
Online Matching in Sparse Random Graphs: Non-Asymptotic Performances of Greedy Algorithm

Nathan Noiry
Télécom Paris, Palaiseau, France
noiry.nathan@gmail.com

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

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

Abstract

Motivated by sequential budgeted allocation problems, we investigate online matching problems where connections between vertices are not i.i.d., but they have fixed degree distributions – the so-called configuration model. We estimate the competitive ratio of the simplest algorithm, GREEDY, by approximating some relevant stochastic discrete processes by their continuous counterparts, which are solutions of an explicit system of partial differential equations. This technique gives precise bounds on the estimation errors, with arbitrarily high probability as the problem size increases. In particular, it allows the formal comparison between different configuration models. We also prove that, quite surprisingly, GREEDY can have better performance guarantees than RANKING, another celebrated algorithm for online matching that usually outperforms the former.

1 Introduction

Finding matchings in bipartite graphs $(\mathcal{U} \cup \mathcal{V}, \mathcal{E})$, where $\mathcal{E} \subset \mathcal{U} \times \mathcal{V}$ is a set of edges, is a long-standing problem with different motivations and approaches [Godsil, 1981, Zdeborová and Mészard, 2006, Lovász and Plummer, 2009, Bordenave et al., 2013]. If \mathcal{U} is seen as a set of resources and \mathcal{V} as demands, the objective is to allocate as many resources to demands (an allocation - or a matching - between u and v is admissible if $(u, v) \in \mathcal{E}$) with the constraint that a resource is allocated to only one demand and vice-versa.

Motivated particularly by practical applications to Internet advertising, the *online* variant of this problem is receiving increasing attention (we refer to the excellent survey [Mehta, 2012] for more applications, specific settings, results and techniques). In this case, the set of vertices \mathcal{U} is present at the beginning and the graph unveils sequentially: vertices $v \in \mathcal{V}$ are observed sequentially, one after the other, along with the edges they belong to. An online algorithm must decide, right after observing v_k and its associated set of edges $\mathcal{E}_k := \{(u, v_k) \in \mathcal{E}\}$ to match it to some other vertex $u \in \mathcal{U}$, at the conditions that $(u, v_k) \in \mathcal{E}_k$ and $u \in \mathcal{U}$ has not been matched yet. The performance of an online algorithm is evaluated by its *competitive ratio*, which is the ratio between the size of the matching it has created and the highest possible matching in hindsight [Feldman et al., 2009].

This theoretical setting is particularly well suited for online advertising: \mathcal{U} is the set of campaigns/ads that an advertiser can run and users v_1, v_2, \dots, v_T arrive sequentially [Mehta, 2012, Manshadi et al., 2012]. Some of them are eligible for a large subset of campaigns, others are not (usually based

on their attributes/features, such as the geographic localization, the browsing history, or any other relevant information). The objective of an advertiser (in this over-simplified model) is to maximize the number of displayed ads. In practice, campaigns/ads are not displayed only once but have a maximal budget of impressions (say, a specific ad can be displayed only 10.000 times each day). A possible trick consists of duplicating the vertices of \mathcal{U} as many times as the budget. However, this results in strong and undesirable correlations between vertices. It is, therefore, more appropriate to consider a bipartite graph with *capacities* and admissible matchings as subsets of edges such that each vertex belongs to several different edges, but not more than their associated capacities $\omega \in \mathbb{N}$ (a vertex $v \in \mathcal{V}$ is matched once while $u \in \mathcal{U}$ can be matched ω_u times).

This online matching problem with capacities has been quite extensively studied. It is known that GREEDY, which matches all incoming vertices to any available neighbor has a competitive ratio of $1/2$ in the worst case, albeit it achieves $1 - 1/e$ as soon as the incoming vertices arrive in Random Order [Goel and Mehta, 2008b]. The worst-case optimal algorithm is the celebrated RANKING, which achieves $1 - 1/e$ on any instance [Karp et al., 1990, Devanur et al., 2013, Birnbaum and Mathieu, 2008], and also has better guarantees in the Random Order setting [Mahdian and Yan, 2011].

Beyond the adversarial setting, the following stochastic setting has been considered: there exist a finite set of L “base” vertices $v^{(1)}, \dots, v^{(L)}$ associated to base edge-sets $\mathcal{E}^{(1)}, \dots, \mathcal{E}^{(L)}$. When a vertex v_k arrives, its type $\theta_k \in \{1, \dots, L\}$ is drawn iid from some distribution (either known beforehand or not) and then its edge set is set as $\mathcal{E}_k = \mathcal{E}^{(\theta_k)}$. In the context where the distribution is known, algorithms with much better competitive ratios than GREEDY or RANKING were designed [Manshadi et al., 2012, Jaillet and Lu, 2014, Brubach et al., 2019], specifically with a competitive ratio of $1 - 2/e^2$ when the expected number of arrival of each type are integral and 0.706 without this assumption. Notably, those competitive ratios still hold with Poisson arrival rates rather than a fixed number of arrivals.

On a side note, a vast line of work considers online matching in weighted graphs [Devanur et al., 2012, Goel and Mehta, 2008a, Mehta, 2012], which is outside the scope of this paper. However, it is still worth noting that the unweighted graph is a weighted graph with all weights equal.

This model of the stochastic setting is quite interesting but rather strong: it lacks flexibility and cannot be used to represent some challenging instances (for example when the degrees of each vertex \mathcal{U} increase linearly with the number of vertices in \mathcal{V} , or when the set \mathcal{U} of campaigns must be fixed so that the model is well specified, etc...). Another tentative is to consider Erdős-Rényi graphs assuming that each possible edge is present in $\mathcal{U} \times \mathcal{V}$ with some fixed probability and independently of the other edges (see [Mastin and Jaillet, 2013]). The most interesting and challenging setting corresponds to the so-called *sparse* regime where each vertex of \mathcal{U} has an expected degree independent of the size n of \mathcal{V} , which amounts to take a probability of connection equal to c/n . Interestingly enough, even the analysis of the simplest GREEDY algorithm is quite challenging and already insightful in those models [Borodin et al., 2018, Arnosti, 2019, Dyer et al., 1993, Mastin and Jaillet, 2013]. Unfortunately, although this Erdős-Rényi model is compatible with growing sets \mathcal{U} and \mathcal{V} , it also turns out to be quite restrictive. The main reason is that the approximate Poisson degree distribution of the vertices has light-tail and does not allow for the appearance of the so-called *scale-free property* satisfied by many real-world networks [Barabási et al., 2000, Van Der Hofstad, 2016].

We, therefore, consider a more appropriate random graphs generation process called *configuration model*, introduced by [Bender and Canfield, 1978] and [Bollobás, 1980]. The optimal matching of this model has been computed in [Bordenave et al., 2013]. The configuration model is particularly well suited to handle different situations such as the following one. Assume that campaigns can either be “intensive” (with many eligible users) or “selective/light” (few eligible users), with an empirical proportion of, say, 20%/80%. Then whether an advertiser handles 100 campaigns at the same time or 10.000, it will always have roughly this proportion of intensive vs. light campaigns. Similarly, some users are more valuable than others, and are thus eligible for more campaigns than the others; the proportion of each type being independent of the total population size. The configuration model accommodates these observations by basically drawing iid degrees for vertices \mathcal{U} and \mathcal{V} (accordingly to some different unknown distributions for \mathcal{U} and \mathcal{V}) and then by finding a graph such that those degrees distribution are satisfied (up to negligible errors); as a consequence, the graphs generated are *sparse*, in the sense that the number of edges grows linearly with the number of vertices.

Additionally, the configuration model is a well-suited random graph model which mimics a number of properties of real-world complex networks, while being analytically tractable. For instance, choosing

power-law distributions for the degrees allows to obtain the so-called scale-free property (often observed in practice, as highlighted for the web by Faloutsos et al. [1999]). The configuration model also displays the so called “small-world phenomenon” (observed for instance in the graph of Facebook by Backstrom et al. [2012]) as its diameter is of logarithmic order.

Main contribution

We investigate the performances (in terms of expected competitive ratio) of the GREEDY matching algorithm in configuration models and we provide explicit quantitative results using stochastic approximation techniques [Wormald, 1995]; we prove that the increasing size of the random matching created is arbitrarily close to the solution of some explicit ODE. Solving the latter then gives in turn the solution to the original problem.

The remaining of the paper is organized as follows. Section 2 describes precisely the problem and Theorem 1 is our first main result: it describes the performances of GREEDY in the capacity-less problem. The proof of Theorem 1 is delayed to Appendix D, but the main ideas and intuitions are provided in Section 3. The online matching with capacities problem is treated in Appendix A.

2 Online Matching Problems; Models and main result

Consider a bipartite graph with capacities $G = (\mathcal{U}, \mathcal{V}, \mathcal{E}, \omega)$ where $\mathcal{U} = \{1, \dots, N\}$ and $\mathcal{V} = \{1, \dots, T\}$ are two finite set of vertices, $\mathcal{E} \subset \{(u, v), u \in \mathcal{U}, v \in \mathcal{V}\}$ is the set of edges and $\omega : \mathcal{U} \rightarrow \mathbb{N}_*$ is a capacity function. A matching M on G is a subset of edges $e \in \mathcal{E}$ such that any vertex $v \in \mathcal{V}$ is the endpoint of at most one edge $e \in M$ and any vertex $u \in \mathcal{U}$ is the endpoint of at most ω_u edges in M . We will denote by \mathcal{M} the set of matchings on G ; the optimal matching $M^* \in \mathcal{M}$ is the one (or any one) with the highest cardinality, denoted by $|M^*|$.

The batched matching problem consists in finding any optimal matching M^* given a graph with capacities G ; the online variant might be a bit more challenging, as the matching is constructed sequentially. Formally, the set of vertices \mathcal{U} and their capacities ω are known from the start, and vertices $v \in \mathcal{V}$ arrive sequentially (with the edges they belong to) and $M_0 = \emptyset$. At stage $t \in \mathbb{N}$ – assuming a matching M_{t-1} has been constructed –, a decision maker observes a new vertex¹ v_t and its associated set of edges $\{(u, v_t); u \in \mathcal{E}\}$. If possible, one of these edges (u_t, v_t) is added to M_{t-1} , with the constraint that $M_t = M_{t-1} \cup \{(u_t, v_t)\}$ is still a matching. The objective is to maximize the size of the constructed matching M_T . The classical way to evaluate the performances of an algorithm is the *competitive ratio*, defined as $|M_T|/|M^*| \in [0, 1]$ (the higher the better).

2.1 Structured online matching via Configuration Model

As mentioned before, the online matching problem can be quite difficult without additional structure. We will therefore assume that the vertex degrees in \mathcal{U} and \mathcal{V} have (at least asymptotically in N and T) some given subGaussian² distributions $\pi_{\mathcal{U}}$ and $\pi_{\mathcal{V}}$, of respective expectation $\mu_{\mathcal{U}}$ and $\mu_{\mathcal{V}}$ and respective proxy-variance $\sigma_{\mathcal{U}}^2$ and $\sigma_{\mathcal{V}}^2$. Those numbers are related in the sense that we assume³ that $T = \frac{\mu_{\mathcal{U}}}{\mu_{\mathcal{V}}} N \in \mathbb{N}$. Given those degree distributions, the graphs we consider are random draws from a bipartite configuration model described below; for the sake of clarity, we first consider the capacity-less case (when $\omega_u = 1$ for all $u \in \mathcal{U}$).

Given $\pi_{\mathcal{U}}$ and $\pi_{\mathcal{V}}$ and $N, T \geq 1$, let $d_1^{\mathcal{U}}, \dots, d_N^{\mathcal{U}} \in \mathbb{N} \stackrel{\text{i.i.d.}}{\sim} \pi_{\mathcal{U}}$ and $d_1^{\mathcal{V}}, \dots, d_T^{\mathcal{V}} \in \mathbb{N} \stackrel{\text{i.i.d.}}{\sim} \pi_{\mathcal{V}}$ be independent random variables; intuitively, those numbers are respectively the number of half-edges attached to vertex in \mathcal{U} and \mathcal{V} . Consider also two extra random variables

$$d_{T+1}^{\mathcal{V}} = \max \left\{ \sum_{i=1}^N d_i^{\mathcal{U}} - \sum_{j=1}^T d_j^{\mathcal{V}}, 0 \right\} \quad \text{and} \quad d_{N+1}^{\mathcal{U}} = \max \left\{ \sum_{j=1}^T d_j^{\mathcal{V}} - \sum_{i=1}^N d_i^{\mathcal{U}}, 0 \right\}$$

¹Although the order of arrival is irrelevant to the models we studied, it could have an impact on other models.

² X is subGaussian with proxy-variance σ^2 if for any $s \in \mathbb{R}$, $\mathbb{E}[\exp(sX)] \leq \exp\left(\frac{\sigma^2 s^2}{2}\right)$. Actually, we only need that $\pi_{\mathcal{U}}$ and $\pi_{\mathcal{V}}$ have some finite moment of order $\gamma > 2$.

³In the general case, consider $T = \lfloor N\mu_{\mathcal{U}}/\mu_{\mathcal{V}} \rfloor$. The proof is identical, up to a negligible $1/N$ error term

so that equality between total degrees holds, i.e., $\sum_{i=1}^{N+1} d_i^{\mathcal{U}} = \sum_{j=1}^{T+1} d_j^{\mathcal{V}}$. Finally, a random (capacity-less) bipartite graph denoted by $\mathbf{CM}(\mathbf{d}^{\mathcal{U}}, \mathbf{d}^{\mathcal{V}})$ is constructed with a uniform pairing of half-edges of $\mathcal{U} \cup \{N+1\}$ with half-edges of $\mathcal{V} \cup \{T+1\}$ and removing vertices $T+1$ and $N+1$ and their associated edges. These two artificially added vertices are just here to define a pairing between half-edges. Notice that, by the law of large numbers and since $T = (\mu_{\mathcal{U}}/\mu_{\mathcal{V}})N$, $d_{T+1}^{\mathcal{V}} = o(N)$ and $d_{N+1}^{\mathcal{U}} = o(N)$ almost surely⁴.

The bipartite configuration model $\mathbf{CM}(\mathbf{d}^{\mathcal{U}}, \mathbf{d}^{\mathcal{V}})$ is then the random graph obtained by a uniform matching between the half-edges of \mathcal{U} and the half-edges of \mathcal{V} , where the random sequences $\mathbf{d}^{\mathcal{U}} = (d_i^{\mathcal{U}})_i$ and $\mathbf{d}^{\mathcal{V}} = (d_j^{\mathcal{V}})_j$ are defined as above.

