left[\sum_{t=1}^{N(1-\epsilon)} c_t[k] \mathbb{1}\{c_t[k] \leq \frac{\eta}{N}\} \right] - \frac{1}{4N} \sum_{t=0}^{N(1-\epsilon)-1} \sum_{\ell=1}^{k\eta} \left(\frac{\ell}{k} \right)^2 g_k(\ell, \frac{t}{N}) \right| \leq \frac{9k\eta^3 e^{(4\eta+\frac{1}{\epsilon})}}{4} \sqrt{\frac{12 \ln(N)}{N}} + 2\eta \left(2Ne^2 e^{-\ln(N)k^2\eta^2} + \frac{4k\eta}{N^4} \right)$$

with $g_k(\ell, t)$ the solution of the following system of differential equations.

$$\frac{\partial g_k(\ell, t)}{\partial t} = - \left(\frac{\ell}{k} + \frac{1}{1-t} \right) g_k(\ell, t) + \frac{1}{1-t} \sum_{\ell'=0}^{+\infty} \frac{\ell'}{k} g_k(\ell', t) g_k(\ell - \ell', t),$$

with the initial conditions:

$$g_k(\ell, 0) = kp_k^2 e^{-\frac{\ell}{k}}.$$

Remark: g_k is the function f_k obtained by setting $c = +\infty$ (the dependency of f_k on c was omitted in the notations).

Proof: At a given iteration, we have:

$$\mathbb{E} \left[c_t \mathbb{1}\{c_t \leq \frac{\eta}{N}\} | \mathcal{F}_t \right] = \frac{1}{4} \sum_{\ell=1}^{k\eta} \left(\frac{\ell}{kN} \right)^2 F_{k,N}(\ell, t).$$

Define the function $\Gamma_{k,N} : \mathbb{R}^{k\eta} \rightarrow \mathbb{R}^{k\eta}$ as:

$$\Gamma_{k,N}^\ell \left(t, (y_\ell)_{\ell=1}^{k\eta} \right) = - \left(\frac{\ell}{k} + \frac{1}{1-t} \right) y_\ell + \frac{\sum_{\ell'=0}^{\ell} \frac{\ell'}{k} y_{\ell'} y_{\ell-\ell'}}{1-t}.$$

By Lemma 4, we have:

$$\left| \Gamma_{k,N}^\ell \left(t, \left(\frac{F_{k,N}(1, t)}{N} \right)_{\ell=1}^{k\eta} \right) - \mathbb{E} \left[F_{k,N}(\ell, t+1) - F_{k,N}(\ell, t) | \mathcal{F}_t \right] \right| \leq \frac{6}{N\epsilon}. \quad (5)$$

By Lemma 5, it holds that with probability at least $1 - 4k\eta/N^4$,

$$\sum_{\ell=1}^{k\eta} \left| F_{N,k}(\ell, 0) - Nk e^{-\frac{\ell}{k}} p_k^2 \right| \leq 3k\eta \sqrt{12N \ln(N)} \leq \frac{12k\eta}{N^{1/4}} N. \quad (6)$$

Note $\mathcal{A}_2$ that event. Define domain $\mathcal{D}_2$ as:

$$\mathcal{D}_2 = \left\{ (y_\ell)_{\ell=1}^{k\eta}, \sum_{\ell} y_\ell \leq 2, \sum_{\ell} \frac{\ell}{k} y_\ell \leq 2 \right\}$$

By Lemmas 22 and 21, $(g_k(\ell, t))_\ell \in \mathcal{D}_2, \forall t \in [0, 1]$. Let us study the regularity of the function $\Gamma_{k,N}$ on $\mathcal{D}_2$. For any two sequences $(y'_1, \dots, y'_{k\eta})$ and $(y_1, \dots, y_{k\eta})$ in $\mathcal{D}_2$, we have:

$$\begin{aligned} \sum_{\ell} |\Gamma_{k,N}^\ell(t, y_1, \dots, y_{k\eta}) - \Gamma_{k,N}^\ell(t, y'_1, \dots, y'_{k\eta})| &\leq \left(\eta + \frac{1}{1-t} \right) \sum_{\ell} |y_\ell - y'_\ell| \\ &\quad + \frac{1}{1-t} \underbrace{\sum_{\ell} \left| \sum_{\ell'=1}^{\ell} \frac{\ell'}{k} (y_{\ell'} y_{\ell-\ell'} - y'_{\ell'} y'_{\ell-\ell'}) \right|}_{(iv)} \end{aligned}$$

We work on bounding (iv).

$$\begin{aligned} (iv) &\leq \sum_{\ell} \left| \sum_{\ell'=1}^{\ell} \frac{\ell'}{k} y_{\ell'} (y_{\ell-\ell'} - y'_{\ell-\ell'}) \right| + \left| \sum_{\ell'=1}^{\ell} \frac{\ell'}{k} y'_{\ell-\ell'} (y_{\ell'} - y'_{\ell'}) \right|, \\ &\leq \eta (\|\mathbf{y}\|_1 + \|\mathbf{y}'\|_1) \|\mathbf{y} - \mathbf{y}'\|_1 \leq 4\eta \|\mathbf{y} - \mathbf{y}'\|_1. \end{aligned}$$

$\Gamma_{k,N}$ is thus $(\underbrace{\eta + \frac{4\eta+1}{\epsilon}}_{L_\eta})$ -lipschitz with respect to the L_1 norm on $\mathcal{D}_2$. We also have:

$$\begin{aligned} \sum_{\ell=1}^{k\eta} |\Gamma_{k,N}^\ell(t, y_1, \dots, y_{k\eta})| &\leq \left| \sum_{\ell=1}^{k\eta} \left(\frac{\ell}{k} + \frac{1}{1-t} \right) y_\ell \right| + \left| \sum_{\ell=1}^{k\eta} \frac{\sum_{\ell'=0}^{\ell} \frac{\ell'}{k} y_{\ell'} y_{\ell-\ell'}}{1-t} \right| \\ &\leq 2 \left(1 + \frac{1}{1-t} \right) + \frac{4}{1-t} \leq 2 + \underbrace{\frac{6}{\epsilon}}_{R_\epsilon}. \end{aligned} \tag{7}$$

Fix $a(N) = k\eta$. The trend hypothesis holds by Equation 5 with $\delta_2(N) = \frac{6}{\epsilon N}$. The boundedness hypothesis holds with $\beta = 3$. By Equation 6, event $\mathcal{A}_2$ holds with probability at least $1 - \frac{4k\eta}{N^4}$, and under event $\mathcal{A}_2$, the initial condition hypothesis holds with $\lambda_2(N) = 3k\eta \sqrt{\frac{12 \ln(N)}{N}}$. For any $N \geq \frac{1}{\epsilon^2}$, $\delta_2(N)/L_\eta + R_\epsilon k\eta/N \leq \lambda_2(N)$. $\Gamma_{k,N}$ is L_η -Lipschitz with respect to the L_1 norm and bounded by R_ϵ on $\mathcal{D}_2$. By Theorem 4, w.p. at least $1 - 2Ne^2 e^{-\ln(N)k^2\eta^2} - 4k\eta/N^4$, for all $t \leq N(1-\epsilon)$:

$$\sum_{\ell=1}^{k\eta} \left| F_{k,N}(\ell, t) - Ng_k\left(\ell, \frac{t}{N}\right) \right| \leq 3\lambda_2(N) e^{(\eta + \frac{4\eta+1}{\epsilon})} N.$$

We denote that event $\mathcal{C}$. Under event $\mathcal{C}$,

$$\begin{aligned} \sum_{t=1}^{N(1-\epsilon)} \left| \frac{1}{4} \sum_{\ell=1}^{k\eta} \left(\frac{\ell}{kN} \right)^2 Ng_k\left(\ell, \frac{t}{N}\right) - \frac{1}{4} \sum_{\ell=1}^{k\eta} \left(\frac{\ell}{kN} \right)^2 F_{k,N}(\ell, t) \right| \\ \leq \sum_{t=1}^{N(1-\epsilon)} \frac{1}{4} \left(\frac{\eta}{N} \right)^2 \sum_{\ell=1}^{k\eta} \left| Ng_k\left(\ell, \frac{t}{N}\right) - F_{k,N}(\ell, t) \right| \\ \leq \frac{9k\eta^3 e^{(\eta + \frac{4\eta+1}{\epsilon})}}{4} \sqrt{\frac{12 \ln(N)}{N}}. \end{aligned}$$

Thus:

$$\left| \mathbb{E} \left[\sum_{t=1}^{N(1-\epsilon)} c_t \mathbb{1}_{\{c_t \leq \frac{\eta}{N}\}} \right] - \frac{1}{4N} \sum_{t=0}^{N(1-\epsilon)-1} \sum_{\ell=1}^{k\eta} \left(\frac{\ell}{k} \right)^2 g_k(\ell, \frac{t}{N}) \right| \leq \frac{9k\eta^3 e^{(\eta + \frac{4\eta+1}{\epsilon})}}{4} \sqrt{\frac{12 \ln(N)}{N}} + 2\eta \mathbb{P}(\bar{\mathcal{C}}).$$

□

Lemma 12 (Link with the continuous equation, metric matching version)

$$\left| \int_{t=0}^{1-\epsilon} \int_{x=1}^{\eta} x^2 g(x, t) dx dt - \frac{1}{N} \sum_{t=0}^{N(1-\epsilon)-1} \sum_{\ell=1}^{k\eta} \left(\frac{\ell}{k} \right)^2 g_k(\ell, \frac{t}{N}) \right| \leq \eta^2 \left[\frac{1}{N} \left(2 + \frac{6}{\epsilon} \right) + \frac{2e^{3\eta + \frac{1}{\epsilon}} + 3 + \eta}{k} \right]$$

with $g(x, t)$ the solution of the following PDE:

$$\frac{\partial g(x, t)}{\partial t} = - \left(x + \frac{1}{1-t} \right) g(x, t) + \frac{1}{1-t} \int_{x'=0}^x x' g(x', t) g(x-x', t) dx',$$

with the initial conditions:

$$g(x, 0) = e^{-x}.$$

The proof of that lemma is a computation deferred to section 4.5.

Lemma 13 We have:

$$\int_0^{1-t} \int_0^{+\infty} \frac{x^2}{4} g(x, t) dx = \frac{1}{2} \left[\frac{1}{1-t} - 1 \right].$$

The proof of that lemma is a computation deferred to section 4.5.

Now take any $\varepsilon \geq 0$. By lemma 2, for any N large enough so that $84 \frac{(\alpha \ln(N))^2}{N^{1/8}} + 24 \sqrt{\frac{\ln(N)}{N}} + \frac{\alpha^5 \ln(N)^3}{18N} \leq \frac{\varepsilon}{2}$, we have:

$$\begin{aligned} \mathbb{P} \left(\left| \sum_t c_t - 0.5 \left[\frac{1}{1-t} - 1 \right] \right| \geq \varepsilon \right) &\leq \underbrace{\mathbb{P} \left(\left| \sum_t c_t[N^{1/8}] - 0.5 \left[\frac{1}{1-t} - 1 \right] \right| \geq \frac{\varepsilon}{2} \right)}_{(i)} \\ &\quad + \underbrace{\frac{8\alpha + 2}{N^2} + \frac{4}{N} + e^{-2N^{-3/4}}}_{\xrightarrow[N \rightarrow +\infty]{} 0}. \end{aligned}$$

If N is also large enough so that $\delta + \frac{4\alpha+1}{N} \leq \frac{\varepsilon}{4}$, we have:

$$(i) \leq \underbrace{\mathbb{1} \left\{ \left| \mathbb{E} \left[\sum_t c_t[N^{1/8}] \right] - 0.5 \left[\frac{1}{1-t} - 1 \right] \right| \geq \frac{\varepsilon}{8} \right\}}_{(ii)} + \underbrace{\frac{4\alpha + 3}{N^2}}_{\xrightarrow[N \rightarrow +\infty]{} 0}.$$

Take η large enough so that, $\frac{C_\epsilon}{C'_\epsilon} e^{-\eta C'_\epsilon} + \int_{t=0}^{1-\epsilon} \int_{x=\eta}^{+\infty} x^2 f(x, t) dx dt \leq \frac{\varepsilon}{16}$. Then by Lemma 10:

$$(ii) = \underbrace{\mathbb{1} \left\{ \left| \mathbb{E} \left[\sum_t c_t[N^{1/8}] \mathbb{1}_{\left\{ c_t[N^{1/8}] \leq \frac{\eta}{N} \right\}} \right] - \int_{t=0}^{1-\epsilon} \int_{x=0}^{\eta} \frac{x^2}{4} f(x, t) dx dt \right| \geq \frac{\varepsilon}{16} \right\}}_{(iii)}.$$

Take N large enough so that: $\frac{9\eta^3 e^{(\eta + \frac{4\eta+1}{\epsilon})}}{4} \sqrt{\frac{12 \ln(N)}{N^{3/4}}} + 2\eta \left(2Ne^2 e^{-\ln(N)N^{1/4}\eta^2} + \frac{4N^{1/8}\eta}{N^4} \right) \leq \frac{\epsilon}{32}$.
Then by Lemma 11:

$$(iii) = \underbrace{\mathbb{1} \left\{ \left| \frac{1}{4} \int_{t=0}^{1-\epsilon} \sum_{\ell=1}^{\eta N^{1/8}} \left(\frac{\ell}{N^{1/8}} \right)^2 f_k(\ell, t) dt - \int_{t=0}^{1-\epsilon} \int_{x=\eta}^{+\infty} x^2 f(x, t) dx dt \right| \geq \frac{\epsilon}{32} \right\}}_{(iv)}.$$

However, by Lemma 12 if N is large enough so that $\frac{\eta^2}{4} \left[\frac{1}{N} \left(2 + \frac{6}{\epsilon} \right) + \frac{2}{N^{1/8}} e^{3\eta + \frac{1}{\epsilon}} + \frac{4}{N^{1/8}} \right] \leq \frac{\epsilon}{32}$,
(iv) = 0, and the theorem follows.