2.2 Competitive ratio of GREEDY algorithm. Main result

The first question to investigate in this structured setting is the computation of the (expected) competitive ratio of the simple algorithm GREEDY. It constructs a matching by sequentially adding any admissible edge uniformly at random. Describing it and stating our results require the following additional notations: for any $e = (u, v) \in E$, $u(e) = u$ (resp. $v(e) = v$) is the extremity of e in \mathcal{U} (resp. \mathcal{V}); the generating series of $\pi_{\mathcal{U}}$ and $\pi_{\mathcal{V}}$ are denoted by $\phi_{\mathcal{U}}$ and $\phi_{\mathcal{V}}$ and are defined as

$$\phi_{\mathcal{U}}(s) := \sum_{k \geq 0} \pi_{\mathcal{U}}(k) s^k \quad \text{and} \quad \phi_{\mathcal{V}}(s) := \sum_{k \geq 0} \pi_{\mathcal{V}}(k) s^k.$$

Our first main theorem, stated below, identifies the asymptotic size of the matching generated by GREEDY on the bipartite configuration model we have just defined. As the batched problem (i.e., computing the size of the optimal matching M^*) is well understood [Bordenave et al., 2013], this quantity is sufficient to derive competitive ratios. Again, for the sake of presentation, we first assume that all capacities are fixed, equal to one; the general case is presented in Appendix A.

Theorem 1. (Performances of GREEDY in the capacity-less case)

Given $N \geq 1$ and $T = \frac{\mu_{\mathcal{U}}}{\mu_{\mathcal{V}}} N$, let M_T be the matching built by GREEDY on $\mathbf{CM}(\mathbf{d}^{\mathcal{U}}, \mathbf{d}^{\mathcal{V}})$ then the following convergence in probability holds:

$$\frac{|M_T|}{N} \xrightarrow[N \rightarrow +\infty]{\mathbf{P}} 1 - \phi_{\mathcal{U}}(1 - G(1)).$$

where G is the unique solution of the following ordinary differential equation:

$$G'(s) = \frac{1 - \phi_{\mathcal{V}}\left(1 - \frac{1}{\mu_{\mathcal{U}}} \phi'_{\mathcal{U}}(1 - G(s))\right)}{\frac{\mu_{\mathcal{V}}}{\mu_{\mathcal{U}}} \phi'_{\mathcal{U}}(1 - G(s))}; \quad G(0) = 0. \quad (1)$$

Moreover, for any $s \in [0, 1]$, if $M_T(s)$ is the matching obtained by GREEDY after seeing a proportion s of vertices of \mathcal{V} , then

$$\frac{|M_T(s)|}{N} \xrightarrow[N \rightarrow +\infty]{\mathbf{P}} 1 - \phi_{\mathcal{U}}(1 - G(s)). \quad (2)$$

Convergence rates are explicit; with probability exponentially large, at least $1 - \zeta N \exp(-\xi N^{c/2})$,

$$\sup_{s \in [0, 1]} \left| \frac{|M_T(s)|}{N} - (1 - \phi_{\mathcal{U}}(1 - G(s))) \right| \leq \kappa N^{-c},$$

where ζ, ξ, κ depend only on the (first two) moments of both $\pi_{\mathcal{V}}$ and $\pi_{\mathcal{U}}$, and c is some universal constant (set arbitrarily as $1/20$ in the proof).

Theorem 1 generalizes to the case with capacities, see Sections A.1 and A.2. The details of the proof of Theorem 1 are postponed to Appendix D, but the main ideas are given in Section 3.

2.3 Examples, Instantiations and Corollaries

We provide in this section some interesting examples and corollaries that illustrate the powerfulness of Theorem 1, and how it can be used to compare different situations.

⁴And even $\mathcal{O}(\sqrt{N})$ with probability exponentially large in N as both distributions are sub-Gaussian. So the effects of those additional vertices can be neglected.

2.3.1 d -regular graphs

The first typical example of random graphs are “ d -regular”, for some $d \in \mathbb{N}$, i.e., graphs such that each vertex has an exact degree of d (to avoid trivial examples, we obviously assume $d \geq 2$).

It is non-trivial to sample a d -regular graph at random, yet it is easy to generate random graphs G_N with the configuration model described above, with the specific choices of $\pi_U = \pi_V = \delta_d$, the Dirac mass at d . The downside is that G_N is not exactly a d -regular bipartite random graph (as some vertices might be connected more than once, i.e., there might exist parallel edges). However, conditioned to be *simple*, i.e., without multiple edges and loops, it has the law of a uniform d -regular bipartite random graph. Moreover, the probability of being simple is bounded away from 0 [Van Der Hofstad, 2016]; as a consequence, any property holding with probability tending to 1 for G_N , holds with probability tending to 1 for uniform d -regular bipartite random graphs. Finally, we also mention that Hall’s Theorem [Frieze and Karoński, 2016] implies that G_N admits a perfect matching, so that $|M^*| = N$.

Instantiating Equation (1) to d -regular graphs yields that the competitive ratio of GREEDY converges, with probability 1, to $1 - (1 - G(1))^d$ where G is the solution of the following ODE

$$\frac{(1 - G(s))^{d-1}}{1 - (1 - (1 - G(s))^{d-1})^d} G'(s) = \frac{1}{d}. \quad (3)$$

As expected, had we taken $d = 1$, then $G(s) = s$ hence the competitive ratio of GREEDY is 1 (but again, $d = 1$ -regular graphs are trivial). More interestingly, if $d = 2$, the ODE has a closed form solution: $G(s) = \exp(\frac{s}{2}) - 1$, so that the competitive ratio of GREEDY converges to $4\sqrt{e} - (e + 3) \simeq 0.877 \gg 1 - \frac{1}{e} \simeq 0.632$, where the latter is a standard bound of the competitive ratio of GREEDY (for general, non-regular graphs) [Mehta, 2012].

Solving Equation (3) In the general case $d \geq 3$, even if Equation (3) does not have a closed form solution, it is still possible to provide some insights. Notice first that the polynomial $P(X) = 1 - (1 - (1 - X)^{d-1})^d$ admits $n := d(d - 1)$ roots, among which there is 1 with multiplicity $d - 1$. If X is another root, then

$$(1 - (1 - X)^{d-1})^d = 1 \Leftrightarrow 1 - (1 - X)^{d-1} = e^{\frac{ik\pi}{d}}, \quad k = 1, \dots, d - 1.$$

Therefore,

$$(1 - X)^{d-1} = 1 - e^{\frac{ik\pi}{d}},$$

which admits $d - 1$ distinct solutions for each $k = 1, \dots, d - 1$. The resulting $n := (d - 1)^2$ distinct complex, denoted x_1, \dots, x_n , are the roots of $P(X)/(1 - X)^{d-1}$, so the ODE reduces to:

$$\frac{y'(t)}{\prod_{1 \leq i \leq n} (y(t) - x_i)} = \frac{1}{d}. \quad (4)$$

Since the following trivially holds:

$$\frac{1}{\prod_{1 \leq i \leq n} (X - x_i)} = \sum_{1 \leq i \leq n} \frac{1}{\prod_{j \neq i} (x_i - x_j)} \frac{1}{X - x_i} =: \sum_{1 \leq i \leq n} \frac{a_i}{X - x_i}.$$

it is possible to integrate Equation (4) in $\sum_{1 \leq i \leq n} a_i \log(y(t) - x_i) = \frac{s}{d} + c$ to finally get

$$\prod_{1 \leq i \leq n} (y(t) - x_i)^{a_i} = C \exp\left(\frac{s}{d}\right),$$

and since $y(0) = 0$, it must hold that $C = \prod_{1 \leq i \leq n} (-x_i)^{a_i}$. As a consequence, $y(1)$ solves:

$$\prod_{1 \leq i \leq n} (y(1) - x_i)^{a_i} = e^{1/d} \prod_{1 \leq i \leq n} (-x_i)^{a_i}.$$

Unfortunately, even for $d = 3$, the solution somehow simplifies but has no closed form; on the other hand, numerical computations indicate that the competitive ratio of GREEDY converges to 0.89 when $d = 3$ and N tends to infinity. We provide in Figure 3 the numerical solutions of the ODE for d -regular graphs (actually, we draw the functions $1 - \phi_U(1 - G(s))$ that are more relevant) for various values of d ; the end-point obtained at $s = 1$ indicates the relative performance of GREEDY. As expected, those functions are point-wise increasing with d (as the problem becomes simpler and simpler for GREEDY when $d \geq 2$).

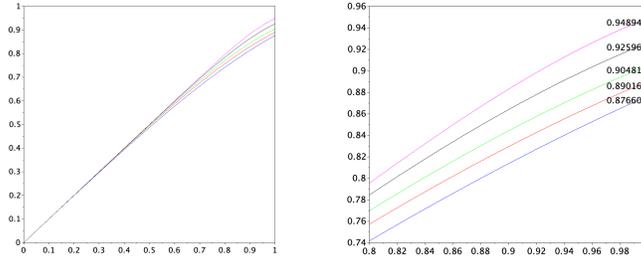


Figure 1: Numerical computations (on Scilab, results are almost instantaneous) of GREEDY performances for $d = 2$ (blue), $d = 3$ (red), $d = 4$ (green), $d = 6$ (black) and $d = 10$ (magenta). On the left, global solution, on the right, zoom-in on the end points with final values.

2.3.2 The Erdős-Rényi case.

In an Erdős-Rényi graph, there is an edge between two vertices $u \in \mathcal{U}$ and $v \in \mathcal{V}$ with some probability $p = \frac{c}{N}$, independently from each others. As N goes to infinity, the number of edges to a vertex follows (approximately) a Poisson law of parameter $c > 1$.

As a consequence, we consider the configuration model where $\pi_{\mathcal{U}}$ and $\pi_{\mathcal{V}}$ are Poisson laws of parameter c , which yields $\mu = c$, $\phi_{\mathcal{U}}(s) = e^{c(s-1)}$. In this case, Equation (1) becomes:

$$\frac{cG'(s)e^{-cG(s)}}{1 - e^{-c}e^{-cG(s)}} = 1.$$

The solutions are given by:

$$G(s) = \frac{1}{c} \log \left(\frac{c}{\log(e^{k-cs} + 1)} \right),$$

yielding

$$\phi_X(1 - G(s)) = \frac{1}{c} \log(e^{k-cs} + 1).$$

The initial condition $\phi_{\mathcal{U}}(1 - G(0)) = \phi_{\mathcal{U}}(1) = 1$ gives $e^k = e^c - 1$, from which we deduce that the number of matches of GREEDY is asymptotically proportional to

$$1 - \phi_{\mathcal{U}}(1 - G(1)) = 1 - \frac{\log(2 - e^{-c})}{c},$$

which recovers, as a sanity check, some existing results [Mastin and Jaillet, 2013].

2.3.3 The comparison of different configuration models

Using Gronwall's Lemma, it is possible to show Theorem 1 can be used to compare different configuration models, as in the following Corollary.

Corollary 1. Consider two configuration models $\text{CM}_1(\mathbf{d}_1^U, \mathbf{d}_1^V)$ and $\text{CM}_2(\mathbf{d}_2^U, \mathbf{d}_2^V)$, s.t. \mathbf{d}_1^U and \mathbf{d}_1^V are both drawn i.i.d. from π_U , \mathbf{d}_1^V is drawn i.i.d. from π_V^1 and \mathbf{d}_2^V is drawn i.i.d. from π_V^2 , with $\sum_x x\pi_V^1(x) = \sum_x x\pi_V^2(x)$. If $\phi_V^1(s) \geq \phi_V^2(s)$ for any $s \in (0, 1)$, then by denoting respectively γ_1 and γ_2 the asymptotic proportion of vertices matched by GREEDY in $\text{CM}_1(\mathbf{d}_1^U, \mathbf{d}_1^V)$ and $\text{CM}_2(\mathbf{d}_2^U, \mathbf{d}_2^V)$, it holds that necessarily $\gamma_2 \geq \gamma_1$.

For instance, let us assume that the degree distribution on the offline side is fixed. Then the matching size obtained by GREEDY is asymptotically larger if vertices on the online side all have exactly the same degree d rather than if those degrees are drawn from a Poisson distribution with expectation d .

A similar result (with a different criterion) holds with fixed degree distribution on the online side and differing one on the offline side.

2.4 GREEDY can outperform RANKING !

We recall that the RANKING algorithm, which is the worse case optimal, chooses at random a ranking over \mathcal{U} and uses it to break ties (i.e., if two vertices u and u' can be matched to v_k , then it is the one with the smallest rank that is matched by RANKING). Quite surprisingly, we get that in the configuration model RANKING can have a worse competitive ratio than GREEDY, which advocates again for its thorough study.

Proposition 1. *Let γ_R and γ_G be the asymptotic performances of RANKING and GREEDY on the 2-regular graph. The following holds:*

$$\gamma_G > \gamma_R.$$

In other words, GREEDY outperforms RANKING in the 2-regular graph.

We conjecture that the above result actually holds for any $d \geq 2$, and more generally for a wide class of distributions $\pi_{\mathcal{U}}$ and $\pi_{\mathcal{V}}$ (finding a general criterion would be very interesting). The proof of Proposition 1 is provided in Appendix G. The main idea is that in the 2-regular graph, RANKING is biased towards selecting as matches vertices with two remaining half-edges rather than just one. Indeed, vertices with only one remaining half-edge were not selected previously and thus have a higher rank. The vertices with only one remaining half-edge will not get matched in the subsequent iterations, so not picking them as matches is suboptimal. On the other hand, GREEDY picks any match uniformly at random and does not exhibit such bias.

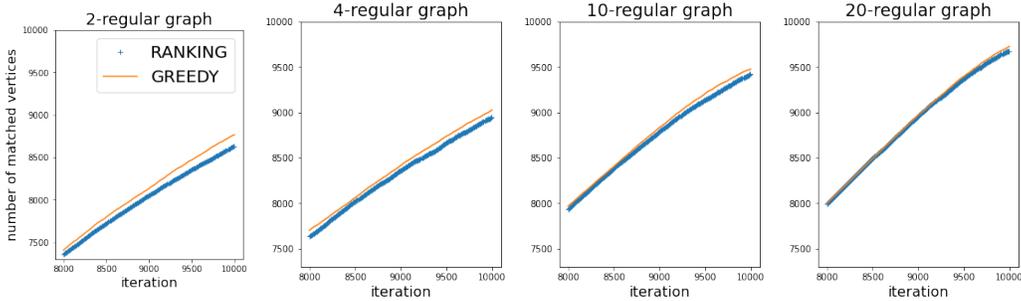


Figure 2: Experimental performances of GREEDY vs. RANKING on d -regular graphs

3 Ideas of proof of Theorem 1

The main idea behind the proof of Theorem 1 (postponed to Section D) is to show that the random deterministic evolution of the matching size generated by GREEDY is closely related to the solution of some ODE (this is sometimes called “the differential equation method” [Wormald, 1995] or “stochastic approximations” [Robbins and Monro, 1951]). Computing the solution of the ODE is easier - if not explicitly, at least numerically in intricate cases - than estimating the performances of GREEDY by Monte-Carlo simulations and it provides qualitative, as well as quantitative, properties.