4 Technical Steps for the proofs of section 3

4.1 Proof of Proposition 2

The proposition is a consequence of four lemmas that bound the impact in matching sizes incurred in each step. We start with two technical lemmas that are used in the proofs of the four following ones.

Lemma 14 *Adding or removing a vertex to the offline side of the graph modifies the score of CLOSEST by at most one. Modifying the match of a vertex modifies the score of CLOSEST by at most two.*

Proof: Let us compare the runs of the CLOSEST algorithm in

$$\text{Geom}(\mathcal{X}, \mathcal{Y}, c, N) \text{ and } \text{Geom}(\mathcal{X} \cup x_0^+, \mathcal{Y}, c, N)$$

when the vertices in $\mathcal{Y}$ arrive in the same order. Let m_t and m_t^+ be the number of matched vertices at iteration t , and $\mathcal{X}_t$ and $\mathcal{X}_t^+$ the sets of free vertices at iteration t , in $\text{Geom}(\mathcal{X}, \mathcal{Y}, c, N)$ and $\text{Geom}(\mathcal{X} \cup x_0^+, \mathcal{Y}, c, N)$ respectively. We will show by induction that at every iteration, one of the following properties holds:

(P1): $\{m_t = m_t^+ \text{ and } \exists x_t^+ \in \mathcal{X} \cup x_0^+ \text{ s.t. } \mathcal{X}_t^+ = \mathcal{X}_t \cup x_t^+\}$, (P2): $\{m_t + 1 = m_t^+ \text{ and } \mathcal{X}_t^+ = \mathcal{X}_t\}$.

If (P2) is true at some iteration t , it remains true until the end of the run and the proof is over. If (P1) is true at iteration t , the following cases are possible:

1. if y_t has no neighbor in $\mathcal{X}_t^+$, it is unmatched in both graphs,
2. if y_t 's closest neighbor in $\mathcal{X}_t^+$ is not x_t^+ , it is matched to the same vertex in both graphs,
3. if y_t 's closest neighbor in $\mathcal{X}_t^+$ is x_t^+ and x_{t+1}^+ in $\mathcal{X}_t$, it is matched to x_t^+ in $\text{Geom}(\mathcal{X} \cup x_0^+, \mathcal{Y}, c, N)$ and to x_{t+1}^+ in $\text{Geom}(\mathcal{X}, \mathcal{Y}, c, N)$,
4. y_t 's only neighbor in $\mathcal{X}_t^+$ is x_t^+ , in which case it is matched to x_t^+ in $\text{Geom}(\mathcal{X} \cup x_0^+, \mathcal{Y}, c, N)$ and unmatched in $\text{Geom}(\mathcal{X}, \mathcal{Y}, c, N)$.

Cases 1 to 3 imply that (P1) remains true at iteration $t+1$, case 4 implies that (P2) is true at iteration $t+1$. (P1) is true at iteration 0, thus either (P1) or (P2) holds at iteration N .

For the second part of the Lemma, modifying the match of a vertex is equivalent to adding and removing a vertex. $\square$

Lemma 15 *Let $\alpha = 3 \max(\frac{12}{\epsilon^2(1-\epsilon)}; \frac{32}{(\epsilon-\epsilon^2)^2})$. For any $N \geq 4$, $k \geq \frac{4}{\epsilon-\epsilon^2}$ probability at least $1 - (4\alpha + 1)/N^2$, For any interval $I = [\frac{i}{N}; \frac{i+\ell}{N}]$ s.t. $\ell \geq \alpha \ln(N)$,*

$$|\mathcal{U} \cap I| \geq |\mathcal{Y}_{1-\epsilon} \cap I| + 16 \ln(N) + 6, |\tilde{\mathcal{U}} \cap I| \geq |\mathcal{Y}_{1-\epsilon} \cap I| + 16 \ln(N) + 6, |\mathcal{X} \cap I| \geq |\mathcal{Y}_{1-\epsilon} \cap I| + 16 \ln(N) + 6$$

$$\text{and } |\mathcal{Y}_{1-\epsilon} \cap I| \leq (1-\epsilon) \left(1 + \frac{\epsilon}{2} \right) \ell, \left| \mathcal{V}_\epsilon \cap \left[\frac{i}{N}; \frac{i+1}{N} \right] \right| \leq 4 \ln(N) + 1.$$

We note $\mathcal{E}_I$ that event.

Proof: We start by proving the two last inequalities. Consider any interval $I = [\frac{i}{N}; \frac{i+\ell}{N}]$. By Chernoff bound for sums of Bernoulli r.v.:

$$\mathbb{P}\left(|\mathcal{Y}_{1-\epsilon} \cap I| \geq (1-\epsilon) \left(1 + \frac{\epsilon}{2}\right) \ell\right) \leq e^{-\frac{\epsilon^2(1-\epsilon)}{12} \ell}.$$

The following bound holds:

$$\sum_{\ell=\alpha \ln(N)}^{+\infty} e^{-\frac{\epsilon^2(1-\epsilon)}{12} \ell} = \frac{e^{-\frac{\epsilon^2(1-\epsilon)}{12}}}{e^{-\frac{\epsilon^2(1-\epsilon)}{12}} - 1} e^{-\frac{\epsilon^2(1-\epsilon)}{12} \alpha \ln(N)} \leq 2 \frac{12}{\epsilon^2(1-\epsilon)} e^{-\frac{\epsilon^2(1-\epsilon)}{12} \alpha \ln(N)} \leq \frac{\alpha}{N^3}.$$

The first inequality holds since $\frac{e^x}{e^x - 1} \leq \frac{2}{x}$ for any $x \leq 1$, and the second one by definition of α . Now, by union bound over all possible values of i , we get that probability at least $1 - \alpha/N^2$, for any interval $I = [\frac{i}{N}; \frac{i+\ell}{N}]$ s.t. $\ell N \geq \alpha \ln(N)$,

$$|\mathcal{Y}_{1-\epsilon} \cap I| \leq (1-\epsilon) \left(1 + \frac{\epsilon}{2}\right) \ell. \quad (8)$$

For any $i \in 0, \dots, N-1$,

$$\mathbb{E} \left[\sum_{j=1}^{N(1-\epsilon)} \mathbf{1} \left\{ v_j \in \left[\frac{i}{N}; \frac{i+1}{N} \right] \right\} \right] = 1 - \epsilon.$$

Again by Chernoff bound, we have:

$$\begin{aligned} \mathbb{P} \left(\sum_{j=1}^{N(1-\epsilon)} \mathbf{1} \left\{ v_j \in \left[\frac{i}{N}; \frac{i+1}{N} \right] \right\} \geq \left(1 + \frac{4 \ln(N)}{1-\epsilon} \right) (1-\epsilon) \right) &\leq e^{-\frac{16 \ln(N)^2}{2(1-\epsilon) + 4 \ln(N)}} \\ &\leq e^{-3 \ln(N)}, \end{aligned}$$

The second inequality holds for any $N \geq 5$. The last inequality of the lemma then follows by union bound over all possible N intervals.

We now lower bound $|\mathcal{X} \cap I|$, $|\mathcal{U} \cap I|$ and $|\tilde{\mathcal{U}} \cap I|$ for any interval $I = [\frac{i}{N}; \frac{i+\ell}{N}]$ s.t. $\ell N \geq \alpha \ln(N)$. First note that:

$$\mathbb{P} \left(|\mathcal{X} \cap I| \leq (1-\epsilon) \left(1 + \frac{\epsilon}{2} \right) \ell + 16 \ln(N) + 6 \right) \leq \mathbb{P} \left(|\mathcal{X} \cap I| \leq \left(1 - \frac{\epsilon}{2} + \frac{\epsilon^2}{2} \right) \ell \right), \quad (9)$$

as, by definition of α , for any $N \geq 5$, $16 \ln(N) + 6 \leq 20 \ln(N) \leq \epsilon^2 \alpha \ln(N)$. The same holds for $\mathcal{U}$ and $\tilde{\mathcal{U}}$. By Chernoff bound:

$$\begin{aligned} \mathbb{P} \left(|\mathcal{X} \cap I| \leq (1-\epsilon) \left(1 + \frac{\epsilon}{2} \right) \ell + 16 \ln(N) + 6 \right) &\leq \mathbb{P} \left(|\mathcal{X} \cap I| \leq \left(1 - \frac{\epsilon}{4} + \frac{\epsilon^2}{4} \right) \ell \right) \\ &\leq e^{-\frac{(\epsilon - \epsilon^2)^2}{32} \ell}. \end{aligned}$$

By tail bounds for Poisson random variables:

$$\mathbb{P} \left(|\mathcal{U} \cap I| \leq \left(1 - \frac{\epsilon}{4} + \frac{\epsilon^2}{4} \right) \ell \right) \leq e^{-\frac{(\epsilon - \epsilon^2)^2}{32} \ell}.$$

As $1 \geq p_k k \geq 1 - \frac{1}{2k}$ and $k \geq \frac{4}{\epsilon - \epsilon^2}$, we also have the bound:

$$\left(1 - \frac{\epsilon - \epsilon^2}{4 \sqrt{k p_k}} \right) k p_k \ell \geq \left(1 - \frac{1}{2k} - \sqrt{k p_k} \frac{\epsilon - \epsilon^2}{4} \right) \ell \geq \left(1 - \frac{\epsilon - \epsilon^2}{2} \right) \ell.$$

Hence, by Chernoff bound:

$$\mathbb{P}\left(|\tilde{\mathcal{U}} \cap I| \leq \left(1 - \frac{\epsilon}{2} + \frac{\epsilon^2}{2}\right) \ell\right) \leq e^{-\frac{(\epsilon - \epsilon^2)^2}{32} \ell}.$$

We now compute the union bound over all possible length ℓ :

$$\sum_{\ell=\alpha N}^{+\infty} e^{-\frac{(\epsilon - \epsilon^2)^2}{32} \ell} = \frac{e^{-\frac{(\epsilon - \epsilon^2)^2}{32}}}{e^{-\frac{(\epsilon - \epsilon^2)^2}{32}} - 1} e^{-\frac{(\epsilon - \epsilon^2)^2}{8} \alpha \ln(N)} \leq \frac{64}{(\epsilon - \epsilon^2)^2} e^{-\frac{(\epsilon - \epsilon^2)^2}{32} \alpha \ln(N)} \leq \frac{\alpha}{N^3}.$$

A union bound over the N possible values of i terminates the proof. $\square$

Lemma 16 (Poissonization) *With probability at least $1 - \frac{2}{N}$, it holds that:*

$$\left| \kappa(\text{Geom}(\mathcal{X}, \mathcal{Y}, c, N)) - \kappa(\text{Geom}(\mathcal{U}, \mathcal{Y}, c, N)) \right| \leq 2\sqrt{N \ln(N)}.$$

If $\mathcal{E}_{\mathcal{I}}$ holds and $|\mathcal{U} - \mathcal{X}| \leq 2\sqrt{N \ln(N)}$, then :

$$\left| \rho(\text{Geom}(\mathcal{X}, \mathcal{Y}_{1-\epsilon})) - \rho(\text{Geom}(\mathcal{U}, \mathcal{Y}_{1-\epsilon})) \right| \leq \frac{8\sqrt{N \ln(N)}(\alpha \ln(N))^2}{N}.$$

Moreover, $|\mathcal{U} - \mathcal{X}| \leq 2\sqrt{N \ln(N)}$ holds with probability at least $1 - \frac{2}{N}$.

Proof: To obtain the first equation, we combine equation 1, which implies that the poissonisation step only adds or removes $2\sqrt{N \ln(N)}$ vertices with probability at least $1 - \frac{2}{N}$, with Lemma 14.

For the second equation, let us first assume that $|\mathcal{X}| \leq |\mathcal{U}|$ and that event $\mathcal{E}_{\mathcal{I}}$ holds. Consider x and x' , two consecutive free vertices remaining at the end of the run of CLOSEST in $\text{Geom}(\mathcal{X}, \mathcal{Y})$. Let us study the impact of adding an offline vertex between x and x' . This can only modify the match of vertices lying in the interval $[x, x']$ to another vertex within that interval. An upper bound on the modification to the length of the matching is thus $|x' - x| \times |\mathcal{V}_{\epsilon} \cap [x, x']|$. We have:

$$\begin{aligned} |\mathcal{V}_{\epsilon} \cap [x, x']| &\leq |\mathcal{V}_{\epsilon} \cap [\frac{\lfloor Nx + 1 \rfloor}{N}, \frac{\lfloor Nx' \rfloor}{N}]| + |\mathcal{V}_{\epsilon} \cap [\frac{\lfloor Nx \rfloor}{N}, \frac{\lfloor Nx + 1 \rfloor}{N}]| + |\mathcal{V}_{\epsilon} \cap [\frac{\lfloor Nx' \rfloor}{N}, \frac{\lfloor Nx' + 1 \rfloor}{N}]| \\ &\leq |\mathcal{V}_{\epsilon} \cap [\frac{\lfloor Nx + 1 \rfloor}{N}, \frac{\lfloor Nx' \rfloor}{N}]| + 6 \ln(N) + 2 \end{aligned}$$

where the second inequality holds under event $\mathcal{E}_{\mathcal{I}}$. On the other hand,

$$|\mathcal{U} \cap [x, x']| \geq |\mathcal{U} \cap [\frac{\lfloor Nx + 1 \rfloor}{N}, \frac{\lfloor Nx' \rfloor}{N}]|.$$

All the vertices in $\mathcal{U} \cap [x, x']$ have been matched to vertices in $\mathcal{V}_{\epsilon} \cap [x, x']$, so $|\mathcal{V}_{\epsilon} \cap [x, x']| \geq |\mathcal{U} \cap [x, x']|$. This implies that $x' - x \leq \frac{\alpha \ln(N) + 3}{N}$. Thus:

$$\begin{aligned} |x' - x| \times |\mathcal{V}_{\epsilon} \cap [x, x']| &\leq (1 - \epsilon) \left(1 + \frac{\epsilon}{2}\right) (\alpha \ln(N) + 5) \frac{\alpha \ln(N) + 3}{N} \\ &\leq \frac{(\alpha \ln(N) + 5)^2}{N}. \end{aligned}$$

To modify $\mathcal{X}$ into $\mathcal{U}$, we need to add $|\mathcal{U} - \mathcal{X}|$ vertices, which upon each addition modifies the length of the matching by at most $\frac{(\alpha \ln(N) + 5)^2}{N}$, which is smaller than $\frac{4(\alpha \ln(N))^2}{N}$. The same reasoning can be applied if $|\mathcal{X}| \geq |\mathcal{U}|$. $\square$

Lemma 17 (Rounding) *With probability at least $1 - \exp(-\frac{2N}{k^2}) - \frac{2}{N}$, it holds that:*

$$\left| \kappa(\text{Geom}(\mathcal{U}[k], \mathcal{Y}, c, N)) - \kappa(\text{Geom}(\mathcal{U}, \mathcal{Y}, c, N)) \right| \leq \frac{20N}{k},$$

and with probability at least $1 - \exp\left(-\frac{2N}{k^2}\right) - \frac{2}{N} - \frac{4\alpha+1}{N^2}$.

$$|\rho(\text{Geom}(\mathcal{U}[k], \mathcal{Y}_{1-\epsilon})) - \rho(\text{Geom}(\mathcal{U}, \mathcal{Y}_{1-\epsilon}))| \leq \frac{80(\alpha \ln(N))^2}{k}.$$

Proof: Consider a i -CLOSEST algorithm, which matches the incoming vertex to the closet rounded vertex up to iteration i , then to the closest vertex. In the case of ties with the rounded coordinates, they are broken following the unrounded ones' order. The runs of algorithms i -CLOSEST and $i+1$ -CLOSEST only differ if y_i 's closest neighbor upon arrival is modified by the rounding of the coordinates, which we denote event $\mathcal{E}_i$. Note that this can only happen if y_i falls within $2/Nk$ of the middle of the segment between two consecutive free vertices at iteration i . As $y_i \sim \mathcal{U}[0, 1]$,

$$\mathbb{P}(\mathcal{E}_i | \mathcal{U}, y_1, \dots, y_{i-1}, |\mathcal{U}| \leq N + 2\sqrt{N \ln(N)}) \leq \frac{4|\mathcal{U}|}{Nk} \leq \frac{8}{k}.$$

By Azuma-Hoeffding,

$$\mathbb{P}\left(\sum_{i=1}^N \mathbb{1}_{\mathcal{E}_i} \geq \frac{10N}{k} \mid |\mathcal{U}| \leq N + 2\sqrt{N \ln(N)}\right) \leq e^{-\frac{2N}{k^2}}. \quad (10)$$

It now remains to combine with Equation 1, Lemma 14 and a union bound to get the first equation of the Lemma. The proof of the second one continues.

Let us assume $\mathcal{E}_i$ holds, and denote u the match of y_i by i -CLOSEST and u' the match by $i+1$ -CLOSEST. In the proof of the Lemma, denote $\mathcal{U}_i \setminus \{u, u'\}$ the set of free vertices obtained when running i -CLOSEST with u, u' , and y_i removed.

Let u_a and u_b be the two closest vertices in $\mathcal{U}_i \setminus \{u, u'\}$ s.t. $u \in [u_a; u_b]$. By definition of i -CLOSEST, all vertices in $]u_a; u_b[\cap \mathcal{U}_i \setminus \{u, u'\}$ are matched to vertices in $\left[\frac{\lfloor Nu_a \rfloor}{N}; u_b\right] \cap \mathcal{Y}_{1-\epsilon}$, which implies that

$$\begin{aligned} & \left|]u_a; u_b[\cap \mathcal{U} \right| - 2 \leq \left| \left[\frac{\lfloor Nu_a \rfloor}{N}; u_b \right] \cap \mathcal{Y}_{1-\epsilon} \right|, \\ \Rightarrow & \left| \left[\frac{\lfloor Nu_a + 1 \rfloor}{N}; \frac{\lfloor Nu_b \rfloor}{N} \right] \cap \mathcal{U} \right| - 2 \leq \left| \left[\frac{\lfloor Nu_a \rfloor}{N}; \frac{\lfloor Nu_a + 1 \rfloor}{N} \right] \cap \mathcal{Y}_{1-\epsilon} \right| + \left| \left[\frac{\lfloor Nu_a + 1 \rfloor}{N}; \frac{\lfloor Nu_b \rfloor}{N} \right] \cap \mathcal{Y}_{1-\epsilon} \right| \\ & \quad + \left| \left[\frac{\lfloor Nu_b \rfloor}{N}; \frac{\lfloor Nu_b + 1 \rfloor}{N} \right] \cap \mathcal{Y}_{1-\epsilon} \right| \end{aligned}$$

Since $\mathcal{E}_i$ holds, this implies that $|u_a - u_b|N \leq \alpha \ln(N) + 2$. Still since $\mathcal{E}_i$ holds,

$$\left| \left[u_a - \frac{1}{N}; u_b + \frac{1}{N} \right] \cap \mathcal{Y}_{1-\epsilon} \right| \leq (12 + \alpha) \ln(N) + 6.$$

Now adding back u can only modify the match of the vertices in $[u_a - \frac{1}{N}; u_b + \frac{1}{N}]$, as u_a and u_b remain unmatched without u . Moreover, the match of those vertices can only be modified to another vertex within that interval. This implies that the length of the matching created with and without u differ by at most $[(12 + \alpha) \ln(N) + 6] \left(\frac{\alpha \ln(N) + 2}{N} \right)$. Repeating the same reasoning with u' , we get that the modification of the match of y_i modifies the length of the matching by at most $2[(12 + \alpha) \ln(N) + 6] \left(\frac{\alpha \ln(N) + 2}{N} \right)$, which is smaller than $8\alpha^2 \frac{\ln(N)^2}{N}$.

This combined with Equations 10, 1 and Lemma 15 terminates the proof of the lemma. $\square$

Lemma 18 (Discard) *With probability at least $1 - \frac{4}{N}$, it holds that:*

$$\left| \kappa(\text{Geom}(\tilde{\mathcal{U}}[k], \mathcal{Y}, c, N)) - \kappa(\text{Geom}(\mathcal{U}[k], \mathcal{Y}, c, N)) \right| \leq \frac{N}{k} + 4\sqrt{N \ln(N)}.$$

If $\mathcal{E}_{\mathcal{I}}$ holds, then

$$\left| \rho \left(\text{Geom} \left(\tilde{\mathcal{U}}[k], \mathcal{Y}_{1-\epsilon} \right) \right) - \rho \left(\text{Geom} \left(\mathcal{U}[k], \mathcal{Y}_{1-\epsilon} \right) \right) \right| \leq \left(\frac{N}{k} + 4\sqrt{N \ln(N)} \right) \frac{4(\alpha \ln(N))^2}{N}.$$

Proof: For any $\ell \in [kN]$, let

$$n_\ell := \left\{ \frac{\lfloor uNk \rfloor}{Nk} = \frac{\ell}{Nk} \mid u \in \mathcal{U} \right\}.$$

The points in $\mathcal{U}$ are generated through a Poisson point process of intensity N in $[0, 1]$, thus, for any $\ell \in [Nk]$:

$$\mathbb{P}(n_\ell > 0) = 1 - e^{-1/k},$$

and the $(n_\ell)_{\ell \in [Nk]}$ are independent of each other. The number of points in $\tilde{\mathcal{U}}$ is

$$|\tilde{\mathcal{U}}| = \sum_{\ell=1}^{Nk} \mathbb{1}_{n_\ell > 0}.$$

We have

$$\begin{aligned} \mathbf{E} \left[|\tilde{\mathcal{U}}| \right] &= Nk(1 - e^{-1/k}) \\ &\geq N - \frac{N}{k}. \end{aligned}$$

By Chernoff bound,

$$\mathbb{P}(|\tilde{\mathcal{U}}| \leq N - \frac{N}{k} - 2\sqrt{N \ln(N)}) \leq \frac{2}{N}$$

We have

$$N_{\text{removed}} := |\mathcal{U}| - |\tilde{\mathcal{U}}|.$$

We have already obtained by concentration of Poisson random variables

$$\mathbb{P} \left(|N - |\mathcal{U}|| \geq 2\sqrt{N \ln(N)} \right) \leq \frac{2}{N}.$$

Thus,

$$\mathbb{P} \left(N_{\text{removed}} \geq 4\sqrt{N \ln(N)} + \frac{N}{k} \right) \leq \frac{4}{N}.$$

Combining this with Lemma 14 terminates the proof of the first equation. For the second, we proceed in the same way as in the proof of Lemma 16, using this time the lower bound on the number of vertices in $\tilde{\mathcal{U}}$ per interval. $\square$

Lemma 19 (Gluing) *With probability at least $1 - \exp(-\sqrt{N})$, it holds:*

$$\left| \kappa \left(\text{Geom} \left(\tilde{\mathcal{U}}[k], \mathcal{Y}, c, N \right) \right) - \kappa \left(G_{\text{glued}} \left(\tilde{\mathcal{U}}[k], \mathcal{Y}, c, N \right) \right) \right| \leq 2\sqrt{N} + 1.$$

Under event $\mathcal{E}_{\mathcal{I}}$, it holds that:

$$\left| \rho \left(\text{Geom} \left(\tilde{\mathcal{U}}[k], \mathcal{Y}_{1-\epsilon} \right) \right) - \rho \left(G_{\text{glued}} \left(\tilde{\mathcal{U}}[k], \mathcal{Y}_{1-\epsilon} \right) \right) \right| \leq \frac{\alpha^5 \ln(N)^3}{18N}$$

Proof: By Lemma 14, adding the vertex at coordinate 0 modifies the score of the algorithm by at most one.

The match of some vertex $y \in \mathcal{Y}$ may be modified by the gluing step only if $y < \frac{c}{N}$ or $y > 1 - \frac{c}{N}$.

$$N_{\text{gluing}} := \sum_{y \in \mathcal{Y}} \mathbb{1}_{\{\text{the match of } y \text{ is modified during the gluing step}\}}.$$

By Chernoff bound,

$$\mathbb{P}(N_{\text{gluing}} \geq 2c + \sqrt{N}) \leq e^{-\sqrt{N}}.$$

Lemma 14 concludes the proof of the first inequality.

For the second, we proceed as in the proof of Lemma 17.

Consider algorithm i -CLOSEST' which matches the incoming vertex to the closest vertex in the glued space up to iteration i , then to the closest on the line. Denote $\mathcal{E}'_i$ the event that y_i 's closest neighbor is modified. This can happen only if y_i lands between the remaining vertex of highest and smallest coordinates at iteration i . This interval is included in the interval $[u_h, u_\ell]$ between the remaining vertex of highest and smallest coordinates at iteration $N(1 - \epsilon)$ when running CLOSEST algorithm with the glued coordinates at every iteration. By the same reasoning we employed before, under event $\mathcal{E}_I$ this interval is at most of length $\frac{\alpha \ln(N) + 2}{N}$ and there are at most $\alpha \ln(N) + 2$ online vertices that have their match directly modified by the gluing step.

Let us now bound the impact of this modification. As before, denote u and u' the two matches of y_i and denote $\tilde{\mathcal{U}}_i \setminus \{u, u'\}$ the remaining free vertices at the end of the run of i -CLOSEST' with u, u' and y_i removed. Denote u_h^i, u_ℓ^i the remaining vertex of highest and smallest coordinates at iteration i . All the vertices in $\mathcal{U} \cap [u_h^i, 0]$ are matched to vertices in $\mathcal{Y}_{1-\epsilon} \cap [u_h^i, u_\ell^i]$. Thus we have:

$$\begin{aligned} |\mathcal{U} \cap [u_h^i, 0]| &\leq |\mathcal{Y}_{1-\epsilon} \cap [u_h^i, u_\ell^i]|, \\ &\leq |\mathcal{Y}_{1-\epsilon} \cap [u_h^i, 0]| + \alpha \ln(N) + 2. \end{aligned}$$

This implies that $|1 - u_h^i| \leq \frac{\alpha^2 \ln(N)}{12N}$. Similarly, we get the bound $u_\ell^i \leq \frac{\alpha^2 \ln(N)}{12N}$. Adding back u or u' to the offline vertices can only modify the match of vertices in the interval $[u_h^i, u_\ell^i]$ to other vertices in that interval. We thus have that for each modification, the impact on the length is bounded by $\frac{\alpha^4 \ln(N)^2}{36N}$. $\square$

4.2 Proof of Lemma 3 (Gaps repartition)

Denote $|H|$ the length of a sequence H and $S([N])$ the ensemble of all permutations over $[N]$. For a sequence $H = (h_i)_{i \in [K]}$ let $\mathcal{A}_t(H)$ be the event:

$$\left\{ N_t = |H|, \exists \sigma \in S([N_t]) \text{ s.t. } \forall i \in [N_t], u_t(i+1) - u_t(i) = \frac{h_{\sigma(i)}}{Nk} \right\},$$

and $\mathcal{A}_t(\sigma, H)$ the event:

$$\left\{ N_t = |H|, \forall i \in [N_t], u_t(i+1) - u_t(i) = \frac{h_{\sigma(i)}}{Nk} \right\}.$$

Note that $\mathcal{A}_t(H) = \cup_{\sigma \in S([H])} \mathcal{A}_t(\sigma, H)$. For a list of sequences $H_1, \dots, H_t$, we denote $\mathcal{A}_{1:t}(H_{1:t})$ the event that $\mathcal{A}_s(H_s)$ hold for all $s \leq t$:

$$\mathcal{A}_{1:t}(H_{1:t}) := \cap_{1 \leq s \leq t} \mathcal{A}_s(H_s).$$

Lemma 3 is a consequence of the following stronger Lemma.