Tracking the matching size is non-trivial because the vertices (in \mathcal{U} and \mathcal{V}) have different degrees, hence some of them are more likely to be matched than others. However, in the configuration model, each vertex has the same distribution of degrees before the sequences $\mathbf{d}^{\mathcal{U}}$ and $\mathbf{d}^{\mathcal{V}}$ are fixed. As a consequence, the proof relies on the three following techniques

1. The graph is built sequentially, along with the matching and not beforehand (fixing the “randomness” at the beginning would be very difficult to handle in the analysis). Thankfully, this does not change the law of the graph generated (this is obviously crucial).
2. We are not only going to track the size of the matching built as we need to handle different probabilities of matching (and pairing the graph) for each vertex. As a consequence, we are going to track the numbers of non-matched vertices which have still i half-edges to be paired and the number of already matched vertices that have j half-edges remaining. This will give one different ODE per value of i of j .

Since $\pi_{\mathcal{U}}$ and $\pi_{\mathcal{V}}$ are sub-Gaussian, we will prove that with arbitrarily high probability - exponential in N -, there are only a polynomial number of such equations

3. All those differential equations are then “aggregated” to build the final ODE satisfied by the matching size. Interestingly, this aggregated ODE has a simple form, while the full system is on the other hand quite intricate.

In the following sub-sections, we separate the proofs in the different building blocks to provide intuitions; the proof of technical lemmas are deferred to the appendix.

3.1 Building the graph together with the matching

The first step in the analysis is to notice that the bipartite configuration model can be constructed by sequentially pairing the half-edges coming from \mathcal{V} . The matching generated by GREEDY is then constructed simultaneously with the graph. More precisely, given two sequences⁵ of non-negative integers $\mathbf{d}^{\mathcal{U}} = (d_1^{\mathcal{U}}, \dots, d_N^{\mathcal{U}})$ and $\mathbf{d}^{\mathcal{V}} \cup \{d_{T+1}^{\mathcal{V}}\} = (d_1^{\mathcal{V}}, \dots, d_T^{\mathcal{V}}, d_{T+1}^{\mathcal{V}})$, we introduce in the following a generating algorithm that simultaneously build the associated bipartite configuration model $\mathbf{CM}(\mathbf{d}^{\mathcal{U}}, \mathbf{d}^{\mathcal{V}})$ together with GREEDY. Recall that the bipartite configuration model is obtained through a uniform matching between the half-edges of \mathcal{U} and the half-edges of \mathcal{V} . To avoid confusion, we will call a *marked matching* a pairing of two half-edges that corresponds to an edge that will belong to the constructed matching \mathbf{M} . This construction pseudo-code is detailed in Algorithm 1.

Algorithm 1: GREEDY MATCHING CONFIGURATION MODEL WITHOUT CAPACITIES

Input: $\mathbf{d}^{\mathcal{U}} = (d_1^{\mathcal{U}}, \dots, d_N^{\mathcal{U}})$ and $\mathbf{d}^{\mathcal{V}} = (d_1^{\mathcal{V}}, \dots, d_T^{\mathcal{V}})$
Initialization. $\mathbf{M}_0 \leftarrow \emptyset$, $\mathcal{E}_0 \leftarrow \emptyset$ and $H_0^{\mathcal{U}} \leftarrow \{\text{half-edges of } \mathcal{U}\}$
for $t = 1, \dots, T$ **do**
 Order uniformly at random the edges emanating from v_t : $e_1^t, \dots, e_{k_t}^t$
 for $i = 1, \dots, k_t$ **do**
 Choose uniformly an half-edge $e_i^{\mathcal{U}}$ in $H^{\mathcal{U}}$
 $\mathcal{E} \leftarrow \mathcal{E} \cup \{u(e_i^{\mathcal{U}}), v_t\}$ // Create an edge between e_i^t and $e_i^{\mathcal{U}}$
 $H^{\mathcal{U}} \leftarrow H^{\mathcal{U}} \setminus \{e_i^{\mathcal{U}}\}$ // Remove the half-edge
 if v_t and $u(e_i^{\mathcal{U}})$ *unmatched* **then**
 $\mathbf{M}_t \leftarrow \mathbf{M}_{t-1} \cup \{u(e_i^{\mathcal{U}}), v_t\}$ // v_t is matched
 end
 end
end
 $\mathbf{CM}(\mathbf{d}^{\mathcal{U}}, \mathbf{d}^{\mathcal{V}}) \leftarrow (\mathcal{U}, \mathcal{V}, \mathcal{E})$.
Output: Bipartite configuration model $\mathbf{CM}(\mathbf{d}^{\mathcal{U}}, \mathbf{d}^{\mathcal{V}})$ and matching \mathbf{M}_T on it.

Since each pairing of each half-edge is done uniformly at random, the graph obtained at the end of the algorithm has indeed the law of a bipartite configuration model. Moreover, it is easy to see that \mathbf{M} corresponds to the matching constructed by GREEDY MATCHING on $\mathbf{CM}(\mathbf{d}^{\mathcal{U}}, \mathbf{d}^{\mathcal{V}})$.

3.2 Differential Equation Method - Stochastic Approximation

As mentioned above, several quantities are going to be tracked through time: for all $k \in \{0, \dots, T\}$ and all $i \geq 0$, we define:

- $F_i(k)$ as the number of vertices $u \in \mathcal{U}$ that are not yet matched at the end of step k and whose remaining degree is i , meaning that $d_u - i$ of their initial half-edges have been paired. We will refer them to as *free* vertices.
- $M_i(k)$ as the number of vertices $u \in \mathcal{U}$ already matched at the end of step k and whose remaining degree is i . We will refer them to as *marked* vertices.

Notice that for all $0 \leq k \leq T$, the sum $F_i(k) + M_i(k)$ corresponds to the total number of vertices of \mathcal{U} with remaining degree i at the end of step k . We also define

- $\widehat{F}(k) := \sum_{i \geq 0} i F_i(k)$ is the number of available half-edges attached to free vertices at the end of step k ,

⁵Without loss of generality, we assume that the additional extra vertex is always on the \mathcal{V} side.

- $\widehat{M}(k) := \sum_{i \geq 0} iM_i(k)$ is the number of available half-edges attached to marked vertices at the end of step k .

We are going to study the evolution of these quantities along with the one of GREEDY. A major ingredient of the proof is to show that $F_i(k)$ and $M_i(k)$ closely follow the solutions of some ODE. This is the so-called *differential equation method* [Wormald, 1995], stated in Appendix C. For instance, it can easily be seen that $\widehat{F}(k) + \widehat{M}(k)$ closely follows the function $t \mapsto \mu_{\mathcal{U}} - t\mu_{\mathcal{V}}$ on $(0, \mu_{\mathcal{U}}/\mu_{\mathcal{V}})$ in the following sense.

Lemma 1. *For every $\varepsilon > 0$, and for all $0 \leq k \leq T$,*

$$\left| \frac{\widehat{F}(k) + \widehat{M}(k)}{N} - \left(\mu_{\mathcal{U}} - \frac{k}{N} \mu_{\mathcal{V}} \right) \right| \leq \varepsilon.$$

with probability at least $1 - \exp\left(-\frac{N\varepsilon^2}{2\sigma_{\mathcal{U}}^2}\right) + \exp\left(-\frac{T\varepsilon^2}{2\sigma_{\mathcal{V}}^2}\right)$.

We now turn to each individual quantity F_i (resp. M_i). We can prove a similar result, yet the limit function is not explicit (unlike for the matching size as in Theorem 1 statement). The following Lemma 2 states that the discrete sequences of (free and marked) half-edges are closely related to the solutions of some system of differential equations.

Before stating it, we first introduce, for any sequence of non-negative numbers $(x_\ell)_{\ell \geq 0}$ and $(y_\ell)_{\ell \geq 0}$ such that $0 < \sum_{\ell} \ell(x_\ell + y_\ell) < \infty$, every $i \geq 0$, the following mappings

$$\Phi_i(x_0, x_1, \dots, y_0, y_1, \dots) := \frac{-i\mu_{\mathcal{V}}x_i + (i+1)\mu_{\mathcal{V}}x_{i+1} - h\left(\frac{\sum_{\ell \geq 0} \ell y_\ell}{\sum_{\ell \geq 0} \ell(x_\ell + y_\ell)}\right)(i+1)x_{i+1}}{\sum_{\ell \geq 0} \ell(x_\ell + y_\ell)} \quad (5)$$

and

$$\Psi_i(x_0, x_1, \dots, y_0, y_1, \dots) := \frac{-i\mu_{\mathcal{V}}y_i + (i+1)\mu_{\mathcal{V}}y_{i+1} + h\left(\frac{\sum_{\ell \geq 0} \ell y_\ell}{\sum_{\ell \geq 0} \ell(x_\ell + y_\ell)}\right)(i+1)x_{i+1}}{\sum_{\ell \geq 0} \ell(x_\ell + y_\ell)},$$

where h is the following function, well-defined on $[0, 1]$,

$$h(s) = \frac{1 - \phi_{\mathcal{V}}(s)}{1 - s}.$$

Lemma 2. *With probability $1 - \zeta N \exp(-\xi N^{c/2})$, there are at most N^c quantities F_i and M_i , and for all $0 \leq k \leq T$ and all $i \geq 0$*

$$\left| \frac{F_i(k)}{N} - f_i\left(\frac{k}{N}\right) \right| \leq \kappa N^{-2c} \quad \text{and} \quad \left| \frac{M_i(k)}{N} - m_i\left(\frac{k}{N}\right) \right| \leq \kappa N^{-2c},$$

where ζ, κ depend only on the (first two) moments of $\pi_{\mathcal{V}}$ and $\pi_{\mathcal{U}}$ and $c = 1/20$.

The continuous mappings f_i and m_i are solutions of the system of differential equations on $[0, \mu_{\mathcal{U}}/\mu_{\mathcal{V}})$

$$\begin{aligned} \frac{df_i}{dt} &= \Phi_i(f_0, f_1, \dots, m_0, m_1, \dots), \\ \frac{dm_i}{dt} &= \Psi_i(f_0, f_1, \dots, m_0, m_1, \dots), \\ f_i(0) &= \pi_{\mathcal{U}}(i), \\ m_i(0) &= 0. \end{aligned} \quad (6)$$

This system is well defined as stated by the following Lemma 3.

Lemma 3. *The system (6) has a unique solution which is well-defined on $[0, \mu_{\mathcal{U}}/\mu_{\mathcal{V}})$. More precisely, denoting by f and m the generating series of the sequences $(f_i)_{i \geq 0}$ and $(m_i)_{i \geq 0}$,*

$$f(t, s) = \sum_{i \geq 0} f_i(t) s^i \quad \text{and} \quad m(t, s) = \sum_{i \geq 0} m_i(t) s^i,$$

it holds that:

$$f\left(\frac{\mu_{\mathcal{U}}}{\mu_{\mathcal{V}}}(1 - e^{-\mu_{\mathcal{V}}t}), s\right) = \phi_{\mathcal{U}}\left((s-1)e^{-\mu_{\mathcal{V}}t} + 1 - F(t)\right), \quad (7)$$

and

$$m\left(\frac{\mu_{\mathcal{U}}}{\mu_{\mathcal{V}}}(1 - e^{-\mu_{\mathcal{V}}t}), s\right) = \int_0^t F'(u)\phi'_{\mathcal{U}}((s-1)e^{-\mu_{\mathcal{V}}u} + 1 - F(u)) du.$$

where F is a solution of the following ODE

$$\frac{\frac{1}{\mu_{\mathcal{U}}}\phi'_{\mathcal{U}}(1 - F(t))}{1 - \phi_{\mathcal{V}}\left(1 - \frac{1}{\mu_{\mathcal{U}}}\phi'_{\mathcal{U}}(1 - F(t))\right)} F'(t) = e^{-\mu_{\mathcal{V}}t}.$$

3.3 Aggregating solutions to compute GREEDY performances

To get Theorem 1, notice that the number of vertices matched by GREEDY is N minus the number of free vertices remaining at the end, which is approximately equal to $Nf(\frac{\mu_{\mathcal{U}}}{\mu_{\mathcal{V}}}, 1)$ by definition of f and because of Lemma 2. This corresponds to $t = +\infty$ in Equation (7), thus the performance of GREEDY is, with arbitrarily high probability, arbitrarily close to

$$N(1 - \phi_{\mathcal{U}}(1 - F(+\infty)))$$

The statement of Theorem 1 just follows from a simple final change of variable.

Conclusion

We studied theoretical performances of GREEDY algorithm on matching problems with different underlying structures. Those precise results are quite interesting and raise many questions, especially since GREEDY actually outperforms RANKING in many different situations (in theory for 2-regular graphs, but empirical evidence indicates that this happens more generically).

Our approach has also successfully been used to unveil some questions on the comparison between different possible models. But more general questions are still open; for instance, assuming that the expected degree is fixed, which situation is the more favorable to GREEDY and online algorithm: small or high variance, or more generally this distribution $\pi_{\mathcal{U}}$ or an alternative one $\pi'_{\mathcal{U}}$? The obvious technique would be to compare the solution of the different associated ODE's. Similarly, the questions of stability/robustness of the solution to variation in the distribution $\pi_{\mathcal{U}}$ and $\pi_{\mathcal{V}}$ are quite challenging and left for future work.

We believe online matching will become an important problem for the machine learning community in the future. Each year, the complexity of the underlying graphs increases and we are considering adding features to the model in future work (such as random variables on the edges, modeling the interest for a consumer for a given product), or connection modeled via some Kernel between vertices features (say, if users and products/campaigns are embedded in the same space). In this context, machine learning tools will certainly be needed to tackle the problem.

Acknowledgments and Disclosure of Funding

V. Perchet acknowledges support from the ANR under grant number #ANR-19-CE23-0026 as well as the support grant as part of the Investissement d'avenir project, reference LabEx Ecodec/ANR-11-LABX-0047. Nathan Noiry also acknowledges support from the Telecom Paris DSAIDIS chair and from the ANR ProGraM (ANR-19-CE40-0025). Flore Sentenac is supported by IP PARIS' PhD Funding.

References

- Nick Arnosti. Greedy matching in bipartite random graphs. *working paper*, 2019.
- Lars Backstrom, Paolo Boldi, Marco Rosa, Johan Ugander, and Sebastiano Vigna. Four degrees of separation, 2012.
- Albert-László Barabási, Réka Albert, and Hawoong Jeong. Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A: statistical mechanics and its applications*, 281(1-4):69–77, 2000.

- Edward A. Bender and E. Rodney Canfield. The asymptotic number of labeled graphs with given degree sequences. *J. Combinatorial Theory Ser. A*, 24(3):296–307, 1978.
- 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>.
- Béla Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European J. Combin.*, 1(4):311–316, 1980.
- Charles Bordenave, Marc Lelarge, and Justin Salez. Matchings on infinite graphs. *Probability Theory and Related Fields*, 157(1-2):183–208, 2013.
- Allan Borodin, Christodoulos Karavasilis, and Denis Pankratov. Greedy bipartite matching in random type poisson arrival model. *arXiv preprint arXiv:1805.00578*, 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. 01 2013. doi: 10.1137/1.9781611973105.7.
- Nikhil R. Devanur, Balasubramanian Sivan, and Yossi Azar. Asymptotically optimal algorithm for stochastic adwords. In *Proceedings of the 13th ACM Conference on Electronic Commerce, EC '12*, page 388–404, New York, NY, USA, 2012. Association for Computing Machinery. ISBN 9781450314152. doi: 10.1145/2229012.2229043. URL <https://doi.org/10.1145/2229012.2229043>.
- Martin Dyer, Alan Frieze, and Boris Pittel. The average performance of the greedy matching algorithm. *The Annals of Applied Probability*, pages 526–552, 1993.
- Nathanaël Enriquez, Gabriel Faraud, Laurent Ménard, and Nathan Noiry. Depth first exploration of a configuration model. *arXiv preprint arXiv:1911.10083*, 2019.
- Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships of the internet topology. In *Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication, SIGCOMM '99*, page 251–262, New York, NY, USA, 1999. Association for Computing Machinery. ISBN 1581131356. doi: 10.1145/316188.316229. URL <https://doi.org/10.1145/316188.316229>.
- Jon Feldman, Aranyak Mehta, Vahab Mirrokni, and Shan Muthukrishnan. Online stochastic matching: Beating $1-1/e$. In *2009 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 117–126. IEEE, 2009.
- Alan Frieze and Michał Karoński. *Introduction to random graphs*. Cambridge University Press, 2016.
- Christopher David Godsil. Matchings and walks in graphs. *Journal of Graph Theory*, 5(3):285–297, 1981.
- Gagan Goel and Aranyak Mehta. Online budgeted matching in random input models with applications to adwords. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '08*, page 982–991, USA, 2008a. Society for Industrial and Applied Mathematics.
- Gagan Goel and Aranyak Mehta. Online budgeted matching in random input models with applications to adwords. pages 982–991, 01 2008b. doi: 10.1145/1347082.1347189.
- 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>.

- László Lovász and Michael D Plummer. *Matching theory*, volume 367. American Mathematical Soc., 2009.
- Mohammad Mahdian and Qiqi Yan. Online bipartite matching with random arrivals: An approach based on strongly factor-revealing lps. 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.
- Herbert Robbins and Sutton Monro. A stochastic approximation method. *The annals of mathematical statistics*, pages 400–407, 1951.
- Remco Van Der Hofstad. *Random graphs and complex networks*, volume 1. Cambridge university press, 2016.
- Nicholas C Wormald. Differential equations for random processes and random graphs. *The annals of applied probability*, 5(4):1217–1235, 1995.
- Lenka Zdeborová and Marc Mézard. The number of matchings in random graphs. *Journal of Statistical Mechanics: Theory and Experiment*, 2006(05):P05003, 2006.

1. For all authors...
 - (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
 - (b) Did you describe the limitations of your work? [Yes]
 - (c) Did you discuss any potential negative societal impacts of your work? [N/A]
 - (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
 - (a) Did you state the full set of assumptions of all theoretical results? [Yes]
 - (b) Did you include complete proofs of all theoretical results? [Yes]
3. If you ran experiments...
 - (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes]
 - (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [N/A]
 - (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes]
 - (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [No] The experiments can be run on any personal laptop.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
 - (a) If your work uses existing assets, did you cite the creators? [Yes]
 - (b) Did you mention the license of the assets? [Yes]
 - (c) Did you include any new assets either in the supplemental material or as a URL? [Yes]
 - (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
 - (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
 - (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
 - (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
 - (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

A General version of the result

A.1 The fixed capacity matching problem

We now investigate the case where vertices $u \in \mathcal{U}$ have capacities, which means that they can be matched to several vertices $v \in \mathcal{V}$. Precisely, if the capacity of u is denoted by ω_u , then this vertex can be matched to at most ω_u vertices in \mathcal{V} (but as before, half-edges of u are going to be paired with d_u half-edges originating from \mathcal{V}). The graph is still constructed using the configuration model introduced in Section 2.1, i.e., the law of d_u is $\pi_{\mathcal{U}}$ (and similarly, degrees of $v \in \mathcal{V}$ are i.i.d., with law $\pi_{\mathcal{V}}$).

For the moment, to simplify the analysis and the results statements, we are going to assume that all vertices $u \in \mathcal{U}$ have the same initial capacity $C \in \mathbb{N}$. We denote the random graph with capacities generated this way by $\mathbf{CM}(d^{\mathcal{U}}, d^{\mathcal{V}}, C)$

Theorem 2. (Performances of GREEDY with fixed capacities)

Given $N \geq 1$ and $T = \frac{\mu_{\mathcal{U}}}{\mu_{\mathcal{V}}}N$, let M_T be the matching built by GREEDY on $\mathbf{CM}(d^{\mathcal{U}}, d^{\mathcal{V}}, C)$ then the following convergence in probability holds:

$$\frac{|M_T|}{CN} \xrightarrow[N \rightarrow +\infty]{\mathbf{P}} 1 - \sum_{k=0}^{C-1} \frac{1-k/C}{k!} G(1)^k \phi_{\mathcal{U}}^{(k)} (1 - G(1)).$$

where G is the unique solution of the following ordinary differential equation

$$G'(s) = \frac{1 - \phi_{\mathcal{V}} \left(1 - \frac{1}{\mu_{\mathcal{U}}} \Gamma_U(G(s)) \right)}{\frac{\mu_{\mathcal{V}}}{\mu_{\mathcal{U}}} \Gamma_U(G(s))}.$$

where

$$\Gamma_U(g) = \phi'_{\mathcal{U}}(1-g) + \sum_{k=1}^{C-1} \frac{g^k}{k!} \phi_{\mathcal{U}}^{(k+1)}(1-g)$$

Moreover, for any $s \in [0, 1]$, if $M_T(s)$ is the matching obtained by GREEDY after seeing a proportion s of vertices of \mathcal{V} , then

$$\frac{|M_T(s)|}{CN} \xrightarrow[N \rightarrow +\infty]{\mathbf{P}} 1 - \sum_{k=0}^{C-1} \frac{1-k/C}{k!} G(s)^k \phi_{\mathcal{U}}^{(k)} (1 - G(s)).$$

The proof of Theorem 2, in Appendix E, has three major differences with the one of Theorem 1:

1. The first one is that more quantities must be tracked, not just the number of vertices with remaining free half-edges, but the number of such vertices for each possible value of remaining capacity; the total number of equations is roughly speaking multiplied by a factor $(C+1)/2$ (since only $F_i(k)$ are affected by the capacities and not $M_i(k)$). We will therefore denote in the remaining by $F_i^{(c)}(k)$ the number of vertices with i remaining half-edges to be paired and with current capacity equal to c (those vertices can still be matched to c different vertices $v \in \mathcal{V}$).
2. The second major difference lies in the resolution of the system of differential equations. The solution was rather direct without capacities (i.e., $c = 1$). Unfortunately, the evolution of $F_i^{(c)}$ strongly depends on $F_i^{(c+1)}$. As a consequence, the trick is to solve this system by induction, starting from $c = C$ (this solution is almost identical to that of the case with no capacities) and then to inject this solution in the PDEs defining $F_i^{(C-1)}$ so on so forth. Indeed, the fluid limits of $\sum_i F_i(c)$ and $\sum_i M_i$, that we denote respectively by $f^{(c)}$ and m satisfy the following coupled equations (up to some time change $\theta(t)$ and where $H(t) = h(q(t))$ for some function $q(\cdot)$ introduced in the proof):

$$\partial_t f^{(c)}(\theta(t), s) = [-\mu_{\mathcal{V}}s + \mu_{\mathcal{V}} - H(t)] \partial_s f^{(c)}(\theta(t), s) + H(t) \partial_s f^{(c+1)}(\theta(t), s),$$

and

$$\partial_t m(\theta(t), s) = [-\mu_{\mathcal{V}}s + \mu_{\mathcal{V}}] \partial_s m(\theta(t), s) + H(t) \partial_s f^{(1)}(\theta(t), s).$$

3. Finally, the third main difference is how the performances of GREEDY are defined. The upper bound is obviously to create the minimum between CN and T matches (where T is the number of vertices in \mathcal{V}). Anyway, those two numbers are within a constant multiplicative factor (recall that $T = \frac{\mu_{\mathcal{U}}}{\mu_{\mathcal{V}}}N$ for a valid configuration model), hence we arbitrarily chose to normalize GREEDY performances by CN . As a consequence, the (normalized) performances of GREEDY now rewrite as

$$\frac{\sum_{i \geq 0} \left(M_i(T) + \sum_{c=1}^C \left(1 - \frac{c}{C}\right) F_i^{(c)}(T) \right)}{N},$$

where $M_i(k)$ still denotes the number of marked vertices, i.e., those whose capacities have been depleted before step k with i remaining half-edges to be paired.

A.2 General case, online matching with capacities

In the general case, we no longer assume that all vertices $u \in \mathcal{U}$ have the same initial capacities, but ω_u can be equal to any value in \mathbb{N} (yet this capacity is independent of the degree). Notice however that the capacities ω_u of vertices could be capped at their degrees d_u (since they would never be depleted otherwise). As a consequence, capacities can be assumed to be bounded by $C < N^\beta$ for some $\beta < 1$ since the maximal degree is also smaller than N^β with arbitrarily high probability.

We therefore denote by $p_c \in [0, 1]$ the fraction of vertices of \mathcal{U} whose initial capacity is exactly $c \in [1, C]$. Notice, we do not need to assume that capacities are drawn i.i.d. accordingly to some distribution, our results hold for any values $(p_c)_c$. We denote by $\text{CM}(\mathbf{d}^{\mathcal{U}}, \mathbf{d}^{\mathcal{V}}, \mathbf{p})$ the random graph with capacities generated.

Quite interestingly, the techniques are exactly the same as in the previous case: we consider the exact same system of differential equations; the only differences are the initial conditions. Similarly, the maximal matching size is no longer NC but $N\mathbb{E}_{\mathbf{p}}[c] := N \sum_c c p_c$. We also denote the cdf of the empirical distribution p_c by $P_c := \sum_{k \leq c} p_c$

Theorem 3. (Performances of GREEDY with different capacities)

Given $N \geq 1$ and $T = \frac{\mu_{\mathcal{U}}}{\mu_{\mathcal{V}}}N$, let M_T be the matching built by GREEDY on $\text{CM}(\mathbf{d}^{\mathcal{U}}, \mathbf{d}^{\mathcal{V}}, \mathbf{p})$ then the following convergence holds in probability:

$$\frac{|M_T|}{N\mathbb{E}_{\mathbf{p}}[c]} \xrightarrow[N \rightarrow +\infty]{\mathbf{P}} 1 - \sum_{k=0}^{C-1} \frac{\sum_{c=1}^C c p_{c+k}}{\mathbb{E}_{\mathbf{p}}[c]} \frac{1}{k!} G(1)^k \phi^{(k)}(1 - G(1)).$$

where G is the unique solution of the following ordinary differential equation

$$G'(s) = \frac{1 - \phi_{\mathcal{V}} \left(1 - \frac{1}{\mu_{\mathcal{U}}} \Gamma_{\mathcal{U}}^{\mathbf{p}}(G(s)) \right)}{\frac{\mu_{\mathcal{V}}}{\mu_{\mathcal{U}}} \Gamma_{\mathcal{U}}^{\mathbf{p}}(G(s))}.$$

with

$$\Gamma_{\mathcal{U}}^{\mathbf{p}}(g) = \phi'_{\mathcal{U}}(1 - g) + \sum_{k=1}^{C-1} \left(\frac{(1 - P_k) g^k}{k!} \phi_{\mathcal{U}}^{(k+1)}(1 - g) \right).$$

Moreover, for any $s \in [0, 1]$, if $M_T(s)$ is the matching obtained by GREEDY after seeing a proportion s of vertices of \mathcal{V} , then

$$\frac{|M_T(s)|}{N\mathbb{E}_{\mathbf{p}}[c]} \xrightarrow[N \rightarrow +\infty]{\mathbf{P}} 1 - \sum_{k=0}^{C-1} \frac{\sum_{c=1}^C c p_{c+k}}{\mathbb{E}_{\mathbf{p}}[c]} \frac{1}{k!} G(s)^k \phi^{(k)}(1 - G(s)).$$

As mentioned before, the proof (delayed to Appendix F) is rather similar to the previous one; the major difference is that the change of initial condition of the system of PDE makes it a bit more complicated to solve (hence the more intricate formulation of the result).

B Additional Numerical Experiments

B.1 Further comparisons between the theoretical result and simulations

We provide in Figure 3 a comparison between the score predicted by the numerical solutions of the ODE (the functions $1 - \phi_{\mathcal{U}}(1 - G(s))$) for 4-regular graphs and the simulated performance of GREEDY for various values of N . As expected, the deviations of the simulated trajectories remain within $\mathcal{O}(\sqrt{N})$ of the expected theoretical trajectory. Figure 4 illustrates the same comparison on an Erdős-Rényi graph whose expected degree equals 4.

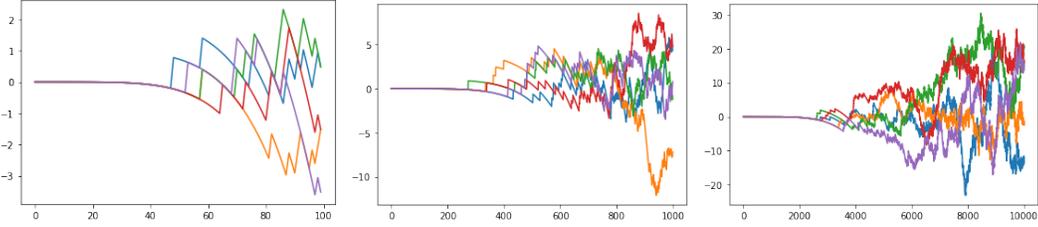


Figure 3: Difference between the theoretical performances and simulated performances of the GREEDY algorithm on the d -regular graph ($d = 4$) on 5 independent runs, with $N = 100, 1000, 10000$.