Lemma 20 *For any sequence H , any iteration t , any two permutations $\sigma, \sigma' \in S([|H|])$,*

$$\mathbb{P} \left(\mathcal{A}_t(\sigma, H_t) \mid \mathcal{A}_{1:t}(H_{1:t}) \right) = \mathbb{P} \left(\mathcal{A}_t(\sigma', H_t) \mid \mathcal{A}_{1:t}(H_{1:t}) \right). \quad (11)$$

Proof: We prove this lemma by induction. Let $K < Nk$ be a non-negative integer and let $H = (h_i)_i$ be any sequence of integers s.t.

$$\sum_{i=1}^{|H|} \frac{h_i}{Nk} = 1.$$

Since $u_0(1) = 0$, the knowledge of the sizes and the ordering of the gaps determines the position of the points. Thus, for any $\sigma \in S([H])$:

$$\mathbb{P}\left(\forall i \in [H], u_0(i+1) - u_0(i) = \frac{h_{\sigma(i)}}{Nk}\right) = p_k^{|H|-1} (1 - p_k)^{Nk - |H|}.$$

This does not depend on the choice of permutation σ . Thus for any sequence H s.t. $\sum_{i=1}^{|H|} \frac{h_i}{Nk} = 1$ and any two permutations $(\sigma, \sigma') \in S([H])^2$:

$$\mathbb{P}\left(\mathcal{A}_0(\sigma, H) \mid \mathcal{A}_0(H)\right) = \mathbb{P}\left(\mathcal{A}_0(\sigma', H) \mid \mathcal{A}_0(H)\right). \quad (12)$$

Equation (11) thus holds at iteration 0. Let us assume that Equation (11) holds at all iterations until the t -th one. Let us show this implies that it also holds at iteration $t+1$. There are two cases possible, depending on whether or not a vertex is matched.

We first show the implication in the case where the incoming vertex is not matched, i.e. it lays at a distance larger than c/N of any free vertex. Note y_t the incoming vertex at iteration t . For any $\sigma \in S([N_t])$, any H_t a sequence of length N_t s.t. $\sum_{i=1}^{|H_t|} \frac{h_i}{Nk} = 1$,

$$\mathbb{P}(y_t \text{ is not matched} \mid \mathcal{A}_t(\sigma, H_t)) = 1 - \sum_{i=1}^{N_t} \frac{\min(2c, h_i/k)}{N} \quad (13)$$

The following therefore holds:

$$\mathbb{P}\left(\mathcal{A}_{t+1}(\sigma, H_t) \mid \mathcal{A}_{1:t}(H_{1:t})\right) = \mathbb{P}\left(y_t \text{ is not matched}, \mathcal{A}_t(\sigma, H_t) \mid \mathcal{A}_{1:t}(H_{1:t})\right).$$

By the induction hypothesis and Equation (13), the right term does not depend on σ , thus, for any σ, σ' in $S([N_t])$,

$$\mathbb{P}\left(\mathcal{A}_{t+1}(\sigma, H_t) \mid \mathcal{A}_{1:t}(H_{1:t}) \cap \mathcal{A}_{t+1}(H_t)\right) = \mathbb{P}\left(\mathcal{A}_{t+1}(\sigma', H_t) \mid \mathcal{A}_{1:t}(H_{1:t}) \cap \mathcal{A}_{t+1}(H_t)\right).$$

We now turn to the case where y_t is matched. We define an admissible sequence for sequence H and a couple $j < j' \leq |H|$ as a sequence $\tilde{H}(j, j')$ s.t. for any $i \in [|H| - 1]$

$$\tilde{h}_i^{j, j'} = \begin{cases} h_i & \text{if } i < j' \text{ and } i \neq j \\ h_j + h_{j'} & \text{if } i = j \\ h_{i+1} & \text{if } i \geq j'. \end{cases}$$

Note it is the sequence of gaps obtained when event $\mathcal{A}_t(H)$ is true and a vertex $u_t(i)$ with $u_t(i) - u_t(i-1) = h_j$ and $u_t(i+1) - u_t(i) = h_{j'}$ is matched.

For any $H_{1:t}$, any $j < j' \leq |H_t|$, any $\sigma \in S([|H_t| - 1])$ define events :

$$\begin{aligned} \mathcal{B}_t(H_t, i, j, j') &= \left\{ u_t(i) - u_t(i-1) = \frac{h_j}{Nk} \text{ and } u_t(i+1) - u_t(i) = \frac{h_{j'}}{Nk} \right\} \\ &\cup \left\{ u_t(i) - u_t(i-1) = \frac{h_{j'}}{Nk} \text{ and } u_t(i+1) - u_t(i) = \frac{h_j}{Nk} \right\}. \end{aligned}$$

and

$$\mathcal{C}_t(\sigma, H_t, i, j, j') = \left\{ \forall k < i-1, u_t(k+1) - u_t(k) = \frac{\tilde{h}_{\sigma(k)}^{j, j'}}{Nk} \right\} \cap \left\{ \forall k > i, u_t(k+1) - u_t(k) = \frac{\tilde{h}_{\sigma(k-1)}^{j, j'}}{Nk} \right\}.$$

Note $\mathcal{E}_t(\sigma, H_t, i, j, j') = \mathcal{B}_t(H_t, i, j, j') \cap \mathcal{C}_t(\sigma, H_t, i, j, j')$. For any $H_{1:t}$, any $j < j' \leq |H_t|$, any $\sigma \in S(|H_t| - 1)$, it holds that:

$$\begin{aligned} \mathbb{P} \left(\mathcal{A}_{t+1}(\sigma, \tilde{H}_t(j, j')) \mid \mathcal{A}_{1:t}(H_{1:t}) \right) &= \sum_{i \text{ s.t. } \tilde{h}_{\sigma(i-1)} = \tilde{h}_j} \mathbb{P} \left(u_t(i) \text{ is matched, } \mathcal{E}_t(\sigma, H_t, i, j, j') \mid \mathcal{A}_{1:t}(H_{1:t}) \right) \\ &= \frac{\min(c, \frac{h_j}{k}) + \min(c, \frac{h_{j'}}{k})}{N} \sum_{i \text{ s.t. } \tilde{h}_{\sigma(i-1)} = \tilde{h}_j} \mathbb{P} \left(\mathcal{E}_t(\sigma, H_t, i, j, j') \mid \mathcal{A}_{1:t}(H_{1:t}) \right). \end{aligned}$$

By the induction property, the right term does not depend on σ . This implies that the induction property remains true when a vertex is matched as well, which implies that Equation (11) holds for all $t \in [N]$. $\square$

We now show that Lemma 20 implies Lemma 3. Let $\mathcal{F}_t$ be the event associated with the values $(F_{k,N}(\ell, t'))_{\ell \in [Nk]}$ for all $t' \leq t$, which also determines the value of N_t . Let H be any sequence of length N_t s.t. for any $\ell \in [Nk]$, $|\{h_i = \ell \mid i \in [N_t]\}| = F_{k,N}(\ell, t)$. Equation (11) implies:

$$\begin{aligned} \mathbb{E} \left[M_{k,N}(\ell_-, \ell_+, t) \mid \mathcal{F}_t \right] &= \frac{1}{|S([N_t])|} \sum_{\sigma \in S([N_t])} \sum_{i=1}^{N_t} \mathbb{1}_{\{h_{\sigma(i)} = \ell_-, h_{\sigma(i+1)} = \ell_+\}} \\ &= \sum_{i=1}^{N_t} \frac{1}{|S([N_t])|} \sum_{\sigma \in S([N_t])} \mathbb{1}_{\{h_{\sigma(i)} = \ell_-, h_{\sigma(i+1)} = \ell_+\}} \\ &= \sum_{i=1}^{N_t} \frac{|S([N_t - 2])|}{|S([N_0])|} [\mathbb{1}_{\{\ell_- \neq \ell_+\}} F_{k,N}(\ell_-, t) F_{k,N}(\ell_+, t)] \\ &\quad + \sum_{i=1}^{N_t} \frac{|S([N_t - 2])|}{|S([N_0])|} \mathbb{1}_{\{\ell_- = \ell_+\}} F_{k,N}(\ell_-, t) (F_{k,N}(\ell_+, t) - 1) \\ &= F_{k,N}(\ell_-, t) \left(\mathbb{1}_{\{\ell_- \neq \ell_+\}} \frac{F_{k,N}(\ell_+, t)}{N_t - 1} + \mathbb{1}_{\{\ell_- = \ell_+\}} \frac{F_{k,N}(\ell_+, t) - 1}{N_t - 1} \right). \end{aligned}$$

$\square$

4.3 Analysis of the PDEs

In this section, we prove some properties of the functions f_k, f, g_k and g that are useful for the proofs of the following sections.

Lemma 21 (Total length invariant) For any $t \in [0, 1]$,

$$\sum_{\ell=1}^{+\infty} \frac{\ell}{k} f_k(\ell, t) = 1 \text{ and } \int_{x=0}^{+\infty} f(x, t) = 1.$$

The same holds for g_k and g .

Proof: This is true for $t = 0$ by definition of the initial condition. We now show that this quantity is an invariant of the system of ODEs. We have:

$$\begin{aligned} \frac{\partial \sum_{\ell=1}^{+\infty} \frac{\ell}{k} f_k(\ell, t)}{\partial t} &= - \sum_{\ell=1}^{+\infty} \frac{\ell}{k} \min(\frac{\ell}{k}, 2c) f(\frac{\ell}{k}, t) - \frac{1}{\sum_{\ell=1}^{+\infty} f_k(\ell, t)} \sum_{\ell=1}^{+\infty} \min(\frac{\ell}{k}, 2c) f(\frac{\ell}{k}, t) \sum_{\ell=1}^{+\infty} \frac{\ell}{k} f(\frac{\ell}{k}, t) \\ &\quad + \frac{1}{\sum_{\ell=1}^{+\infty} f_k(\ell, t)} \sum_{\ell=1}^{+\infty} \sum_{\ell'=0}^{+\infty} \min(\frac{\ell'}{k}, 2c) f_k(\ell', t) (\ell - \ell') f_k(\ell - \ell', t) \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{\sum_{\ell=1}^{+\infty} f_k(\ell, t)} \sum_{\ell=1}^{+\infty} \sum_{\ell'=0}^{+\infty} \ell' \min\left(\frac{\ell'}{k}, 2c\right) f_k(\ell', t) f_k(\ell - \ell', t), \\
& = - \sum_{\ell=1}^{+\infty} \frac{\ell}{k} \min\left(\frac{\ell}{k}, 2c\right) f\left(\frac{\ell}{k}, t\right) - \frac{1}{\sum_{\ell=1}^{+\infty} f_k(\ell, t)} \sum_{\ell=1}^{+\infty} \min\left(\frac{\ell}{k}, 2c\right) f\left(\frac{\ell}{k}, t\right) \sum_{\ell=1}^{+\infty} \frac{\ell}{k} f\left(\frac{\ell}{k}, t\right) \\
& \quad + \frac{1}{\sum_{\ell=1}^{+\infty} f_k(\ell, t)} \sum_{\ell=1}^{+\infty} \min\left(\frac{\ell}{k}, 2c\right) f\left(\frac{\ell}{k}, t\right) \sum_{\ell=1}^{+\infty} \frac{\ell}{k} f\left(\frac{\ell}{k}, t\right) \\
& \quad + \frac{1}{\sum_{\ell=1}^{+\infty} f_k(\ell, t)} \sum_{\ell=1}^{+\infty} \frac{\ell}{k} \min\left(\frac{\ell}{k}, 2c\right) f\left(\frac{\ell}{k}, t\right) \sum_{\ell=1}^{+\infty} f_k(\ell, t), \\
& = 0.
\end{aligned}$$

The computation is similar for f , g and g_k . □

Lemma 22 (Bound L_1 norm) For any $t \in [0, 1]$,

$$e^{-4c} \leq \sum_{\ell=1}^{+\infty} f_k(\ell, t) \leq 1 \text{ and } e^{-4c} \leq \int_0^{+\infty} f(x, t) dx \leq 1.$$

Also, for any $t \in [0, 1]$,

$$\sum_{\ell=1}^{+\infty} g_k(\ell, t) = 1 - t = \int_0^{+\infty} g(x, t) dx.$$

Proof:

$$\begin{aligned}
\frac{\partial \sum_{\ell=1}^{+\infty} f_k(\ell, t)}{\partial t} & = - \sum_{\ell=1}^{+\infty} \min\left(\frac{\ell}{k}, 2c\right) f\left(\frac{\ell}{k}, t\right) - \frac{1}{\sum_{\ell=1}^{+\infty} f_k(\ell, t)} \sum_{\ell=1}^{+\infty} \min\left(\frac{\ell}{k}, 2c\right) f\left(\frac{\ell}{k}, t\right) \sum_{\ell=1}^{+\infty} \frac{\ell}{k} f\left(\frac{\ell}{k}, t\right) \\
& \quad + \frac{1}{\sum_{\ell=1}^{+\infty} f_k(\ell, t)} \sum_{\ell=1}^{+\infty} \sum_{\ell'=0}^{+\infty} \min\left(\frac{\ell'}{k}, 2c\right) f_k(\ell', t) f_k(\ell - \ell', t), \\
& = - \sum_{\ell=1}^{+\infty} \min\left(\frac{\ell}{k}, 2c\right) f\left(\frac{\ell}{k}, t\right).
\end{aligned}$$

This implies that for any $t \in [0, 1]$, $e^{-4c} \leq e^{-4ct} \leq \sum_{\ell=1}^{+\infty} f_k(\ell, t) \leq \sum_{\ell=1}^{+\infty} f_k(\ell, 0) = 1$.

We have:

$$\frac{\partial \sum_{\ell=1}^{+\infty} g_k(\ell, t)}{\partial t} = - \sum_{\ell=1}^{+\infty} \frac{\ell}{k} g_k\left(\frac{\ell}{k}, t\right) = -1.$$

So, $\sum_{\ell=1}^{+\infty} g_k(\ell, t) = 1 - t$, and the same holds for g . □

4.4 Proofs of the technical lemmas for section 3.3

Proof of lemma 6 : In this proof, we use the shorthand $a_\ell = \frac{F_{k,N}(\ell, t)}{N}$. For any $t \leq N$, $\sum_{\ell=1}^{Nk} \frac{\ell}{k} a_\ell = 1$. This and Lemma 21 imply:

$$\sum_{\ell=kN^{1/8}}^{Nk} a_\ell \leq N^{-1/8} \quad \text{and} \quad r_{N,k}(t) \leq N^{-1/8}, \quad (14)$$

so that $|\sum_{\ell=kN^{1/8}}^{Nk} a_\ell - r_{N,k}(t)| \leq 2N^{-1/8}$.

If $(a_\ell)_{\ell=1}^{kN^{1/8}} \in \mathcal{D}$, for any N large enough s.t. $N^{-1/8} \leq \frac{1}{4}e^{-4c}$,

$$\sum_{\ell=1}^{kN^{1/8}} a_\ell \geq \frac{1}{2}e^{-4c} - \sum_{\ell=kN^{1/8}+1}^{kN} a_\ell \geq \frac{1}{4}e^{-4c}.$$

Therefore, for any N s.t. $N^{-1/8} \leq \frac{1}{4}e^{-4c}$, if $(a_\ell)_{\ell=1}^{kN^{1/8}} \in \mathcal{D}$,

$$\left| \frac{1}{N_t/N} - \frac{1}{\sum_{\ell=1}^{kN^{1/8}} a_\ell + r_{N,k}(t)} \right| \leq w_2 N^{-1/8}, \quad (15)$$

with w_2 a constant that does not depend on N . Now, for any $\ell \leq kN^{1/8}$, we have:

$$\begin{aligned} |A_{\ell,t} - \Phi_{k,N}^\ell(t, (a_\ell)_{\ell=1}^{kN^{1/8}})| &\leq 2c \left| \frac{1}{N_t/N} - \frac{1}{\sum_{\ell'=1}^{kN^{1/8}} a_{\ell'} + r_{N,k}(t)} \right| \sum_{\ell'=0}^{\ell} a_{\ell'} a_{\ell-\ell'} \\ &\quad + \underbrace{\left| \sum_{\ell'=1}^{kN} \min\left(\frac{\ell'}{k}, 2c\right) \frac{N}{N_t} a_{\ell'} - \frac{\sum_{\ell'=1}^{kN^{1/8}} \min\left(\frac{\ell'}{k}, 2c\right) a_{\ell'} + \tilde{r}_{N,k}(t)}{\sum_{\ell'=1}^{kN^{1/8}} a_{\ell'} + r_{N,k}(t)} \right|}_{(i)} a_\ell. \end{aligned}$$

We start by bounding (i). First, since $\sum_{\ell'} a_{\ell'} \leq 1$ and $r_{N,k}(t) \leq N^{-1/8}$, we have $\sum_{\ell'=1}^{kN^{1/8}} \min\left(\frac{\ell'}{k}, 2c\right) a_{\ell'} + \tilde{r}_{N,k}(t) \leq 4c$. By Equation 15, this entails:

$$\begin{aligned} (i) &\leq \left| \sum_{\ell'} \min\left(\frac{\ell'}{k}, 2c\right) \frac{a_{\ell'}}{N_t/N} - \frac{\sum_{\ell'=1}^{kN^{1/8}} \min\left(\frac{\ell'}{k}, 2c\right) a_{\ell'} + \tilde{r}_{N,k}(t)}{N_t/N} \right| + 4cw_2 N^{-1/8} \\ &= \left| \sum_{\ell'=kN^{1/8}}^{Nk} \min\left(\frac{\ell'}{k}, 2c\right) \frac{a_{\ell'}}{N_t/N} - \frac{\tilde{r}_{N,k}(t)}{N_t/N} \right| + 4cw_2 N^{-1/8} \\ &\leq 4e^{4c} \left| \sum_{\ell'=kN^{1/8}}^{Nk} \min\left(\frac{\ell'}{k}, 2c\right) a_{\ell'} - \tilde{r}_{N,k}(t) \right| + 4cw_2 N^{-1/8} \\ &\leq 4c(4e^{4c} + w_2) N^{-1/8}. \end{aligned}$$

The third inequality holds as $N_t/N \geq 4e^{4c}$, and the last one by equation 14. This entails:

$$\left| A_{\ell,t} - \Phi_{k,N}^\ell(t, (a_\ell)_{\ell=1}^{kN^{1/8}}) \right| \leq 2cw_2 N^{-1/8} \sum_{\ell'=0}^{\ell} a_{\ell'} a_{\ell-\ell'} + 4c(4e^{4c} + w_2) N^{-1/8} N^{-1/8} a_\ell.$$

By Young's convolution inequality, $\sum_{\ell} \sum_{\ell'=0}^{\ell} a_{\ell'} a_{\ell-\ell'} \leq (\sum_{\ell} a_\ell)^2 \leq 1$, so:

$$\sum_{\ell=1}^{kN^{1/8}} \left| A_{\ell,t} - \Phi_{k,N}^\ell(t, (a_\ell)_{\ell=1}^{kN^{1/8}}) \right| \leq (8ce^{4c} + 6cw_4) N^{-1/8}$$

Combining with Lemma 4, we obtain:

$$\sum_{\ell=1}^{kN^{1/8}} \left| \mathbb{E} \left[F_{k,N}(\ell, t+1) - F_{k,N}(\ell, t) \middle| \mathcal{F}_t, (a_\ell) \in \mathcal{D} \right] - \Phi_{k,N}^\ell(t, (a_\ell)_{\ell=1}^{kN^{1/8}}) \right| \leq \frac{2c(4e^{4c} + 3w_2)}{N^{1/8}} + \frac{24e^{4c}k}{N^{7/8}}$$

$$\leq \frac{w_5}{N^{1/8}}.$$

Note that w_5 depends on c and k but not on N . □

Proof of lemma 7 : Consider two sequences $\mathbf{a} = (a_\ell)_{\ell=1}^{kN^{1/8}}$ and $\mathbf{b} = (b_\ell)_{\ell=1}^{kN^{1/8}}$ in $\mathcal{D}$.

$$\begin{aligned} & \|\Phi_{k,N}(t, \mathbf{a}) - \Phi_{k,N}(t, \mathbf{b})\|_1 \leq 2c\|\mathbf{a} - \mathbf{b}\|_1 \\ & + \sum_{\ell} \underbrace{\left| \frac{\sum_{\ell'=1}^{kN^{1/8}} \min(\frac{\ell'}{k}, 2c)a_{\ell'} + \tilde{r}_{N,k}(t)}{\sum_{\ell'=1}^{kN^{1/8}} a_{\ell'} + r_{N,k}(t)} a_{\ell} - \frac{\sum_{\ell'=1}^{kN^{1/8}} \min(\frac{\ell'}{k}, 2c)b_{\ell'} + \tilde{r}_{N,k}(t)}{\sum_{\ell'=1}^{kN^{1/8}} b_{\ell'} + r_{N,k}(t)} b_{\ell} \right|}_{(a)} \\ & + \sum_{\ell} \underbrace{\left| \frac{\sum_{\ell'=0}^{\ell} \min(\frac{\ell'}{k}, 2c)a_{\ell'} a_{\ell-\ell'}}{\sum_{\ell'=1}^{kN^{1/8}} a_{\ell'} + r_{N,k}(t)} - \frac{\sum_{\ell'=0}^{\ell} \min(\frac{\ell'}{k}, 2c)b_{\ell'} b_{\ell-\ell'}}{\sum_{\ell'=1}^{kN^{1/8}} b_{\ell'} + r_{N,k}(t)} \right|}_{(b)}. \end{aligned}$$

Assume w.l.o.g. $\sum_{\ell=1}^{kN^{1/8}} b_{\ell} \geq \sum_{\ell=1}^{kN^{1/8}} a_{\ell}$. We have:

$$\begin{aligned} \frac{1}{\sum_{\ell} a_{\ell} + r_{N,k}(t)} & \geq \frac{1}{\sum_{\ell} b_{\ell} + r_{N,k}(t)} \\ & = \frac{1}{\sum_{\ell} a_{\ell} + r_{N,k}(t)} \left(\frac{1}{1 + \frac{\sum_{\ell} (b_{\ell} - a_{\ell})}{\sum_{\ell} a_{\ell} + r_{N,k}(t)}} \right) \\ & \geq \frac{1}{\sum_{\ell} a_{\ell} + r_{N,k}(t)} \left(1 - \frac{\sum_{\ell} b_{\ell} - a_{\ell}}{\sum_{\ell} a_{\ell} + r_{N,k}(t)} \right). \end{aligned}$$

As $\sum_{\ell} a_{\ell} \geq \frac{1}{2}e^{-4c}$, we have:

$$\left| \frac{1}{\sum_{\ell} a_{\ell} + r_{N,k}(t)} - \frac{1}{\sum_{\ell} b_{\ell} + r_{N,k}(t)} \right| \leq 4e^{8c}\|\mathbf{a} - \mathbf{b}\|_1. \quad (16)$$

Bound of (a) It holds that $\sum_{\ell=1}^{kN^{1/8}} \min(\frac{\ell}{k}, 2c)b_{\ell} + \tilde{r}_{N,k}(t) \leq 6c$, which together with Equation 16 implies:

$$\left| \frac{\sum_{\ell'=1}^{kN^{1/8}} \min(\frac{\ell'}{k}, 2c)b_{\ell'} + \tilde{r}_{N,k}(t)}{\sum_{\ell'=1}^{kN^{1/8}} a_{\ell'} + r_{N,k}(t)} b_{\ell} - \frac{\sum_{\ell'=1}^{kN^{1/8}} \min(\frac{\ell'}{k}, 2c)b_{\ell'} + \tilde{r}_{N,k}(t)}{\sum_{\ell'=1}^{kN^{1/8}} b_{\ell'} + r_{N,k}(t)} b_{\ell} \right| \leq 24ce^{8c}\|\mathbf{a} - \mathbf{b}\|_1 b_{\ell}$$

so that:

$$\begin{aligned} (a) & \leq 24ce^{8c}\|\mathbf{a} - \mathbf{b}\|_1 \underbrace{\sum_{\ell} b_{\ell}}_{\leq 2} + \underbrace{\frac{\tilde{r}_{N,k}(t)}{\sum_{\ell'=1}^{kN^{1/8}} a_{\ell'} + r_{N,k}(t)}}_{\leq 2c} \|\mathbf{a} - \mathbf{b}\|_1 \\ & + \sum_{\ell} \frac{1}{\sum_{\ell'=1}^{kN^{1/8}} a_{\ell'} + r_{N,k}(t)} \underbrace{\left| \sum_{\ell'=1}^{kN^{1/8}} \min(\frac{\ell'}{k}, 2c)a_{\ell'} a_{\ell} - \sum_{\ell'=1}^{kN^{1/8}} \min(\frac{\ell'}{k}, 2c)b_{\ell'} b_{\ell} \right|}_{(ii)} \end{aligned}$$

We now bound (ii).

$$(ii) \leq \left| \sum_{\ell'=1}^{kN^{1/8}} \min(\frac{\ell'}{k}, 2c)a_{\ell'} a_{\ell} - \sum_{\ell'=1}^{kN^{1/8}} \min(\frac{\ell'}{k}, 2c)a_{\ell'} b_{\ell} \right| + \left| \sum_{\ell'=1}^{kN^{1/8}} \min(\frac{\ell'}{k}, 2c)a_{\ell'} b_{\ell} - \sum_{\ell'=1}^{kN^{1/8}} \min(\frac{\ell'}{k}, 2c)b_{\ell'} b_{\ell} \right|$$

$$\leq 2c \sum_{\ell'=1}^{kN^{1/8}} a_{\ell'} |a_{\ell} - b_{\ell}| + 2cb_{\ell} \|\mathbf{a} - \mathbf{b}\|_1$$

so

$$\begin{aligned} (a) &\leq 2c(24e^{8c} + 1) \|\mathbf{a} - \mathbf{b}\|_1 + 2c \sum_{\ell} \frac{\sum_{\ell'=1}^{kN^{1/8}} a_{\ell'} |a_{\ell} - b_{\ell}| + b_{\ell} \|\mathbf{a} - \mathbf{b}\|_1}{\sum_{\ell'=1}^{kN^{1/8}} a_{\ell'} + r_{N,k}(t)} \\ &\leq 2c(24e^{8c} + 4e^{4c} + 2) \|\mathbf{a} - \mathbf{b}\|_1. \end{aligned}$$