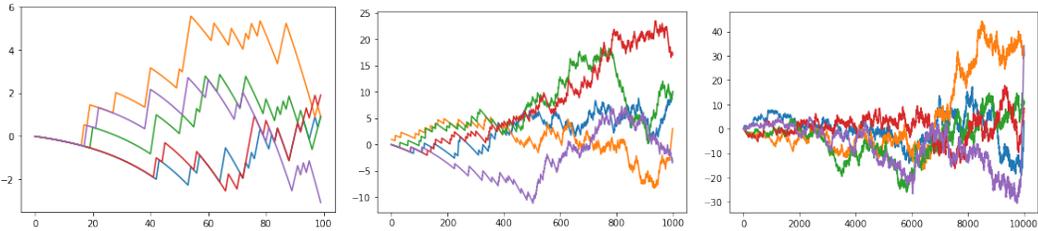


Figure 4: Difference between the theoretical value 2 and simulated performances of the GREEDY algorithm on the Erdős-Rényi graph, $c = 4$, on 5 independent runs, with $N = 100, 1000, 10000$.

In Figure 5, we plot the theoretical performance of the GREEDY algorithm along with its experimental performance on the d -regular graph for various values of d . We also plot the competitive ratio of GREEDY predicted by the ODE as a function of d . As expected, the score increases with d (as the problem becomes simpler and simpler for GREEDY when $d \geq 2$).

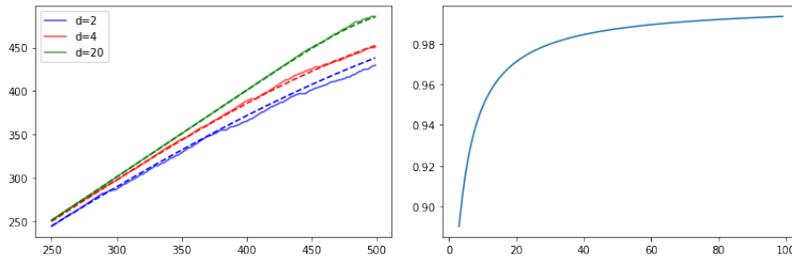


Figure 5: On the left, the expected theoretical performance of the GREEDY algorithm (dashed line) along with the simulated performance (full line) for various values of d . On the right, the expected competitive ratio of GREEDY on the d -regular graph as a function of d .

B.2 GREEDY vs RANKING

We further illustrate in this section the quite surprising fact that, in some configuration models, GREEDY actually outperforms RANKING.

In the adversarial configuration, it is known that the competitive ratio of RANKING is $1 - \frac{1}{e}$ which is bigger than the one of GREEDY, equal to $1/2$, see [Mehta, 2012]. In the following figures, we also plot the performances of two other “algorithms” SMALLEST and HIGHEST, for the sake of comparison; indeed, those are not admissible algorithms as they use the (future) knowledge of the number of half-edges of each vertex $u \in \mathcal{U}$.

More precisely, SMALLEST matches a vertex $v_k \in \mathcal{V}$ to the vertex $u \in \mathcal{U}$ with the smallest number of remaining half-edges (under the constraints obviously that $(u, v_k) \in \mathcal{E}$). As a consequence SMALLEST could be seen as an upper limit for an online algorithm.

HIGHEST does the opposite: it matches v_k to the vertex $u \in \mathcal{U}$ with the highest remaining number of half-edges. So HIGHEST should serve as a lower bound/sanity check for any online algorithm.

In Figure 6, the performances of those 4 matching “algorithms” (again SMALLEST and HIGHEST are not admissible as they use extra knowledge) are illustrated on configuration models with $d = 2, 4, 10$ and 20 .

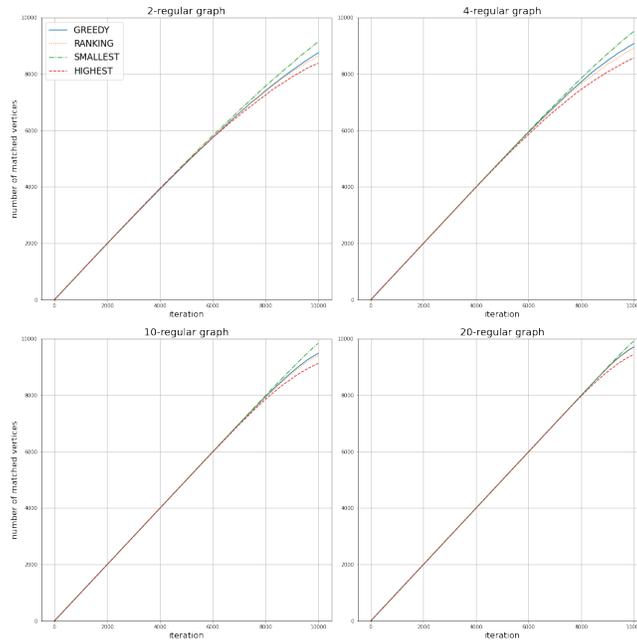


Figure 6: GREEDY outperforms RANKING in d -regular graphs

As mentioned before, GREEDY surprisingly outperforms RANKING in some configuration models, with a relative performance that decreases with d (which is rather natural on the other hand, since the relative performance of HIGHEST and SMALLEST also decreases).

Figure 6 also illustrates the different time steps at which algorithms fail to match new vertices v_k (because all the u they are paired with are already matched with another vertex v_j for some $j < k$). This happens later and later as d increases (as expected), at around half the horizon for $d = 2$ and roughly 82% with $d = 20$.

On the other hand, RANKING and GREEDY have the same performance on Erdos-Renyi graphs, which is a consequence of the memory-less property of those graphs, i.e. the probability of creating a match at each iteration depends only on the number of matched vertices, as shown in Figure 7.

In Figure 8, we plot the relative performance of RANKING and GREEDY on bi-degrees graph, where half the vertices have degree x , the other half $2x$. The plots illustrate that the best algorithm is not always the best one depending on the value of x .

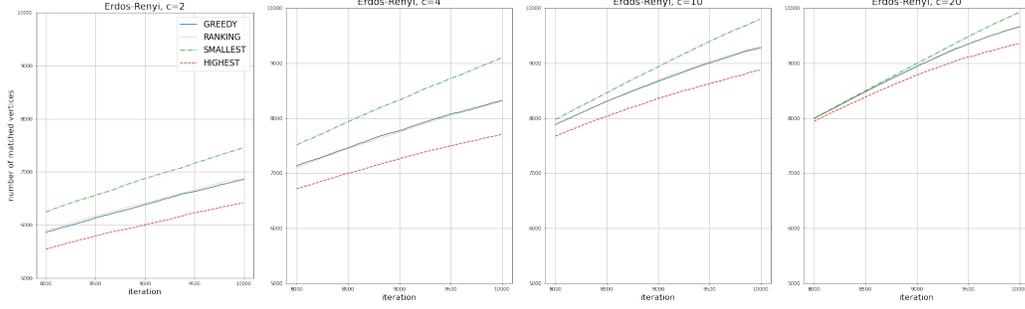


Figure 7: Experimental performances of GREEDY vs. RANKING on Erdos-Renyi graphs

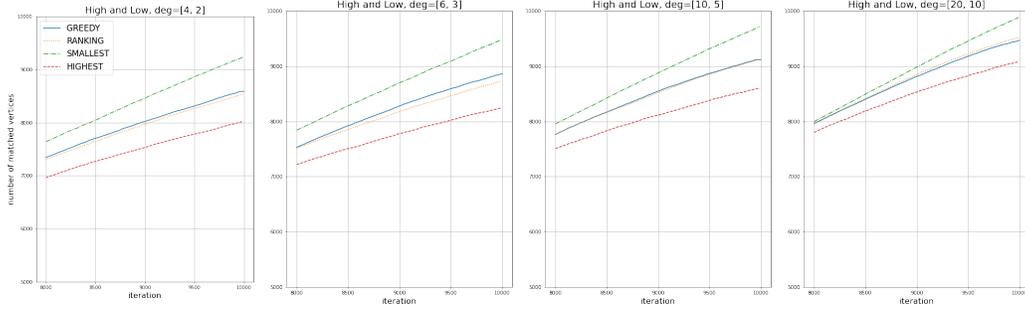


Figure 8: Experimental performances of GREEDY vs. RANKING on bi-degrees graphs

B.3 A few vertices with high capacity vs many vertices with low capacity

In this section, we investigate how nodes' capacities affect GREEDY's expected performance. The baseline is its performance on a random graph where all vertices have capacity 1 and the vertices degrees in \mathcal{U} and \mathcal{V} follow the distributions $\pi_{\mathcal{U}}$ and $\pi_{\mathcal{V}}$. The comparison graph with capacity C has $|\mathcal{U}|/C$ "in-place" vertices, each with a capacity C , and their degrees follows the modified distribution $\tilde{\pi}_C^{\mathcal{U}}$ where $\tilde{\pi}_C^{\mathcal{U}}(x = k) = \pi_{\mathcal{U}}(x = k/C)$. Informally, the graph with capacity C is built from the baseline graph by merging C vertices of equal degree d into a single vertex of degree dC .

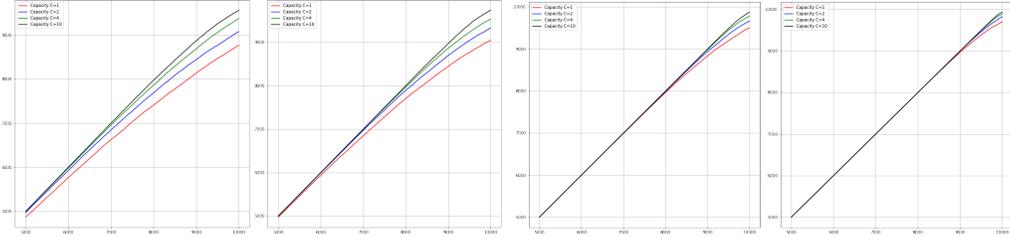


Figure 9: GREEDY performs better in high capacity graphs in d -regular graphs, from left to right $d = [2, 4, 10, 20]$

The results of the simulation illustrate that the GREEDY performs better on graphs with vertices of high capacity.

C Stochastic approximation & Differential equation method

The following theorem is an improved version of Wormald's Theorem [Enriquez et al., 2019].

Theorem 4. *Let $a > 0$. For all $N \geq 1$ and all $1 \leq k \leq N^a$, let $Y_k(i) = Y_k^{(N)}(i)$ be a Markov chain with respect to a filtration $\{\mathcal{F}_i\}_{i \geq 1}$. Suppose that, for all $k \geq 1$, there exists a function f_k such that:*

- $Y_k(0)/N = z_k(0)$;

- $|Y_k(i+1) - Y_k(i)| \leq N^\beta$;
- $|\mathbf{E} [Y_k(i+1) - Y_k(i) \mid \mathcal{F}_i] - f_k \left(\frac{i}{N}, \frac{(Y_k(i))_{1 \leq k \leq N^a}}{N} \right)| \leq cN^{-\lambda}$, for some constant $c > 0$

where $0 < \beta < 1/2$, $\lambda > 0$. Suppose that the following infinite system of differential equations with initial conditions $(z_k(0))_{k \geq 1}$ has a unique solution $(z_k)_{k \geq 1}$:

$$\forall k \geq 1, \quad z_k'(t) = f_k(t, (z_k(t))_{k \geq 1}).$$

Then, for all $k \geq 1$, $Y_k(\lfloor tN \rfloor)/N$ converges in probability towards z_k for the topology of uniform convergence.

More precisely, for every $1 < \varepsilon < \frac{1-\beta}{\beta}$, for every $\frac{(1+\varepsilon)\beta}{2} < \alpha < \varepsilon\beta$ and for every $0 \leq i \leq \frac{N}{\omega}$ where $\omega = N^{(1+\varepsilon)\beta}$, it holds that

$$\mathbb{P} \left(\left| Y(i\omega) - z \left(\frac{i\omega}{N} \right) N \right| \leq i \left(N^{\alpha+\beta} + cN^{(1+\varepsilon)\beta-\lambda} + N^{2(1+\varepsilon)\beta-2} \right) \right) \leq i \exp \left(- \frac{N^{2\alpha-(1+\varepsilon)\beta}}{2} \right)$$

D Proofs of technical steps of Theorem 1

D.1 Proof of Lemma 1

It is an application of (maximal) Hoeffding-Azuma inequality since, for every $0 \leq k \leq M-1$,

$$\mathbf{E} \left[\left(\widehat{F}(k+1) + \widehat{M}(k+1) \right) - \left(\widehat{F}(k) + \widehat{M}(k) \right) \mid \mathcal{F}_k \right] = -\mathbf{E} [d_k^\mathcal{V}] = -\mu_\mathcal{V}.$$

D.2 Proof of Lemma 2

Since $\pi_\mathcal{V}$ is $\sigma_\mathcal{V}$ subGaussian, then for any $\beta > 0$,

$$\mathbf{P} \left(\exists i \in \{1, \dots, T\}, d_i^\mathcal{V} \geq \mu_\mathcal{V} + N^\beta \right) \leq T \exp \left(- \frac{N^{2\beta}}{2\sigma_\mathcal{V}^2} \right).$$

In particular, for some $\beta < 1/2$ to be chosen later on, if $\mu_\mathcal{V} \leq N^\beta/2$, then all degrees are smaller than N^β with probability at least $1 - T \exp \left(- \frac{N^{2\beta}}{8\sigma_\mathcal{V}^2} \right)$; from now on, we will place ourselves on that event.

We also denote by $(\mathcal{F}_k)_{0 \leq k \leq M}$ the natural filtration associated to the GREEDY MATCHING algorithm. In order to apply Theorem 4, it remains to control for every $i \geq 0$ and $0 \leq k \leq M-1$,

$$\left| \mathbf{E} [F_i(k+1) - F_i(k) \mid \mathcal{F}_k] - \Phi_i \left(F_0 \left(\frac{k}{N} \right), F_1 \left(\frac{k}{N} \right), \dots, M_0 \left(\frac{k}{N} \right), M_1 \left(\frac{k}{N} \right), \dots \right) \right|$$

and

$$\left| \mathbf{E} [M_i(k+1) - M_i(k) \mid \mathcal{F}_k] - \Psi_i \left(F_0 \left(\frac{k}{N} \right), F_1 \left(\frac{k}{N} \right), \dots, M_0 \left(\frac{k}{N} \right), M_1 \left(\frac{k}{N} \right), \dots \right) \right|$$

Let $0 \leq k \leq T - \frac{2iN^\gamma}{\mu^\gamma}$, with $\gamma > 1/2$ some parameter to be fixed later, so that, according to Lemma 1, with probability at least $1 - \exp \left(- \frac{N^{2\gamma-1}}{2\sigma_\mathcal{U}^2} \right) - \exp \left(- \frac{\mu_\mathcal{U} N^{2\gamma-1}}{2\mu_\mathcal{V} \sigma_\mathcal{V}^2} \right)$ it holds that $\widehat{F}(k) + \widehat{M}(k) \geq N^\gamma$.