Bound of (b) By Young's inequality, it holds that $\sum_{\ell} |\sum_{\ell'=0}^{\ell} \min(\frac{\ell'}{k}, 2c) b_{\ell'} b_{\ell-\ell'}| \leq 8c$, so we have by equation 16:

$$\sum_{\ell} \left| \frac{\sum_{\ell'=0}^{\ell} \min(\frac{\ell'}{k}, 2c) b_{\ell'} b_{\ell-\ell'}}{\sum_{\ell'=1}^{kN^{1/8}} a_{\ell'} + r_{N,k}(t)} - \frac{\sum_{\ell'=0}^{\ell} \min(\frac{\ell'}{k}, 2c) b_{\ell'} b_{\ell-\ell'}}{\sum_{\ell'=1}^{kN^{1/8}} b_{\ell'} + r_{N,k}(t)} \right| \leq 32ce^{8c} \|\mathbf{a} - \mathbf{b}\|_1$$

This implies:

$$\begin{aligned} (b) &\leq 32ce^{8c} \|\mathbf{a} - \mathbf{b}\|_1 + \sum_{\ell} \left| \frac{\sum_{\ell'=0}^{\ell} \min(\frac{\ell'}{k}, 2c) a_{\ell'} a_{\ell-\ell'}}{\sum_{\ell'=1}^{kN^{1/8}} a_{\ell'} + r_{N,k}(t)} - \frac{\sum_{\ell'=0}^{\ell} \min(\frac{\ell'}{k}, 2c) b_{\ell'} b_{\ell-\ell'}}{\sum_{\ell'=1}^{kN^{1/8}} a_{\ell'} + r_{N,k}(t)} \right| \\ &\leq 32ce^{8c} \|\mathbf{a} - \mathbf{b}\|_1 + 2e^{4c} \underbrace{\sum_{\ell} \left| \sum_{\ell'=0}^{\ell} \min(\frac{\ell'}{k}, 2c) a_{\ell'} a_{\ell-\ell'} - \sum_{\ell'=0}^{\ell} \min(\frac{\ell'}{k}, 2c) b_{\ell'} b_{\ell-\ell'} \right|}_{(iii)}. \end{aligned}$$

We have:

$$\begin{aligned} (iii) &\leq \sum_{\ell} \left| \sum_{\ell'=0}^{\ell} \min(\frac{\ell'}{k}, 2c) a_{\ell'} (a_{\ell-\ell'} - b_{\ell-\ell'}) \right| + \sum_{\ell} \left| \sum_{\ell'=0}^{\ell} \min(\frac{\ell'}{k}, 2c) b_{\ell-\ell'} (a_{\ell'} - b_{\ell'}) \right| \\ &\leq 2c (\|\mathbf{a}\|_1 + \|\mathbf{b}\|_1) \|\mathbf{a} - \mathbf{b}\|_1. \end{aligned}$$

So:

$$(b) \leq 48ce^{8c} \|\mathbf{a} - \mathbf{b}\|_1.$$

□

Proof of Lemma 8 Define

$$\mathcal{L}_c := \{g \in L_1 \text{ s.t. } e^{-4c} \leq \|g\|_1 \leq 1 \text{ and } g \geq 0\}.$$

By Lemmas 21 and 22, $\forall t \in [0, 1]$, $f_k(\cdot, t), f(\cdot, t) \in \mathcal{L}_{\eta}$. Define application $A : \mathcal{L}_c \rightarrow \mathcal{L}_c$ as

$$\begin{aligned} A(f)(x) &= -\min(x, 2c) f(x, t) - \frac{1}{\int_0^{+\infty} f(x', t) dx'} \int_0^{+\infty} \min(x', 2c) f(x', t) dx' f(x, t) \\ &\quad + \frac{1}{\int_0^{+\infty} f(x', t) dx'} \int_0^x \min(x', 2c) f(x', t) f(x - x', t) dx'. \end{aligned}$$

Let us show it is Lipschitz with respect to the L_1 norm, and derive the Lipschitz constant. First, for any two functions $f_1, f_2 \in \mathcal{L}_c^2$,

$$\left| \frac{1}{\int_0^{+\infty} f_1(x, t) dx} - \frac{1}{\int_0^{+\infty} f_2(x, t) dx} \right| \leq e^{4c} \|f_1 - f_2\|_1.$$

Also, noting $\tilde{f}$ the function $\min(x, 2c)f(x)$, we have $\|\tilde{f}_1 - \tilde{f}_2\|_1 \leq 2c \|f_1 - f_2\|_1$, and

$$\begin{aligned} \|\tilde{f}_1 * f_1 - \tilde{f}_2 * f_2\|_1 &\leq \|(\tilde{f}_1 - \tilde{f}_2) * f_1\|_1 + \|\tilde{f}_2 * (f_2 - f_1)\|_1 \\ &\leq \|\tilde{f}_1 - \tilde{f}_2\|_1 \|f_1\|_1 + \|\tilde{f}_2\|_1 \|f_2 - f_1\|_1 \\ &\leq 4c \|f_2 - f_1\|_1, \end{aligned}$$

where the second line comes from Young's inequality. Putting everything together, we get:

$$\begin{aligned} \|A(f_1) - A(f_2)\|_1 &\leq \|\tilde{f}_2 - \tilde{f}_1\|_1 + (e^{4c} + 2c) \|f_1 - f_2\|_1 \\ &\quad + e^{4c} \|f_1 - f_2\|_1 \|f_1 * \tilde{f}_1\|_1 + e^{2c} \|\tilde{f}_1 * f_1 - \tilde{f}_2 * f_2\|_1 \\ &\leq (4c + (2c + 1)e^{4c} + 4ce^{2c}) \|\tilde{f}_2 - \tilde{f}_1\|_1. \end{aligned}$$

We note this Lipschitz constant lip_c . Extend the function in Lemma 3 to $\mathbb{R}_+$ as:

$$f_k(x, t) = f_k\left(\frac{\lceil kx \rceil}{k}, t\right),$$

Define application $B_k : \mathcal{L}_c \rightarrow \mathcal{L}_c$

$$\begin{aligned} B_k(f)(x) &= -\min\left(\frac{\lceil kx \rceil}{k}, 2c\right) f(x, t) - \frac{1}{\int_0^{+\infty} f(x', t) dx'} \int_0^{+\infty} \min\left(\frac{\lceil kx' \rceil}{k}, 2c\right) f(x', t) dx' f(x, t) \\ &\quad + \frac{1}{\int_0^{+\infty} f(x', t) dx'} \int_0^{\frac{\lceil kx \rceil}{k}} \min\left(\frac{\lceil kx' \rceil}{k}, 2c\right) f(x', t) f(x - x', t) dx'. \end{aligned}$$

Which is defined s.t. for any $x \in \mathbb{R}_+$, $\frac{\partial f_k(x, t)}{\partial t} = B_k(f_k(\cdot, t))(x)$. For any $t \in [0, 1]$, we have:

$$\begin{aligned} &\|B_{k,2}(f_k(\cdot, t))(x) - A_2(f_k(\cdot, t))(x)\|_1 \\ &\leq \underbrace{\int_0^{2c} \left| \frac{\lceil kx \rceil}{k} - x \right| f_k(x, t) dx}_{\leq \frac{1}{k} \|f_k\|_1 \leq \frac{1}{k}} + \frac{1}{\int_0^{+\infty} f_k(x', t) dx'} \int_0^{2c} \int_x^{\frac{\lceil kx \rceil}{k}} 2c f_k(x', t) \underbrace{f_k(x - x', t)}_{\leq 1} dx' dx \\ &\quad + \frac{1}{\int_0^{+\infty} f_k(x', t) dx'} \underbrace{\int_0^{+\infty} \left| \int_0^x \left(\frac{\lceil kx' \rceil}{k} - x' \right) f_k(x', t) f_k(x - x', t) dx' \right| dx}_{\leq \frac{1}{k} \int_0^{+\infty} \int_0^x f_k(x', t) f_k(x - x', t) dx' dx \leq \frac{1}{k} \|f_k\|_1^2 \leq \frac{1}{k} \|f_k\|_1 \leq \frac{1}{k}} \\ &\leq \frac{2}{k} + \frac{2c}{\int_0^{+\infty} f_k(x', t) dx'} \int_0^\eta \int_x^{x + \frac{1}{k}} f_k(x', t) dx' dx \leq \frac{2c + 2}{k}. \end{aligned}$$

We also have:

$$\|f(\cdot, 0) - f_k(\cdot, 0)\|_1 \leq \frac{2}{k}.$$

Thus, by application of Gronwall's Lemma, for any $t \in [0, 1]$

$$\|f(\cdot, t) - f_k(\cdot, t)\|_1 \leq \frac{2}{k} e^{\text{lip}_c t} + \frac{2c + 2}{k}.$$

□

4.5 Proofs of the technical lemmas for section 3.3.2

Proof of Lemma 12 : By equation 7, it holds that:

$$\left| \frac{\partial \sum_{\ell=1}^{k\eta} g_k(\ell, t)}{\partial t} \right| \leq \sum_{\ell=1}^{k\eta} \left| \Gamma_{k,N}^\ell \left(t, (g_k(\ell, t))_{\ell=1}^{k\eta} \right) \right| \leq \left(2 + \frac{6}{1-t} \right).$$

This implies:

$$\left| \int_{t=0}^{1-\epsilon} \sum_{\ell=1}^{k\eta} \left(\frac{\ell}{k} \right)^2 g_k(\ell, t) dt - \frac{1}{N} \sum_{t=0}^{N(1-\epsilon)-1} \sum_{\ell=1}^{k\eta} \left(\frac{\ell}{k} \right)^2 g_k(\ell, \frac{t}{N}) \right| \leq \eta^2 \int_{t=0}^{1-\epsilon} \left| g_k(\ell, t) dt - g_k \left(\ell, \frac{\lfloor tN \rfloor}{N} \right) \right|, \\ \leq \frac{\eta^2}{N} \left(2 + \frac{6}{1-t} \right). \quad (17)$$

Define domain:

$$\mathcal{L}_\eta := \{g : [0, \eta] \rightarrow [0, 1] \text{ s.t. } \|g\|_1 \leq 1, \|xg\|_1 \leq 1 \text{ and } 1 \geq g \geq 0\}.$$

By Lemmas 21 and 22, $\forall t \in [0, 1]$, $g_k(\cdot, t), g(\cdot, t) \in \mathcal{L}_\eta$.

Define application $A_2 : \mathcal{L}_\eta \rightarrow \mathcal{L}_\eta$ by

$$A_2(g)(x) = - \left(x + \frac{1}{1-t} \right) g(x, t) + \frac{1}{1-t} \int_0^x x' g(x', t) g(x - x', t) dx'.$$

Let us show that A_2 is Lipschitz with respect to the L_1 norm on $\mathcal{L}_\eta$, and derive the Lipschitz constant.

Take any $(g_1, g_2) \in \mathcal{L}_\eta^2$. We have:

$$\begin{aligned} \| (xg_1) * g_1 - (xg_2) * g_2 \|_1 &\leq \| (xg_1 - xg_2) * g_1 \|_1 + \| (xg_2) * (g_1 - g_2) \|_1, \\ &\leq \| xg_1 - xg_2 \|_1 \| g_1 \|_1 + \| xg_2 \|_1 \| g_1 - g_2 \|_1, \\ &\leq 2\eta \| g_1 - g_2 \|_1, \end{aligned}$$

where the second line comes from Young's convolution inequality. We thus get:

$$\| A_2(g_1) - A_2(g_2) \|_1 \leq \left(3\eta + \frac{1}{1-t} \right) \| g_1 - g_2 \|_1 \leq \left(3\eta + \frac{1}{\epsilon} \right) \| g_1 - g_2 \|_1.$$

Extend the function in Lemma 11 $g_k(\ell, t)$ to $x \in \mathbb{R}_+$ as:

$$g_k(x, t) = g_k \left(\frac{\lfloor kx \rfloor}{k}, t \right),$$

Define application $B_{k,2} : \mathcal{L}_\eta \rightarrow \mathcal{L}_\eta$

$$B_{k,2}(g)(x) = - \left(\frac{\lfloor kx \rfloor}{k} + \frac{1}{1-t} \right) g(x, t) + \frac{1}{1-t} \int_0^{\lfloor kx \rfloor / k} \frac{\lfloor kx' \rfloor}{k} g_k(x', t) g(x - x', t) dx'.$$

Which is defined s.t. for any $x \in \mathbb{R}_+$, $\frac{\partial g_k(x, t)}{\partial t} = B_{k,2}(g_k(\cdot, t))(x)$. For any $t \in [0, 1]$, we have:

$$\begin{aligned} &\| B_{k,2}(g_k(\cdot, t))(x) - A_2(g_k(\cdot, t))(x) \|_1 \\ &\leq \underbrace{\int_0^\eta \left| \frac{\lfloor kx \rfloor}{k} - x \right| g_k(x, t) dx}_{\leq \frac{1}{k} \| g_k \|_1 \leq \frac{1}{k}} + \frac{1}{1-t} \int_0^\eta \int_x^{\lfloor kx \rfloor / k} \underbrace{\frac{\lfloor kx' \rfloor}{k} g_k(x', t)}_{\leq \eta+1} \underbrace{g_k(x - x', t)}_{\leq 1} dx' dx \\ &\quad + \frac{1}{1-t} \underbrace{\int_0^\eta \left| \int_0^x \left(\frac{\lfloor kx' \rfloor}{k} - x' \right) g_k(x', t) g_k(x - x', t) dx' \right| dx}_{\leq \frac{1}{k} \int_0^\eta \int_0^x g_k(x', t) g_k(x - x', t) dx' dx \leq \frac{1}{k} \| g_k \|^2 \leq \frac{1}{k} \| g_k \|_1 \leq \frac{1}{k}} \\ &\leq \frac{2}{k} + \frac{\eta+1}{1-t} \int_0^\eta \int_x^{x+\frac{1}{k}} g_k(x', t) dx' dx \leq \frac{3+\eta}{k}. \end{aligned}$$

We also have:

$$\| g(\cdot, 0) - g_k(\cdot, 0) \|_1 \leq \frac{2}{k}.$$

Thus, by application of Gronwall's Lemma, for any $t \in [0, 1]$

$$\int_0^\eta |f(x, t) - f_k(x, t)| dx \leq \frac{2}{k} e^{3\eta + \frac{1}{\epsilon}} + \frac{3 + \eta}{k}.$$

□

Proof of lemma 13 : We have:

$$\frac{d}{dt} \int_0^{+\infty} \frac{x^2}{4} g(x, t) dx = - \int_0^{+\infty} \frac{x^2}{4} \left(x + \frac{1}{1-t} \right) g(x, t) dx + \frac{1}{1-t} \underbrace{\int_0^{+\infty} \frac{x^2}{4} \int_0^x x' g(x', t) g(x - x', t) dx'}_{(a)}.$$

Let us simplify (a):

$$\begin{aligned} (a) &= \int_0^{+\infty} \int_0^x x' \frac{(x - x')^2}{4} g(x', t) g(x - x', t) dx' - \int_0^{+\infty} \int_0^x \frac{(x')^3}{4} g(x', t) g(x - x', t) dx' \\ &\quad + \int_0^{+\infty} \int_0^x \frac{(x')^2}{2} x g(x', t) g(x - x', t) dx', \\ &= \int_0^{+\infty} \frac{x^2}{4} g(x, t) dx' \underbrace{\int_0^{+\infty} x g(x, t) dx'}_{=1} - \int_0^{+\infty} \int_0^x \frac{(x')^3}{4} g(x', t) g(x - x', t) dx' \\ &\quad + \int_0^{+\infty} \int_0^x \frac{(x')^2}{2} (x - x') g(x', t) g(x - x', t) dx' + \int_0^{+\infty} \int_0^x \frac{(x')^3}{2} g(x', t) g(x - x', t) dx', \\ &= 3 \int_0^{+\infty} \frac{x^2}{4} g(x, t) dx + \int_0^{+\infty} \frac{x^3}{4} g(x, t) dx \underbrace{\int_0^{+\infty} g(x, t) dx}_{=1-t}. \end{aligned}$$

Reinjecting in the previous equation we obtain:

$$\frac{d}{dt} \int_0^{+\infty} \frac{x^2}{4} g(x, t) dx = \frac{2}{1-t} \int_0^{+\infty} \frac{x^2}{4} g(x, t) dx.$$

Define $z(t) = \int_0^{+\infty} \frac{x^2}{4} g(x, t) dx$. Note that $z(0) = 0.5$. We get:

$$\frac{z'(t)}{z(t)} = \frac{2}{1-t},$$

which integrates to $\ln(z(t)) = -2 \ln(1-t) + \ln(z(0))$. Thus $z(t) = \frac{z(0)}{(1-t)^2}$, and the total length of the matching created is:

$$\int_0^t z(t) dt = z(0) \left[\frac{1}{1-t} - 1 \right].$$

□

References

- Anders Aamand, Justin Chen, and Piotr Indyk. (optimal) online bipartite matching with degree information. *Advances in Neural Information Processing Systems*, 35:5724–5737, 2022.
- Mohammad Akbarpour, Yeganeh Alimohammadi, Shengwu Li, and Amin Saberi. The value of excess supply in spatial matching markets, 2021. URL <https://arxiv.org/abs/2104.03219>.
- Eric Balkanski, Yuri Faenza, and Noemie Perivier. The power of greedy for online minimum cost matching on the line, 2022.

- Benjamin Birnbaum and Claire Mathieu. On-line bipartite matching made simple. *SIGACT News*, 39(1):80–87, March 2008. ISSN 0163-5700. doi: 10.1145/1360443.1360462. URL <https://doi.org/10.1145/1360443.1360462>.
- Bartłomiej Blaszczyński. Lecture Notes on Random Geometric Models — Random Graphs, Point Processes and Stochastic Geometry. Lecture, December 2017. URL <https://hal.inria.fr/cel-01654766>.
- Allan Borodin, Christodoulos Karavasilis, and Denis Pankratov. An experimental study of algorithms for online bipartite matching, 2018.
- Brian Brubach, Karthik Abinav Sankararaman, Aravind Srinivasan, and Pan Xu. Online stochastic matching: New algorithms and bounds, 2019.
- Nikhil Devanur, Kamal Jain, and Robert Kleinberg. Randomized primal-dual analysis of ranking for online bipartite matching. In *Proc. SODA '13*, 01 2013. doi: 10.1137/1.9781611973105.7.
- Nathanaël Enriquez, Gabriel Faraud, Laurent Ménard, and Nathan Noiry. Depth first exploration of a configuration model. *arXiv preprint arXiv:1911.10083*, 2019.
- Alan Frieze, Colin McDiarmid, and Bruce Reed. Greedy matching on the line. *SIAM Journal on Computing*, 19(4):666–672, 1990.
- Gagan Goel and Aranyak Mehta. Online budgeted matching in random input models with applications to adwords. In *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 982–991, 01 2008. doi: 10.1145/1347082.1347189.
- Anupam Gupta, Guru Guruganesh, Binghui Peng, and David Wajc. Stochastic online metric matching, 2019. URL <https://arxiv.org/abs/1904.09284>.
- Thomas P Hayes. A large-deviation inequality for vector-valued martingales. *Combinatorics, Probability and Computing*, 2005.
- Zhiyi Huang, Xinkai Shu, and Shuyi Yan. The power of multiple choices in online stochastic matching. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2022, page 91–103, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450392648. doi: 10.1145/3519935.3520046. URL <https://doi.org/10.1145/3519935.3520046>.
- Patrick Jaillet and Xin Lu. Online stochastic matching: New algorithms with better bounds. *Mathematics of Operations Research*, 39(3):624–646, 2014.
- R. M. Karp, U. V. Vazirani, and V. V. Vazirani. An optimal algorithm for on-line bipartite matching. In *Proceedings of the Twenty-Second Annual ACM Symposium on Theory of Computing*, STOC '90, page 352–358, New York, NY, USA, 1990. Association for Computing Machinery. ISBN 0897913612. doi: 10.1145/100216.100262. URL <https://doi.org/10.1145/100216.100262>.
- Mohammad Mahdian and Qiqi Yan. Online bipartite matching with random arrivals: An approach based on strongly factor-revealing lps. In *Proceedings of the Annual ACM Symposium on Theory of Computing*, pages 597–606, 01 2011. doi: 10.1145/1993636.1993716.
- Vahideh H Manshadi, Shayan Oveis Gharan, and Amin Saberi. Online stochastic matching: Online actions based on offline statistics. *Mathematics of Operations Research*, 37(4):559–573, 2012.
- Andrew Mastin and Patrick Jaillet. Greedy online bipartite matching on random graphs. *arXiv preprint arXiv:1307.2536*, 2013.
- Aranyak Mehta. Online matching and ad allocation. *Theoretical Computer Science*, 8(4):265–368, 2012.
- Błażej Miasojedow. Hoeffding’s inequalities for geometrically ergodic markov chains on general state space. *Statistics & Probability Letters*, 87:115–120, 2014.

Nathan Noiry, Flore Sentenac, and Vianney Perchet. Online matching in sparse random graphs: Non-asymptotic performances of greedy algorithm. In *NeurIPS*, 2021.

Lutz Warnke. On wormald’s differential equation method, 2019.

Nicholas C Wormald. Differential equations for random processes and random graphs. *The annals of applied probability*, 5(4):1217–1235, 1995.

Nicholas C Wormald et al. The differential equation method for random graph processes and greedy algorithms. *Lectures on approximation and randomized algorithms*, 73(155):0943–05073, 1999.

A Poisson Point Processes

The following definitions and properties of point processes come from lecture notes Blaszczyzyn [2017], and are reported here for clarity.

Definition 1 (*Homogeneous Poisson point process*). A point process Φ on $[0, 1]$ is an homogeneous Poisson point process of intensity λ if the following two conditions are satisfied:

1. For any $(a, b] \in [0, 1]$, $\Phi(a, b]$, the number of points in interval $(a, b]$, is a Poisson random variable of intensity $\lambda(b - a)$, i.e.;

$$P\{\Phi(a, b] = n\} = \frac{[\lambda(b - a)]^n}{n!} e^{-\lambda(b - a)}.$$

2. The number of points in any two disjoint intervals are independent of each other, and this extends to any finite number of disjoint intervals, i.e.;

$$P\{\Phi(a_i, b_i] = n_i, i = 1, \dots, k\} = \prod_{i=1}^k \frac{[\lambda(b_i - a_i)]^{n_i}}{n_i!} e^{-\lambda(b_i - a_i)},$$

for any integer $k \geq 2$, any $a_1 < b_1 \leq a_2 < b_2 \leq \dots \leq a_k < b_k$.

Note that the second point of the definition implies the following property: given there are n points of the homogeneous Poisson process in the window B , these points are independently and uniformly distributed in B .

Let Φ^N be a Poisson point process of intensity N on $[0, 1]$, and $\mathcal{U}^N \sim \Phi^N$. Let us enumerate the points of the point process $\mathcal{U}^N$ according to their coordinates. The sequence $\{u_k\}$ can be constructed as a renewal process with exponential holding times, i.e., $u_k = \sum_{i=1}^k F_i$ for $k \geq 1$, where $\{F_k : k = 1, \dots\}$ is a sequence of independent, identically distributed exponential random variables of parameter N . Indeed,

$$\mathbb{P}\{F_1 > t\} = \mathbb{P}\{u_1 > t\} = \mathbb{P}\{\Phi((0, t]) = 0\} = e^{-Nt},$$

and, for $k \geq 2$ by independence (second point of the definition),

$$\begin{aligned} \mathbb{P}\{F_k > t \mid F_1, \dots, F_{k-1}\} &= \mathbb{P}\{u_k - u_{k-1} > t \mid u_1, \dots, u_{k-1}\} \\ &= \mathbb{P}\{\Phi(u_{k-1}, u_{k-1} + t] = 0 \mid u_{k-1}\} \\ &= e^{-Nt}. \end{aligned}$$

B A generalized version of Wormald’s theorem

The following theorem is a generalized version of Wormald’s theorem [Wormald et al., 1999]. The main difference with the original theorem is that we apply the concentration inequality to the whole vector rather than coordinate by coordinate, that is Lemma 23 rather than the classical Azuma-Hoeffding inequality. We repeat here the whole proof for completeness. It is largely based on the one in Warnke [2019].

Lemma 23 (Hayes [2005]) Let $\mathbf{X}$ be a real-value martingale taking values in $\mathbb{R}^d$ s.t. $X_0 = 0$ and $\|X_i - X_{i-1}\| \leq 1$. Then for every $t > 0$

$$\mathbb{P}(\|X_m\| \geq a) < 2e^2 e^{-a^2/2m}.$$

Theorem 4 Take $a(n)$, $n \geq 1$, a collection of variables $(Y_k(i))_{1 \leq k \leq a(n)}$, a bounded domain $\mathcal{D} \subseteq \mathbb{R}^{a(n)+1}$, and a function $F : \mathcal{D} \rightarrow \mathbb{R}^{a(n)}$ that is L -Lip with respect to the L_1 norm on $\mathcal{D}$ and each coordinate is bounded by R on $\mathcal{D}$. Assume the random variables $(Y_k(i))_{1 \leq k \leq a(n)}$ are $\mathcal{F}_i$ -measurable and the following holds when $\left(\frac{i}{n}, \frac{Y_1(i)}{n}, \dots, \frac{Y_{a(n)}(i)}{n}\right) \in \mathcal{D}$:

- (Trend, i) $\sum_{k \leq a(n)} \left| \mathbb{E}[Y_k(i+1) - Y_k(i) | \mathcal{F}_i] - F_k\left(\frac{i}{n}, \frac{Y_1(i)}{n}, \dots, \frac{Y_{a(n)}(i)}{n}\right) \right| \leq \delta(n)$,
- (Bounded jumps, ii) $\sum_{k \leq a(n)} |Y_k(i+1) - Y_k(i)| \leq \beta$,
- (Initial condition, iii) $\sum_{k \leq a(n)} |Y_k(0) - ny_k(0)| \leq \lambda(n)n$.

Then there is $T = T(\mathcal{D}) \in (0, \infty)$ such that, whenever $\lambda(n) \geq (\delta(n) + RL a(n)/n) \min\{T, L^{-1}\}$, with probability at least $1 - 2Tne^2 e^{-n\lambda(n)^2/(8T\beta^2)}$ we have

$$\max_{0 \leq i \leq \sigma n} \sum_{1 \leq k \leq a(n)} \left| Y_k(i) - y_k\left(\frac{i}{n}\right) n \right| < 3e^{LT} \lambda(n)n$$

where $(y_k(t))_{1 \leq k \leq a(n)}$ is the unique solution to the system of differential equations $y'_k(t) = F_k(t, y_1(t), \dots, y_{a(n)}(t))$ with $y_k(0)$ defined in (iii) $\forall 1 \leq k \leq a(n)$, and $\sigma = \sigma(y_1(0), \dots, y_{a(n)}(0)) \in [0, T]$ is any choice of $\sigma \geq 0$ with the property that $(t, y_1(t), \dots, y_{a(n)}(t))$ has L_1 -distance at least $3e^{LT} \lambda(n)$ from the boundary of $\mathcal{D}$ for all $t \in [0, \sigma]$.