Recall that, in the k -th step of the algorithm, half-edges of the k -th vertex of V_N are ordered uniformly at random: $(e_i^k)_i$ for $i = 1, \dots, d_k^\mathcal{V}$. Then, each of these half-edges is sequentially paired uniformly at random with half-edges of $d^\mathcal{V}$ that are not yet paired. Let u_i^k be the vertex to which e_i^k is paired and let I_k be the first integer i such that u_i^k belongs to the free vertices of \mathcal{U} at time k , that is to the vertices that are not yet matched. If such an integer does not exist, that is when all u_i^k are already

matched, we set $I_k = +\infty$. As a consequence, we aim at estimating $\mathbf{P}(I_k = i \mid \mathcal{F}_k)$ for the different admissible values, where this probability has the following explicit definition

$$\begin{aligned} \mathbf{P}(I_k = i \mid \mathcal{F}_k) &= \frac{\widehat{M}(k)}{\widehat{F}(k) + \widehat{M}(k)} \frac{\widehat{M}(k) - 1}{\widehat{F}(k) + \widehat{M}(k) - 1} \cdots \frac{\widehat{M}(k) - (i-2)}{\widehat{F}(k) + \widehat{M}(k) - (i-2)} \frac{\widehat{F}(k)}{\widehat{F}(k) + \widehat{M}(k) - (i-1)} \\ &= \frac{\widehat{M}(k)!}{(\widehat{M}(k) - (i-1))!} \frac{(\widehat{F}(k) + \widehat{M}(k) - i)!}{(\widehat{F}(k) + \widehat{M}(k))!} \widehat{F}(k) \end{aligned}$$

First, assume that $\widehat{M}(k) \geq 2N^\theta$ for some parameter $\theta > 2\beta$ to be chosen later, so that those probabilities are all strictly positive. Using Stirling approximation formula, we get that, with $p(k) = \frac{\widehat{M}(k)}{\widehat{F}(k) + \widehat{M}(k)}$ and for any i ,

$$0 \geq \frac{\mathbf{P}(I_k = i \mid \mathcal{F}_k) - (1 - p(k))^{i-1} p(k)}{(1 - p(k))^{i-1} p(k)} \geq -2 \frac{N^{2\beta}}{N^\theta} - \frac{N^\beta}{N^\gamma}$$

Second, assume that $\widehat{M}(k) < 2N^\theta$ for some $\theta > \beta$. This immediately implies that, for i ,

$$0 \geq \mathbf{P}(I_k = i \mid \mathcal{F}_k) - (1 - p(k))^{i-1} p(k) \geq -2 \frac{N^\theta}{N^\gamma}$$

Similar inequalities holds for $\mathbf{P}(I_k = +\infty \mid \mathcal{F}_k)$, except that it is approximately equal to $\mathbf{E}[(1 - p(k))^{d_k^\nu}] = \phi_\nu(1 - p(k))$.

It remains to control the evolution of the processes $F_i(k)$ and $M_i(k)$. Notice that, by their very definition, on the event $I_k = x$ for some $1 \leq x \leq d_k^\nu$, the following happens:

1. The first $x - 1$ half-edges e_k^1, \dots, e_k^x are paired uniformly at random with marked half-edges of \mathcal{U} . If the corresponding vertex has a remaining degree equal to i , then M_i decreases by one, and M_{i-1} increases by one.
2. The x -th half-edge e_k^x is paired uniformly at random with free half-edge of \mathcal{U} . If the corresponding vertex has a remaining degree i , then F_i decreases by one, and M_{i-1} increases by one.
3. The $d_k^\nu - x$ remaining half-edges $e_k^{x+1}, \dots, e_k^{d_k^\nu}$ are paired uniformly at random with half-edges of \mathcal{U} . If the corresponding vertex is free with remaining degree i , then F_i decreases by one and F_{i-1} increases by one. Otherwise, if the corresponding vertex is marked with remaining degree i , then M_i decreases by one and M_{i-1} increases by one.

Notice that, after the pairing of each half-edges, the quantity $\widehat{F}(k)$ (resp. $\widehat{M}(k)$) may decrease (resp. increase) by one. Therefore, working on the event where $d_k^\nu \leq N^\beta$, we deduce that \widehat{F} and \widehat{M} are affected by an additive term of order at most N^β . The same argument holds on F_i and M_i .

All of these considerations imply that

$$\begin{aligned} &\left| \mathbf{E}[F_i(k+1) - F_i(k) \mid \mathcal{F}_k, I_t = x] - \left(-\frac{iF_i(k)}{\widehat{F}(k)} + (\mu_\nu - x) \left(-\frac{iF_i(k)}{\widehat{F}(k) + \widehat{M}(k)} + \frac{(i+1)F_{i+1}(k)}{\widehat{F}(k) + \widehat{M}(k)} \right) \right) \right| \\ &\leq 2\sigma_{\mu_\nu}^2 \frac{N^\beta}{N^\gamma} \end{aligned}$$

and similarly

$$\begin{aligned} &\left| \mathbf{E}[M_i(k+1) - M_i(k) \mid \mathcal{F}_k, I_t = x] - \left((x-1) \left(-\frac{iM_i}{\widehat{M}} + \frac{(i+1)M_{i+1}}{\widehat{M}} \right) + \frac{(i+1)F_{i+1}}{\widehat{F}} \right) \right. \\ &\quad \left. + (\mu_\nu - x) \left(-\frac{iM_i}{\widehat{F} + \widehat{M}} + \frac{(i+1)M_{i+1}}{\widehat{F} + \widehat{M}} \right) \right| \\ &\leq 2\sigma_\mu^2 \frac{N^\beta}{N^\gamma} + 2 \sum_{j=1}^{x-1} \frac{j}{\widehat{M}(k) - j} \end{aligned}$$

Finally, the case $I_t = +\infty$ is handled similarly, as by definition

$$\mathbf{E}[F_i(k+1) - F_i(k) \mid \mathcal{F}_k, I_t = +\infty] = 0.$$

and the following also holds also holds:

$$\left| \mathbf{E}[M_i(k+1) - M_i(k) \mid \mathcal{F}_k, I_t = +\infty] - \mu_{\mathcal{V}} \left(-\frac{iM_i}{\widehat{M}} + \frac{(i+1)M_{i+1}}{\widehat{M}} \right) \right| \leq 2\sigma_{\mu}^2 \frac{N^{\beta}}{N^{\gamma}} + 2 \sum_{j=1}^{d_k^{\mathcal{V}}-1} \frac{j}{\widehat{M}(k) - j}.$$

It remains to compute the expected variation in $F_i(k)$ and $M_i(k)$. It is a bit simpler for the former, but still, to lighten the notations, we write $p = p_k$ and $q = q_k$ in the following computation.

$$\begin{aligned} & \mathbf{E}[F_i(k+1) - F_i(k) \mid \mathcal{F}_k] \\ &= \mathbf{E}_{d_k^{\mathcal{V}} \sim \pi_{\mathcal{V}}} \left[\sum_{x=1}^{d_k^{\mathcal{V}}} q^{x-1} \left(-\frac{iF_i}{\widehat{F} + \widehat{M}} \right) + p \sum_{x=1}^{d_k^{\mathcal{V}}} q^{x-1} (d_k^{\mathcal{V}} - x) \left(-\frac{iF_i}{\widehat{F} + \widehat{M}} + \frac{(i+1)F_{i+1}}{\widehat{F} + \widehat{M}} \right) \right] + \eta_N \\ &= \frac{1}{\widehat{F} + \widehat{M}} \mathbf{E}_{d_k^{\mathcal{V}} \sim \pi_{\mathcal{V}}} \left[-iF_i \left(\sum_{x=1}^{d_k^{\mathcal{V}}} q^{x-1} + p d_k^{\mathcal{V}} \sum_{x=1}^{d_k^{\mathcal{V}}} q^{x-1} - p \sum_{x=1}^{d_k^{\mathcal{V}}} x q^{x-1} \right) \right. \\ & \quad \left. + (i+1)F_{i+1} \left(p d_k^{\mathcal{V}} \sum_{x=1}^{d_k^{\mathcal{V}}} q^{x-1} - p \sum_{x=1}^{d_k^{\mathcal{V}}} x q^{x-1} \right) \right] + \eta_N \\ &= \frac{1}{\widehat{F} + \widehat{M}} \mathbf{E}_{d_k^{\mathcal{V}} \sim \pi_{\mathcal{V}}} \left[-iF_i \left(\frac{1 - q^{d_k^{\mathcal{V}}}}{1 - q} + d_k^{\mathcal{V}}(1 - q^{d_k^{\mathcal{V}}}) - \frac{d_k^{\mathcal{V}} q^{d_k^{\mathcal{V}}+1} - (d_k^{\mathcal{V}} + 1)q^{d_k^{\mathcal{V}} + 1}}{1 - q} \right) \right. \\ & \quad \left. + (i+1)F_{i+1} \left(q(1 - q^{d_k^{\mathcal{V}}}) - \frac{d_k^{\mathcal{V}} q^{d_k^{\mathcal{V}}+1} - (d_k^{\mathcal{V}} + 1)q^{d_k^{\mathcal{V}} + 1}}{1 - q} \right) \right] + \eta_N \\ &= \frac{1}{\widehat{F} + \widehat{M}} \mathbf{E}_{d_k^{\mathcal{V}} \sim \pi_{\mathcal{V}}} \left[-iF_i d_k^{\mathcal{V}} + (i+1)F_{i+1} \frac{q^{d_k^{\mathcal{V}}} - d_k^{\mathcal{V}} q + (d_k^{\mathcal{V}} - 1)}{1 - q} \right] + \eta_N \\ &= \frac{1}{\widehat{F} + \widehat{M}} \mathbf{E}_{d_k^{\mathcal{V}} \sim \pi_{\mathcal{V}}} \left[-iF_i + (i+1)d_k^{\mathcal{V}} F_{i+1} + (i+1)F_{i+1} \frac{1 - q^{d_k^{\mathcal{V}}}}{1 - q} \right] + \eta_N \\ &= \frac{-i\mu_{\mathcal{V}} F_i + (i+1)\mu_{\mathcal{V}} F_{i+1} - (i+1)h(q)F_{i+1}}{\widehat{F} + \widehat{M}} + \eta_N, \end{aligned}$$

which is exactly (5), up to error term η_N that satisfies, if $\widehat{M}(k) < 2N^{\theta}$,

$$|\eta_N| \leq 2\sigma_{\mu_{\mathcal{V}}}^2 \frac{N^{\beta}}{N^{\gamma}} + 2\mu_{\mathcal{V}}^2 \frac{N^{\theta}}{N^{\gamma}}$$

and, if $\widehat{M}(k) \geq 2N^{\theta}$,

$$|\eta_N| \leq 3\sigma_{\mu_{\mathcal{V}}}^2 \frac{N^{\beta}}{N^{\gamma}} + 2\mu_{\mathcal{V}}^2 \frac{N^{2\beta}}{N^{\theta}} + \mu_{\mathcal{V}}^2 \frac{N^{\beta}}{N^{\gamma}}$$

Computations are quite similar for the difference in $M_i(k)$ and the error term still depends whether $\widehat{M}(k)$ is bigger, or smaller, than $2N^\theta$:

$$\begin{aligned}
& \mathbf{E}[M_i(k+1) - M_i(k) \mid \mathcal{F}_k] \\
&= \mathbf{E}_{d_k^\nu \sim \pi_\nu} \left[\left(\frac{d_k^\nu(1-p)^{d_k^\nu}}{\widehat{M}} + \sum_{x=1}^{d_k^\nu} pq^{x-1} \left(\frac{(x-1)}{\widehat{M}} + \frac{d_k^\nu - x}{\widehat{M} + \widehat{F}} \right) \right) \left((i+1)M_{i+1} - iM_i \right) \right] \\
&\quad + \mathbf{E}_{d_k^\nu \sim \pi_\nu} \left[\sum_{x=1}^{d_k^\nu} pq^{x-1} \frac{(i+1)F_{i+1}}{\widehat{F}} \right] + \varepsilon_N \\
&= \frac{1}{\widehat{M} + \widehat{F}} \mathbf{E}_{d_k^\nu \sim \pi_\nu} \left[\left(d_k^\nu(1-p)^{d_k^\nu-1} + \sum_{x=1}^{d_k^\nu} (pq^{x-2}(x-1) + pq^{x-1}(d_k^\nu - x)) \right) \left((i+1)M_{i+1} - iM_i \right) \right] \\
&\quad + \frac{(i+1)h(q)F_{i+1}}{\widehat{M} + \widehat{F}} + \varepsilon_N \\
&= \frac{\mu_\nu((i+1)M_{i+1} - iM_i) + (i+1)h(q)F_{i+1}}{\widehat{F} + \widehat{M}} + \varepsilon_N,
\end{aligned}$$

where ε_N satisfies, if $\widehat{M}(k) < 2N^\theta$,

$$|\varepsilon_n| \leq 2\sigma_\nu^2 \frac{N^\theta}{N^\gamma} + 2\sigma_\nu^2 \frac{N^\beta}{N^\gamma} + 2\mu_\nu^2 \frac{N^\theta}{N^\gamma};$$

and, if $\widehat{M}(k) \geq 2N^\theta$, it satisfies

$$|\varepsilon_n| \leq 3\sigma_\nu^2 \frac{N^\beta}{N^\gamma} + 2\mu_\nu^2 \frac{N^{2\beta}}{N^\theta} + \mu_\nu^2 \frac{N^\beta}{N^\gamma};$$

We used in the above computations (at the third equality) the following observation:

$$kq^{k-1} + \sum_{x=1}^k pq^{x-1} \left(\frac{x-1}{q} + k-x \right) = k$$

Summing error terms over all the $2N^\beta$ equations relating F_i to f_i and M_j to m_j , the error terms coming from the differential equation method Theorem 4, and using the fact that m is $\mu_\mathcal{U}$ -Lipschitz, we get that the total error, defined by,

$$\text{Err} := \sup_{s \in [0,1]} \left| \frac{|M_T(s)|}{N} - (1 - \phi_\mathcal{U}(1 - G(s))) \right|$$

satisfies

$$\text{Err} \leq N^\beta N^{(1+\varepsilon)\beta} (N^{\alpha+\beta} + 4(\sigma_\nu^2 + \mu_\nu^2) N^{(1+\varepsilon)\beta - \gamma/2 + \beta} + N^{2(1+\varepsilon)\beta - 2}) + \mu_\mathcal{U} N^{(1+\varepsilon)\beta}$$

as soon as $\theta = \beta + \frac{\gamma}{2}$.

It remains to pick admissible values for the different parameters, such as the following ones (checking admissibility follows from immediate computations):

$$\beta = 1/20, \epsilon = 10, \gamma = 21/40, \theta = 25/80, \alpha.23/80$$

Those choices ensures that $\text{Err} = \mathcal{O}(N^{-1/20})$.