Proof: Define $I_{\mathcal{D},n}$ as the minimum of $\lfloor Tn \rfloor$ and the smallest integer $i \geq 0$ where $(i/n, Y_1(i)/n, \dots, Y_{a(n)}(i)/n) \notin \mathcal{D}$ holds. Set $\Delta Y_k(i) := \mathbb{1}_{\{i < I_{\mathcal{D}}\}} [Y_k(i+1) - Y_k(i)]$ and

$$M_k(j) := \sum_{0 \leq i < j} [\Delta Y_k(i) - \mathbb{E}(\Delta Y_k(i) | \mathcal{F}_i)].$$

Since the event $\{i < I_{\mathcal{D}}\}$ is $\mathcal{F}_i$ -measurable (determined by all information of the first i steps), we have

$$Y_k(j) = M_k(j) + Y_k(0) + \sum_{0 \leq i < j} \mathbb{E}(Y_k(i+1) - Y_k(i) | \mathcal{F}_i) \quad \text{for all } 0 \leq j \leq I_{\mathcal{D}}.$$

Furthermore, for all $i \geq 0$, the 'tower property' of conditional expectations implies

$$\mathbb{E}(M_k(i+1) - M_k(i) | \mathcal{F}_i) = \mathbb{1}_{\{i < I_{\mathcal{D}}\}} \mathbb{E}(Y_k(i+1) - Y_k(i) - \mathbb{E}(Y_k(i+1) - Y_k(i) | \mathcal{F}_i) | \mathcal{F}_i) = 0.$$

The 'Boundedness hypothesis' (ii) implies

$$\sum_{k \leq n(a)} |M_k(i+1) - M_k(i)| = \sum_{k \leq n(a)} |\Delta Y_k(i+1) - \mathbb{E}(\Delta Y_k(i+1) | \mathcal{F}_i)| \leq 2\beta.$$

Defining $\mathcal{M}$ as the event

$$\max_{0 \leq j \leq I_{\mathcal{D}}} \sum_{k \leq n(a)} |M_k(j)| < \lambda(n)n,$$

the vector Azuma-Hoeffding inequality (Lemma 23 with $m := \lfloor Tn \rfloor$) and a union bound thus yields $\mathbb{P}(\neg \mathcal{M}) \leq 2Tne^2 e^{-n\lambda(n)^2/(8T\beta^2)}$.

The final deterministic part of the argument is based on a discrete variant of Gronwall's inequality (and induction).

Assuming that the event $\mathcal{M}$ holds, for all $(0, y_1(0), \dots, y_{a(n)}(0)) \in \mathcal{D}$ satisfying $\sum_{1 \leq k \leq a(n)} |Y_k(0) - y_k(0)n| \leq \lambda(n)n$ it remains to prove by induction that, for all integers $0 \leq m \leq \sigma n$, we have

$$\sum_{1 \leq k \leq a(n)} \left| Y_k(m) - y_k\left(\frac{m}{n}\right)n \right| < 3\lambda(n)ne^{LT}.$$

The base case $m = 0$ holds since $\sum_{1 \leq k \leq a(n)} |Y_k(0) - y_k(0)n| \leq \lambda(n)n$ by assumption.

Turning to the induction step $1 \leq m \leq \sigma n$, note that $m - 1 < \lfloor \sigma n \rfloor \leq \lfloor Tn \rfloor$ by definition. So, by choice of σ , the induction hypothesis implies $m - 1 < I_{\mathcal{D}}$ and thus $m \leq I_{\mathcal{D}}$.

Fix $0 \leq j \leq m$. We have:

$$\begin{aligned} \left| Y_k(j) - y_k\left(\frac{j}{n}\right)n \right| &\leq |M_k(j)| + |Y_k(0) - y_k(0)n| \\ &\quad + \sum_{0 \leq i < j} \left| \mathbb{E}(Y_k(i+1) - Y_k(i) \mid \mathcal{F}_i) - \left[y_k\left(\frac{i+1}{n}\right) - y_k\left(\frac{i}{n}\right) \right] n \right| \end{aligned}$$

Under event $\mathcal{M}$ and (iii), we have:

$$\sum_{k \leq n(a)} |M_k(j)| + |Y_k(0) - y_k(0)n| \leq 2\lambda(n)n.$$

On the other hand:

$$\begin{aligned} \sum_{k \leq a(n)} \sum_{0 \leq i < j} \left| \mathbb{E}(Y_k(i+1) - Y_k(i) \mid \mathcal{F}_i) - \left[y_k\left(\frac{i+1}{n}\right) - y_k\left(\frac{i}{n}\right) \right] n \right| \\ \leq \delta(n) + \sum_{k \leq a(n)} \sum_{0 \leq i < j} \sup_{\varepsilon \in [\frac{i}{n}, \frac{i+1}{n}]} \left| F_k(i/n, \tilde{Y}_i/n) - y'_k(\varepsilon) \right| \end{aligned}$$

For any $\varepsilon \in [\frac{i}{n}, \frac{i+1}{n}]$, we have:

$$|y'_k(\varepsilon) - y'_k(\frac{i}{n})| = |F_k(\varepsilon, \tilde{y}(\varepsilon)) - F_k(\frac{i}{n}, \tilde{y}(\frac{i}{n}))| \leq \frac{LR}{n}.$$

It follows that:

$$\begin{aligned} \sum_{k \leq a(n)} \left| Y_k(j) - y_k\left(\frac{j}{n}\right)n \right| &\leq \sum_{0 \leq i < j} \left(LR \frac{a(n)}{n} + \delta(n) + \sum_{k \leq a(n)} \left| F_k(i/n, \tilde{Y}_i/n) - F_k(i/n, y(i/n)) \right| \right) \\ &\quad + 2\lambda(n)n \\ &\leq \sum_{0 \leq i < j} \left(LR \frac{a(n)}{n} + \delta(n) + \frac{L}{n} \sum_{k \leq a(n)} |Y_k(i/n) - y_k(i/n)n| \right) \\ &\quad + 2\lambda(n)n. \end{aligned}$$

We now use the following discrete version of Gronwall's Lemma:

Lemma 24 Assume that there are $b, c \geq 0$ and $a > 0$ such that $x_j < c + \sum_{0 \leq i < j} (ax_i + b)$ for all $0 \leq j \leq m$. Then $x_m < (c + b \min\{m, a^{-1}\}) e^{am}$.

It yields:

$$\sum_{k \leq a(n)} \left| Y_k(m) - y_k\left(\frac{m}{n}\right)n \right| \leq \left(2\lambda(n)n + \left(LR \frac{a(n)}{n} + \delta(n) \right) \min(m, \frac{n}{L}) \right) e^{\frac{Lm}{n}}$$

$$\leq 3\lambda(n)ne^{LT}.$$

□

C Proof of Lemma 10

Proof: Consider any interval $[\frac{a}{kN}; \frac{b}{kN}]$. Denote z the length of that interval. We have $z = \frac{b}{kN} - \frac{a}{kN}$ if $b > a$, else $z = \frac{b}{kN} + 1 - \frac{1}{kN}$. Assume

$$z \geq \frac{4\epsilon - \epsilon^2}{16N}.$$

The number of vertices from the online side that have fallen in this interval before y_t , $|\{y_i \in [\frac{a}{kN}; \frac{b}{kN}] | i < t\}|$ follows a binomial distribution $\mathcal{B}(t-1, z)$. By Chernoff bound:

$$\begin{aligned} \mathbb{P}\left(\left|\left\{y_i \in \left[\frac{a}{kN}; \frac{b}{kN}\right] \mid i < t\right\}\right| \geq \left(1 + \frac{\epsilon}{4}\right)(t-1)z\right) &\leq \mathbb{P}\left(\left|\left\{y_i \in \left[\frac{a}{kN}; \frac{b}{kN}\right] \mid i \neq t\right\}\right| \geq \left(1 + \frac{\epsilon}{4}\right)(N(1-\epsilon) - 1)z\right) \\ &\leq e^{-\frac{\epsilon^2}{48}(N(1-\epsilon)-1)z} \leq e^{-\frac{\epsilon^2}{50}N(1-\epsilon)z} \end{aligned}$$

The last inequality holds as $N \geq \frac{25}{1-\epsilon}$. The number of vertices from the offline side, $|\{u \in [\frac{a}{kN}; \frac{b}{kN}] | u \in \tilde{\mathcal{U}}\}|$ follows a binomial distribution $\mathcal{B}(Nkz + 1, p_k)$.

$$\mathbb{P}\left(\left|\left\{u \in \left[\frac{a}{kN}; \frac{b}{kN}\right] \mid u \in \tilde{\mathcal{U}}\right\}\right| \leq \left(1 - \frac{\epsilon}{4}\right)(Nkz - 1)p_k\right) \leq e^{-\frac{\epsilon^2}{32}(Nkz-1)p_k} \leq e^{-\frac{\epsilon^2}{50}N(1-\epsilon)z}.$$

We have:

$$\begin{aligned} \left(1 - \frac{\epsilon}{4}\right)(Nkz - 1)p_k - \left(1 + \frac{\epsilon}{4}\right)(N(1-\epsilon) - 1)z &\geq zN \left[\underbrace{\left(1 - \frac{\epsilon}{4}\right)kp_k}_{\geq 1 - \frac{\epsilon}{2}} - \left(1 + \frac{\epsilon}{4}\right)(1-\epsilon) \right] - 1, \\ &\geq zN \left[\frac{\epsilon}{2} - \frac{\epsilon}{4}\left(1 + \frac{\epsilon}{2}\right) \right] - 1, \\ &\geq zN \left[\frac{4\epsilon - \epsilon^2}{4} \right] - 1 \geq 0. \end{aligned}$$

This implies that if $\left|\left\{u \in [\frac{a}{kN}; \frac{b}{kN}] \mid u \in \tilde{\mathcal{U}}\right\}\right| \geq \left(1 - \frac{\epsilon}{4}\right)(Nkz - 1)p_k$ and $\left|\left\{y_i \in [\frac{a}{kN}; \frac{b}{kN}] \mid i < t\right\}\right| \leq \left(1 + \frac{\epsilon}{4}\right)(t-1)z$, then:

$$\left|\left\{y_i \in \left[\frac{a}{kN}; \frac{b}{kN}\right] \mid i < t\right\}\right| \leq \left|\left\{u \in \left[\frac{a}{kN}; \frac{b}{kN}\right] \mid u \in \tilde{\mathcal{U}}\right\}\right|.$$

Let us now bound $\mathbb{P}(c_t \geq z|y_t)$. Denote $\frac{a^L}{kN}$ and $\frac{b^H}{kN}$ the positions of the two free vertices closest to y_t upon arrival. Note that $c_t \geq \eta$ can only hold if $\min(|\frac{a^L}{kN} - y_t|; |\frac{b^H}{kN} - y_t|) \geq \eta$ and all online vertices that arrived before y_t in the interval $[\frac{a^L}{kN}; \frac{b^H}{kN}]$ have been matched strictly inside that interval.

$$\mathbb{P}(c_t \geq \eta|y_t) \leq \mathbb{P}\left(\bigcup_{\substack{y_t \in [\frac{a^L}{kN}; \frac{b^H}{kN}] \\ \min(|\frac{a^L}{kN}; y_t|; |y_t, \frac{b^H}{kN}|) \geq \eta}} \left\{\left|\left\{u \in \left[\frac{a}{kN}; \frac{b}{kN}\right] \mid u \in \tilde{\mathcal{U}}\right\}\right| \leq \left|\left\{y_i \in \left[\frac{a}{kN}; \frac{b}{kN}\right] \mid i < t\right\}\right|\right\} \mid y_t\right)$$

$$\begin{aligned}
&\leq \sum_{\ell=\eta kN}^{kN} \sum_{\substack{\min(|[\frac{a}{kN}; y_t]; [y_t, \frac{b}{kN}]|) \geq \eta \\ |[\frac{a}{kN}; \frac{b}{kN}]| = \frac{\ell}{kN}}} \mathbb{P} \left(\left| \left\{ u \in [\frac{a}{kN}; \frac{b}{kN}] \mid u \in \tilde{\mathcal{U}} \right\} \right| \leq \left| \left\{ y_i \in \left[\frac{a}{kN}; \frac{b}{kN} \right] \mid i < t \right\} \right| \right) \\
&\leq \sum_{\ell=\eta kN}^{kN} \sum_{\substack{\min(|[\frac{a}{kN}; y_t]; [y_t, \frac{b}{kN}]|) \geq \eta \\ |[\frac{a}{kN}; \frac{b}{kN}]| = \frac{\ell}{kN}}} 2e^{-\frac{\epsilon^2}{50} N(1-\epsilon) \frac{\ell}{kN}} \\
&\leq e^{-C'_\epsilon N \eta} \sum_{\ell=1}^{kN} \sum_{\substack{\min(|[\frac{a}{kN}; y_t]; [y_t, \frac{b}{kN}]|) \geq \eta \\ |[\frac{a}{kN}; \frac{b}{kN}]| = \frac{\ell}{kN} + \eta}} 2e^{-\frac{\epsilon^2}{50} N(1-\epsilon) \frac{\ell}{kN}} \\
&\leq e^{-C'_\epsilon N \eta} \underbrace{\sum_{\ell=1}^{kN} 2\ell e^{-\frac{\epsilon^2}{50} N(1-\epsilon) \frac{\ell}{kN}}}_{C''_\epsilon}.
\end{aligned}$$

We have:

$$\begin{aligned}
\mathbb{E} \left[\sum_{t \leq N(1-\epsilon)} c_t[k] \mathbb{1}_{\{c_t[k] \geq \frac{\eta}{N}\}} \right] &\leq \sum_{t=1}^N \int_{z=\eta}^1 \mathbb{P}(c_t[k] \geq z) dz, \\
&\leq \sum_{t=1}^N \int_{z=\frac{\eta}{N}}^1 C''_\epsilon e^{-zNC'_\epsilon} dz, \\
&\leq \frac{C''_\epsilon}{C'_\epsilon} e^{-\eta C'_\epsilon}.
\end{aligned}$$

□