All those arguments hold with probability at least (summing all the bad event probabilities)

$$1 - T \exp\left(-\frac{N^{2\beta}}{2\sigma_\nu^2}\right) - \exp\left(-\frac{N^{2\gamma-1}}{2\sigma_\mathcal{U}^2}\right) - \exp\left(-\frac{\mu_\mathcal{U} N^{2\gamma-1}}{2\mu_\nu \sigma_\nu^2}\right) - 2N^\beta N^{(1+\varepsilon)\beta} \exp(-N^{2\alpha-(1+\varepsilon)\beta}) \geq 1 - \mathcal{O}(N \exp(-\zeta N^{1/40}))$$

where the equality holds because of the choice of parameters.

D.3 Proof of Lemma 3

Notice that the functions f and m satisfy the following partial differential equations:

$$\partial_t f(t, s) = \frac{1}{\mu_U - t\mu_V} [-\mu_V s + \mu_V - h(q(t))] \partial_s f(t, s),$$

and

$$\partial_t m(t, s) = \frac{1}{\mu_U - t\mu_V} [-\mu_V s + \mu_V] \partial_s m(t, s) + h(q(t)) \partial_s f(t, s),$$

where $q(t) = \partial_s f(t, 1) / (\mu_U - t\mu_V)$.

To solve these equations, we first perform a time change to get rid of the denominator. Let

$$\theta(t) = \frac{\mu_U}{\mu_V} (1 - e^{-\mu_V t})$$

so that $\theta'(t) = \mu_U - \theta(t)\mu_V$. In order to simplify notations, we set:

$$H(t) := h(q(\theta(t))).$$

Then, the new functions

$$g(t, s) := f(\theta(t), s) \quad \text{and} \quad o(t, s) := m(\theta(t), s)$$

satisfy the following PDEs:

$$\partial_t g(t, s) = [-\mu_V s + \mu_V - H(t)] \partial_s g(t, s), \quad (8)$$

and

$$\partial_t o(t, s) = [-\mu_V s + \mu_V] \partial_s o(t, s) + H(t) \partial_s g(t, s). \quad (9)$$

These two equations fall into the classical framework of *transport* differential equation and can be explicitly solved. We give the details for the reader's convenience.

Solution of (8). Let s be a solution of the following ODE:

$$s'(t) = \mu_V s(t) - \mu_V + H(t). \quad (10)$$

Then, the function g is constant along the curve $(t, s(t))$. Indeed:

$$\frac{d}{dt} g(t, s(t)) = \partial_t g(t, s) + s'(t) \partial_s g(t, s) = 0.$$

The differential equation (10) admits the following general solutions:

$$s_c(t) = \left[c + e^{-\mu_V t} - 1 + \int_0^t e^{-\mu_V u} H(u) du \right] e^{\mu_V t}.$$

Therefore,

$$(t, s) = (t, s_c(t)) \iff c = c(t, s) = (s - 1)e^{-\mu_V t} + 1 - \int_0^t e^{-\mu_V u} H(u) du,$$

and we deduce that (the initial condition is $g(0, s) = \phi_U(s)$):

$$g(t, s) = g(0, c(t, s)) = \phi_U(c(t, s)) = \phi_U \left((s - 1)e^{-\mu_V t} + 1 - \int_0^t e^{-\mu_V u} H(u) du \right). \quad (11)$$

Solution of (9). Let $s_\gamma(t) = \gamma e^{\mu_V t} + 1$. Then, $s'_\gamma(t) = \mu_V s(t) - \mu_V$ and we deduce that, along the curves $(t, s_\gamma(t))$, $o(t, s)$ satisfies the following ODE:

$$\frac{d}{dt} o(t, s_\gamma(t)) = \frac{1 - q(t)^d}{1 - q(t)} \partial_s g(t, s_\gamma(t)).$$

Since

$$(t, s) = (t, s_\gamma(t)) \iff \gamma = \gamma(t, s) = (s - 1)e^{-\mu_V t},$$

we deduce that:

$$o(t, s) = \int_0^t H(u) \partial_s g(u, (s-1)e^{-\mu\nu(t-u)} + 1) du. \quad (12)$$

We now define the function $F(\cdot)$ as

$$F(t) := \int_0^t e^{-\mu\nu u} H(u) du. \quad (13)$$

Using Equations (11) and (12), one can easily deduce that

$$\partial_s g(t, 1) = e^{-\mu\nu t} \phi'_{\mathcal{U}}(1 - F(t))$$

and

$$\begin{aligned} \partial_s o(t, 1) &= \int_0^t H(u) e^{-\mu\nu u} \phi''_{\mathcal{U}}(1 - F(u)) du \\ &= \phi'_{\mathcal{U}}(1) - \phi'_{\mathcal{U}}(1 - F(t)) = (\mu_{\mathcal{U}} - \phi'_{\mathcal{U}}(1 - F(t))) e^{-\mu\nu t}. \end{aligned}$$

In particular,

$$\partial_s g(t, 1) + \partial_s o(t, 1) = \partial_s f(\theta(t), 1) + \partial_s m(\theta(t), 1) = \mu_{\mathcal{U}} e^{-\mu\nu t}.$$

Therefore,

$$H(t) = \frac{1 - \phi_{\mathcal{V}} \left(\frac{\partial_s o(t, 1)}{\partial_s g(t, 1) + \partial_s o(t, 1)} \right)}{1 - \frac{\partial_s o(t, 1)}{\partial_s g(t, 1) + \partial_s o(t, 1)}} = \mu_{\mathcal{U}} \frac{1 - \phi_{\mathcal{V}} \left(1 - \frac{1}{\mu_{\mathcal{U}} \phi'_{\mathcal{U}}(1 - F(t))} \right)}{1 - \phi_{\mathcal{U}}(1 - F(t))},$$

which yields the following ordinary differential equation on F :

$$\frac{\frac{1}{\mu_{\mathcal{U}}} \phi'_{\mathcal{U}}(1 - F(t))}{1 - \phi_{\mathcal{V}} \left(1 - \frac{1}{\mu_{\mathcal{U}}} \phi'_{\mathcal{U}}(1 - F(t)) \right)} F'(t) = e^{-\mu\nu t}. \quad (14)$$

E Proof of Theorem 2

We recall the notations introduced. For all $k \in \{0, \dots, T\}$, all $c \in \{0, \dots, C\}$ and all $i \geq 0$, we define:

- $F_i^{(c)}(k)$ the number of vertices of \mathcal{U} that still have capacity c at the end of step k and whose remaining degree is i . Those vertices are referred to as *free* (with remaining degree i and capacity c at the end of step k).
- $M_i(k)$ the number of vertices of \mathcal{U} that have capacity $c = 0$ at the end of step k and whose remaining degree is i . Those vertices are referred to as *marked* (with remaining degree i at the end of step k).

We also define as before the number of remaining half-edges to respectively free and marked vertices as

$$\widehat{F}(k) = \sum_{c=1}^C \sum_{i \geq 0} i F_i^{(c)}(k), \quad \text{and} \quad \widehat{M}(k) = \sum_{i \geq 0} i M_i(k).$$

The normalized performance of GREEDY is the ratio between the matched vertices in \mathcal{V} and its maximal number, equal to CN :

$$A = \frac{\sum_{i \geq 0} \left(C M_i(T) + \sum_{c=1}^C (C - c) F_i^{(c)}(T) \right)}{CN}$$

As in the proof of Theorem 1:

1. we will place ourselves on the event where all vertices have bounded degrees, smaller than N^β for some small $\beta > 0$
2. we will stop the analysis at N^γ steps of the horizon T so that $\widehat{F}(k) + \widehat{M}(k) > N^\gamma$ with arbitrarily high probability
3. we will distinguish the cases where $\widehat{M}(k) > 2N^\theta$ (with $\theta = \beta + \gamma/2$)

As a consequence, the errors are going to be of the same order of magnitude with the same order of probability (up to a multiplicative factor C) (hence those computations are skipped and replaced by $\mathcal{O}(\cdot)$ notations). The interesting new component in this proof is the new system of differential equations and their solutions.

E.1 The Differential equations

Using the same notations than in the proof of Theorem 1, we get that for all $0 \leq k \leq T$, $i \geq 0$ and $c \leq C$,

$$\begin{aligned}
& \mathbf{E}[F_i^{(c)}(k+1) - F_i^{(c)}(k) \mid \mathcal{F}_k] \\
&= \mathbf{E}_{d_k \sim \pi_\nu} \left[\sum_{x=1}^{d_k} -q^{x-1} \frac{iF_i^{(c)}}{\widehat{F} + \widehat{M}} + p \sum_{x=1}^{d_k} q^{x-1} (d_k - x) \left(\frac{(i+1)F_{i+1}^{(c)} - iF_i^{(c)}}{\widehat{F} + \widehat{M}} \right) \right] \\
&\quad + \mathbf{E}_{d_k \sim \pi_\nu} \left[\sum_{x=1}^{d_k} q^{x-1} \frac{(i+1)F_{i+1}^{(c+1)}}{\widehat{F} + \widehat{M}} \right] + \mathcal{O}(N^{\theta-\gamma}) \\
&= \frac{\mu_\nu \left(-iF_i^{(c)} + (i+1)F_{i+1}^{(c)} \right) - (i+1)h(q)F_{i+1}^{(c)} + (i+1)h(q)F_{i+1}^{(c+1)}}{\widehat{F} + \widehat{M}} + \mathcal{O}(N^{\theta-\gamma})
\end{aligned}$$

where the function h is still defined as $h(q) = \frac{1-\phi_\nu(q)}{1-q}$. Similarly, we can compute the expected increment in M_i as

$$\begin{aligned}
& \mathbf{E}[M_i(k+1) - M_i(k) \mid \mathcal{F}_k] \\
&= \mathbf{E}_{d_k \sim \pi_\nu} \left[\left(\frac{d_k(1-p)^{d_k}}{\widehat{M}} + \sum_{x=1}^{d_k} pq^{x-1} \left(\frac{(x-1)}{\widehat{M}} + \frac{d_k-x}{\widehat{M} + \widehat{F}} \right) \right) ((i+1)M_{i+1} - iM_i) \right] \\
&\quad + \mathbf{E}_{d_k \sim \pi_\nu} \left[\sum_{x=1}^{d_k} pq^{x-1} \frac{(i+1)F_{i+1}^{(1)}}{\widehat{F}} \right] + \mathcal{O}(N^{\theta-\gamma}) \\
&= \frac{\mu_\nu((i+1)M_{i+1} - iM_i) + (i+1)h(q)F_{i+1}^{(1)}}{\widehat{F} + \widehat{M}} + \mathcal{O}(N^{\theta-\gamma})
\end{aligned}$$

From this, we get the following system of differential equations:

$$\partial_t f^{(c)}(t, s) = \frac{1}{\mu_\mathcal{U} - t\mu_\nu} \left[(-\mu_\nu s + \mu_\nu - h(q(t))) \partial_s f^{(c)}(t, s) + \frac{1}{\mu_\mathcal{U} - t\mu_\nu} h(q(t)) \partial_s f^{(c+1)}(t, s) \right], \quad (15)$$

and

$$\partial_t m(t, s) = \frac{1}{\mu_\mathcal{U} - t\mu_\nu} \left[(-\mu_\nu s + \mu_\nu) \partial_s m(t, s) + h(q(t)) \partial_s f^{(1)}(t, s) \right] \quad (16)$$

With those notations, the normalized performances of GREEDY rewrite then into:

$$\begin{aligned}
A &= m\left(\frac{\mu_\mathcal{U}}{\mu_\nu}, 1\right) + \sum_{c=1}^C \left(1 - \frac{c}{C}\right) f^{(c)}\left(\frac{\mu_\mathcal{U}}{\mu_\nu}, 1\right) \\
&= 1 - \sum_{c=1}^C \frac{c}{C} f^{(c)}\left(\frac{\mu_\mathcal{U}}{\mu_\nu}, 1\right)
\end{aligned}$$

E.2 Solving the PDEs

As in the previous section, we start with a time change. Let

$$\theta(t) = \frac{\mu u}{\mu v} (1 - e^{-\mu v t}) \quad (17)$$

so that $\theta'(t) = \mu u - \theta(t)\mu v$. In order to simplify notations, we set:

$$H(t) := h(q(\theta(t))). \quad (18)$$

Then, the new functions

$$g^{(c)}(t, s) := f^{(c)}(\theta(t), s) \quad \text{and} \quad o(t, s) := m(\theta(t), s)$$

satisfy the following PDEs:

$$\partial_t g^{(c)}(t, s) = [-\mu v s + \mu v - H(t)] \partial_s g^{(c)}(t, s) + H(t) \partial_s g^{(c+1)}(t, s),$$

and

$$\partial_t o(t, s) = [-\mu v s + \mu v] \partial_s o(t, s) + H(t) \partial_s g^{(1)}(t, s). \quad (19)$$

We distinguish:

$$\partial_t g^{(C)}(t, s) = [-\mu v s + \mu v - H(t)] \partial_s g^{(C)}(t, s) \quad (20)$$

We define:

$$F(t) = \int_0^t e^{-\mu v u} H(u) du$$

Solution of (20). This equation is the same as the one satisfied by $g(t, s)$, with the same initial conditions. Thus, we can write:

$$g^{(C)}(t, s) = \phi_u((s-1)e^{-\mu v t} + 1 - F(t)).$$

Solution of (20). Lets define the curves:

$$s_{t,s}(u) = [(s-1)e^{-\mu v t} - F(t) + F(u)] e^{\mu v u} + 1.$$

Along those curves, we have:

$$\frac{d}{dt} g^{(c)}(u, s_{t,s}(u)) = H(u) \partial_s g^{(c+1)}(u, s_{t,s}(u)).$$

So:

$$g^{(c)}(t, s) = \int_0^t H(u) \partial_s g^{(c+1)}(u, s_{t,s}(u)) du$$

Solution for $c = C - 1$. We have:

$$\begin{aligned} g^{(C-1)}(t, s) &= \int_0^t H(u) \partial_s g^{(C)}(u, s_{t,s}(u)) du \\ &= \int_0^t F'(u) \phi'_U((s-1)e^{-\mu v u} + 1 - F(u)) du \\ &= F(t) \phi'_U((s-1)e^{-\mu v t} + 1 - F(t)) \end{aligned}$$

Solution for $c = C - k$, general formula. We will prove by induction:

$$g^{(C-k)}(t, s) = \frac{1}{k!} (F(t))^k \phi^{(k)}((s-1)e^{-\mu v t} + 1 - F(t))$$

If it is true for rank k , we have:

$$\partial_s g^{(C-k)}(u, s_{t,s}(u)) = \frac{e^{-\mu v u}}{k!} (F(u))^k \phi^{(k+1)}((s-1)e^{-\mu v t} + 1 - F(t))$$

Which gives:

$$\begin{aligned} g^{(C-(k+1))}(t, s) &= \frac{1}{k!} \left(\int_0^t F'(u) (F(u))^k du \right) \phi^{(k+1)}((s-1)e^{-\mu v t} + 1 - F(t)) \\ &= \frac{1}{(k+1)!} (F(t))^{k+1} \phi^{(k+1)}((s-1)e^{-\mu v t} + 1 - F(t)) \end{aligned}$$

Solution of (19). Let's define the curves:

$$\gamma_{s,t}(u) = 1 + (s-1)e^{-\mu\nu(t-u)}$$

Along those curves:

$$\frac{d}{du}o(u, \gamma_{t,s}(u)) = H(u)\partial_s g^{(1)}(u, \gamma_{t,s}(u))$$

So:

$$o(t, s) = \int_0^t F'(u) \frac{(F(u))^{(C-1)}}{(C-1)!} \phi^{(C)}((s-1)e^{-\mu\nu t} + 1 - F(u)) du$$

Formula for GREEDY performances. Recall that the normalized performances of GREEDY are

$$\begin{aligned} A &= 1 - \sum_{c=1}^C \frac{c}{C} g^{(c)}(+\infty, 1) \\ &= 1 - \sum_{k=0}^{C-1} \frac{1 - \frac{k}{C}}{k!} (F(+\infty))^k \phi^{(k)}(1 - F(+\infty)) \end{aligned}$$

E.3 ODE for F

We have as before:

$$F'(t) = H(t)e^{-\mu\nu t}, \quad H(t) = \frac{1 - \phi_V(Q(t))}{1 - Q(t)}$$

And we also have:

$$Q(t) = \frac{\partial_s o(t, 1)}{\partial_s o(t, 1) + \sum_{c=1}^C \partial_s g^{(c)}(t, 1)}$$

According to the previous section :

$$\begin{aligned} \partial_s o(t, 1) &= \left(\int_0^t F'(u) \frac{(F(u))^{(C-1)}}{(C-1)!} \phi_U^{(C+1)}(1 - F(u)) du \right) e^{-\mu\nu t} \\ &= \left(\int_0^{F(t)} \frac{x^{(C-1)}}{(C-1)!} \phi_U^{(C+1)}(1 - x) dx \right) e^{-\mu\nu t} \\ &= \left[\phi'_U(1) - \phi'_U(1 - F(t)) - \sum_{k=1}^{C-1} \frac{F(t)^k}{k!} \phi_U^{(k+1)}(1 - F(t)) \right] e^{-\mu\nu t} \end{aligned}$$

Which gives:

$$Q(t) = 1 - \frac{1}{\mu U} \left(\phi'_U(1 - F(t)) + \sum_{k=1}^{C-1} \frac{F(t)^k}{k!} \phi_U^{(k+1)}(1 - F(t)) \right)$$

We define:

$$\Gamma_U(F(t)) = \frac{1}{\mu U} \left(\phi'_U(1 - F(t)) + \sum_{k=1}^{C-1} \frac{F(t)^k}{k!} \phi_U^{(k+1)}(1 - F(t)) \right)$$

This yields the following differential equation for F :

$$\frac{\Gamma_U(F(t))}{1 - \phi_V(1 - \Gamma_U(F(t)))} F'(t) = e^{-\mu\nu t}.$$

Theorem 2 then follows from the same arguments in the proof of Theorem 1 (except that errors are C times bigger as there are C more equations to handle).

F Proof of Theorem 3

As mentioned in the main text, the only difference with Theorem 2 is that C could be of the order of N^β (but not bigger on the event where all degrees are smaller than N^β). As a consequence, one must take β even smaller than $1/20$ to have sublinear errors terms (choosing $\beta = 1/40$ is admissible for instance) with exponentially high probability.

Solution of (20). This equation is the same as the one satisfied by $g(t, s)$, the new initial condition is $g^{(C)}(t, s) = p_C \phi_U(s)$. Thus, we can write:

$$g^{(C)}(t, s) = p_C \phi_U((s-1)e^{-\mu\nu t} + 1 - F(t)).$$

Solution for $c = C - 1$. We have:

$$\begin{aligned} g^{(c-1)}(t, s) &= \int_0^t H(u) \partial_s g^{(C)}(u, s_{t,s}(u)) du + g^{(c-1)}(0, s_{t,s}(0)) \\ &= \int_0^t F'(u) \phi'_U((s-1)e^{-\mu\nu t} + 1 - F(t)) du + p_{(C-1)} \phi_U((s-1)e^{-\mu\nu t} - F(t) + 1) \\ &= p_C F(t) \phi'_U((s-1)e^{-\mu\nu t} + 1 - F(t)) + p_{(C-1)} \phi_U((s-1)e^{-\mu\nu t} + 1 - F(t)) \end{aligned}$$

Solution for $c = C - 2$.

$$\partial_s g^{(C-2)}(u, s_{t,s}(u)) = p_C e^{-\mu\nu u} F(u) \phi''_U((s-1)e^{-\mu\nu t} + 1 - F(t)) + p_{(C-1)} e^{-\mu\nu u} \phi'_U(s_{t,s}(u))$$

Let's define:

$$c(t, s) = (s-1)e^{-\mu\nu t} + 1 - F(t)$$

Which gives:

$$\begin{aligned} g^{(C-2)}(t, s) &= p_{(C-1)} \left(\int_0^t F'(u) F(u) du \right) \phi''(c(t, s)) \\ &\quad + p_{(C-1)} \left(\int_0^t F'(u) e^{\mu\nu u} du \right) \phi'(c(t, s)) + p_{(C-2)} \phi_U(s_{t,s}(0)) \\ &= \frac{p_C}{2} (F(t))^2 \phi''((s-1)e^{-\mu\nu t} \\ &\quad + 1 - F(t)) + p_{(C-1)} F(t) \phi'((s-1)e^{-\mu\nu t} + 1 - F(t)) + p_{(C-2)} \phi_U(c(t, s)) \end{aligned}$$

Solution for $c = C - k$, general formula. We will prove by induction:

$$g^{(C-k)}(t, s) = \sum_{l=0}^k p_{C-l} \frac{1}{(k-l)!} (F(t))^{k-l} \phi^{(k-l)}(c(t, s))$$

If it is true for rank k , we have:

$$\partial_s g^{(C-k)}(u, s_{t,s}(u)) = \sum_{l=0}^k p_{C-l} \frac{1}{(k-l)!} (F(t))^{k-l} e^{-\mu\nu u} \phi^{(k+1-l)}(c(t, s))$$

Which gives:

$$\begin{aligned} g^{(C-(k+1))}(t, s) &= p_{(C-(k+1))} \phi_U(c(t, s)) + \sum_{l=0}^k p_{(C-l)} \frac{1}{(k-l)!} \left(\int_0^t F'(u) (F(u))^{(k-l)} du \right) \phi^{(k+1-l)}(c(t, s)) \\ &= \sum_{l=0}^{k+1} p_{(C-l)} \frac{1}{(k+1-l)!} (F(t))^{k+1-l} \phi^{(k+1-l)}(c(t, s)) \end{aligned}$$

Solution of (19).

$$o(t, s) = \sum_{c=1}^C p_c \int_0^t F'(u) \frac{(F(u))^{(c-1)}}{(c-1)!} \phi^{(c)}((s-1)e^{-\mu\nu t} + 1 - F(u)) du$$

$$g^{(c)}(t, s) = \sum_{k=0}^{C-c} p_{c+k} \frac{1}{k!} (F(t))^k \phi^k(c(t, s))$$

Quantity of interest.

$$A = \frac{\mu\nu}{\mu\mathcal{U}} \sum_{c=1}^C c(p_c - g^{(c)}(+\infty, 1))$$

$$= \frac{\mu\nu}{\mu\mathcal{U}} \left(\sum_{c=1}^C cp_c - \sum_{k=0}^{C-1} \left(\frac{1}{k!} (F(+\infty))^k \phi^{(k)}(1 - F(+\infty)) \sum_{c=1}^C cp_{c+k} \right) \right)$$

ODE for the function F.

$$\partial_s o(t, 1) = \left(\sum_{c=1}^C p_c \int_0^t F'(u) \frac{(F(u))^{(c-1)}}{(c-1)!} \phi_{\mathcal{U}}^{(c+1)}(1 - F(u)) du \right) e^{-\mu\nu t}$$

$$= \left[\phi'_{\mathcal{U}}(1) - \phi'_{\mathcal{U}}(1 - F(t)) + \sum_{k=1}^{C-1} \left(\frac{F(t)^k}{k!} \phi_{\mathcal{U}}^{(k+1)}(1 - F(t)) \sum_{c=k+1}^C p_c \right) \right] e^{-\mu\nu t}$$

Which yields:

$$Q(t) = 1 - \frac{1}{\mu\mathcal{U}} \left(\phi'_{\mathcal{U}}(1 - F(t)) + \sum_{k=1}^{C-1} \left(\frac{F(t)^k}{k!} \phi_{\mathcal{U}}^{(k+1)}(1 - F(t)) \sum_{c=k+1}^C p_c \right) \right)$$

We define:

$$\Gamma_U(F(t)) = \frac{1}{\mu\mathcal{U}} \left(\phi'_{\mathcal{U}}(1 - F(t)) + \sum_{k=1}^{C-1} \left(\frac{F(t)^k}{k!} \phi_{\mathcal{U}}^{(k+1)}(1 - F(t)) \sum_{c=k+1}^C p_c \right) \right)$$

This yields the following differential equation for F :

$$\frac{\Gamma_U(F(t))}{1 - \phi_{\mathcal{V}}(1 - \Gamma_U(F(t)))} F'(t) = e^{-\mu\nu t}.$$

G Proof of Proposition 1

Lemma 4. *On the 2-regular graph, the law of the matches generated by the algorithm Ranking equals the law of the matches generated by a biased Greedy algorithm, that chooses a free vertex of degree 2 over one of degree 1 with probability at least 2/3. This is biased as the classical Greedy algorithm chooses it with probability 1/2.*

Proof: Two vertices of the same degree are interchangeable, they are both equally likely to have the smallest rank. Thus Ranking and Greedy behave the same on arriving vertices whose potential neighbors all have the same degree. Let $r(v)$ be the rank of vertex v and $deg(v)$ its residual number of unpaired half-edges.

Let $\mathcal{A}_t(u, 2)$ be the following event:

- u had one of its half-edges paired to the incoming vertex v_t at iteration t ,

- u was not matched and v_t was instead matched to a vertex of residual degree 2.

Let $\mathcal{A}_t(u, 1)$ be the similar event with v_t instead matched to a vertex of residual degree 1.

$$\begin{aligned} \mathbf{P}(r(u) \geq k | \deg(u) = 1 \text{ and } u \text{ free at } t) &= \sum_{t' < t} \mathbf{P}(r(u) \geq k | \mathcal{A}_{t'}(u, 2)) \mathbf{P}(\mathcal{A}_{t'}(u, 2) | \deg(u) = 1 \text{ and } u \text{ free at } t) \\ &\quad + \sum_{t' < t} \mathbf{P}(r(u) \geq k | \mathcal{A}_{t'}(u, 1)) \mathbf{P}(\mathcal{A}_{t'}(u, 1) | \deg(u) = 1 \text{ and } u \text{ free at } t) \end{aligned}$$

Now, assume

$$\forall t' < t, \forall a \in [N], \mathbf{P}(r(a) \geq k | \deg(a) = 1 \text{ and } a \text{ free at } t) \geq \mathbf{P}(r(a) \geq k | \deg(a) = 2). \quad (21)$$

Hypothesis 21 implies:

$$\forall t' < t, \mathbf{P}(r(u) \geq k | \mathcal{A}_{t'}(u, 1)) \geq \mathbf{P}(r(u) \geq k | \mathcal{A}_{t'}(u, 2)),$$

thus, the following inequality holds:

$$\mathbf{P}(r(u) \geq k | \deg(u) = 1 \text{ and } u \text{ free at } t) \geq \mathbf{P}(r(u) \geq k | \cup_{t' < t} \mathcal{A}_{t'}(u, 2)).$$

Let a and b be two different numbers randomly chosen in $[n]$. Two vertices with two remaining half-edges were not affected by the run before, and thus could have any rank. It therefore holds:

$$\begin{aligned} \mathbf{P}(r(u) \geq k | \deg(u) = 1 \text{ and } u \text{ free at } t) &\geq \mathbf{P}(\max(a, b) \geq k) \\ &= 1 - \frac{\binom{k-1}{2}}{\binom{n}{2}} \\ &= 1 - \frac{(k-1)(k-2)}{n(n-1)}. \end{aligned}$$

This inequality implies $\mathbf{P}(r(u) \geq k | \deg(u) = 1 \text{ and } u \text{ free at } t) \geq \mathbf{P}(r(u) \geq k | \deg(u) = 2)$, thus 21 is true by induction.

Therefore, vertices with only one remaining half-edges are likely to have a higher rank than those with two remaining half-edges:

$$\begin{aligned} \mathbf{P}(r(b) < r(a) | \deg(b) = 2, \deg(a) = 1) &= \sum_{k=1}^n \mathbf{P}(\deg = k-1, \deg(a) \geq k | \deg(b) = 2, \deg(a) = 1) \\ &\geq 1 - \sum_{k=1}^n \frac{(k-1)(k-2)}{n^2(n-1)} \\ &\geq \frac{2}{3} + O\left(\frac{1}{n}\right). \end{aligned}$$

□

Let $M_1^G(t)$ and $M_1^R(t)$ be the number of marked vertices of degree 1 by GREEDY and RANKING algorithms respectively. Note that the number of vertices of degree 2 is the same for both algorithms, $F_2^G(t) = F_2^R(t)$. Also, the following always holds

$$F_1^G(t) = 2N - 2t - 2F_2^G(t) - M_1^G(t).$$

This shows that in 2-regular graphs, the number of half-edges in each ensemble is a deterministic quantity of $M_1^R(t+1)$ and $M_1^G(t+1)$.

Suppose it holds at time t that $M_1^G(t) = M_1^R(t) = M_1(t)$ (event \mathcal{A}), then

$$\begin{aligned} \mathbb{E}[M_1^R(t+1)|\mathcal{A}] - \mathbb{E}[M_1^G(t+1)|\mathcal{A}] &= \mathbb{E}[\mathbb{1}_{\{\text{RANKING marks a vertex in } F_2^R(t)\}}|\mathcal{A}] - \mathbb{E}[\mathbb{1}_{\{\text{GREEDY marks a vertex in } F_2^R(t)\}}|\mathcal{A}] \\ &= \frac{1}{6} \cdot \frac{F_1(t)F_2(t)}{2(N-t)} > 0 \end{aligned}$$

The first equality holds since the probability of pairing an half-edge in $M_1^R(t+1)$ and $M_1^G(t+1)$ only depends on the number of half edges in each ensembles, not on the algorithm.

Therefore, by application of Gronwald's lemma, RANKING generates strictly more marked vertices of degree 1. As the probability that an incoming vertice is matched only to non-available vertices increases with M_1 , RANKING performs strictly worse than GREEDY on 2-regular graphs.