

# Subtractive random forests with two choices

Francisco Calvillo

LPSM, Sorbonne Université, 4 Place Jussieu, 75005 Paris, France

Luc Devroye

School of Computer Science, McGill University,  
Montreal, Canada

Gábor Lugosi \*

Department of Economics and Business,  
Pompeu Fabra University, Barcelona, Spain  
ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain  
Barcelona Graduate School of Economics

April 21, 2026

## Abstract

Recommendation systems are pivotal in aiding users amid vast online content. Broutin, Devroye, Lugosi, and Oliveira proposed Subtractive Random Forests (*SURF*), a model that emphasizes temporal user preferences. Expanding on *SURF*, we introduce a model for a multi-choice recommendation system, enabling users to select from two independent suggestions based on past interactions. We evaluate its effectiveness and robustness across diverse scenarios, incorporating heavy-tailed distributions for time delays. By analyzing user topic evolution, we assess the system's consistency. Our study offers insights into the performance and potential enhancements of multi-choice recommendation systems in practical settings.

**Keywords:** subtractive random forests, multi-choice recommendation

---

\*Corresponding author; email: gabor.lugosi@gmail.com

---

## 1 Introduction

Today’s online platforms have a record number of items available for their users. In many cases, presenting them as a catalogue is no longer a good option and can make it difficult for the user to find what they are interested in. Recommendation systems have become an essential tool to cope with the overload of information available on the web.

A well-known way to approach recommendation systems today is through deep learning, and many of the most effective recommendation systems use this principle, see [14]. However, the mechanism of these systems is notable for its opacity. In the realm of online sequential recommendation systems, Broutin, Devroye, Lugosi, and Oliveira [6] present an approach, offering a framework that considers the temporal aspect of user preferences by a simple mechanism. They define a model recommending topics based on a random time delay. The topic that is recommended at time  $n$  is the same that was recommended at time  $n - Z_n$ , where  $(Z_k)_{k \geq 1}$  is a sequence of i.i.d. random variables identically distributed on  $\{1, 2, \dots\}$ . This approach led them to study a family of random forests called subtractive random forests (SURF), allowing a detailed structural study. However, one expects a recommendation system to make several recommendations at a time, not just one.

To model this more realistic scenario, our paper proposes a two-choice recommendation system inspired by the same idea. We envision a scenario where users are presented with two independent recommendations, drawn from the same temporal recommendation mechanism and allowing users to select the most appealing option. The resulting model has some intriguing mathematical properties, and the main goal of this paper is to analyze the model in order to understand the long-term behavior of such recommendation systems.

While our model provides a mathematically tractable framework to study sequential recommendation dynamics, it is intentionally abstract and does not incorporate key aspects of modern recommender systems such as content information, user-specific features, or contextual data. This abstraction limits its direct applicability to real-world platforms, where personalization and diversity are cen-

---

tral goals. However, we view this simplification as a necessary step: by first understanding the behavior of a minimal, analytically transparent model, we lay the groundwork for tackling more complex and realistic settings in which such additional factors can be systematically incorporated.

Beyond serving as a recommender system in its own right, our model can also be viewed as a theoretical tool to gain insight into more complex algorithms. In particular, it offers a simple framework to reason about phenomena such as the repeated occurrence of topics in classical recommender systems, where machine learning mechanisms are often treated as black boxes. By studying this simplified yet transparent model, we can uncover structural patterns that would otherwise remain hidden, offering insight into the dynamics of more sophisticated recommendation methods.

### 1.1 A two-choice recommendation model

We assume that the initial pool of topics is infinite and represented by the set of non-positive integers  $\{0, -1, -2, \dots\}$ . We now consider independent and identically distributed random variables  $Z, W, (Z_n)_{n \geq 1}$  and  $(W_n)_{n \geq 1}$  on the set of positive integers  $\mathbb{N}$ . Define

$$q_i = \mathbb{P}(Z = i) \quad \text{and} \quad p_i = \mathbb{P}(Z \geq i), \quad i \geq 1 .$$

Each topic  $i \leq 0$  is assigned a preference value  $U_i$  within the range  $[0, 1]$ , where we assume that  $(U_n)_{n \leq 0}$  is a sequence of random variables independent of the sequences  $(Z_n)_{n \geq 1}$  and  $(W_n)_{n \geq 1}$ . In the context of our model, a user is a representative individual interacting with the system, and the values  $U_i$  represent how appealing each topic is to this user.

Following [6], we can define a sequential random colouring of the positive integers as follows. For each non-positive integer  $i \leq 0$ , define its colour  $C_i = i$  and its preference value  $V_i = U_i$ . We assign colours and preference values to the

---

positive integers  $n \in \mathbb{N}$  by the recursion

$$\begin{cases} C_n = C_{n-Z_n} \text{ and } V_n = V_{n-Z_n} & \text{if } V_{n-Z_n} < V_{n-W_n} \\ C_n = C_{n-W_n} \text{ and } V_n = V_{n-W_n} & \text{otherwise .} \end{cases}$$

Thus, by identifying the colour  $C_n$  as the topic recommended at time  $n \geq 1$ , this definition means that at the time instant  $n \geq 1$ , the user receives two recommendations (the topic  $C_{n-Z_n}$ , and the topic  $C_{n-W_n}$ ), and chooses the one with the lowest (best) preference value.

This process naturally defines a random directed graph whose vertex set is  $\mathbb{Z}$  by drawing an edge from any positive integer  $n \geq 1$  to any integer  $m < n$  if and only if  $m = n - Z_n$  or  $m = n - W_n$ . Vertices with negative index are called *leaves*, as is customary for final nodes in dags (i.e., directed acyclic graphs). Let  $T_n$  denote the set of vertices that are reached from the vertex  $n \geq 1$ . This set can be defined recursively by saying that  $T_n = \{n\}$  for any  $n \leq 0$ , and

$$T_n = \{n\} \cup T_{n-Z_n} \cup T_{n-W_n}$$

for any  $n \geq 1$ . Define

$$\mathcal{L}_n = T_n \cap \{0, -1, -2, \dots\}$$

as the set of vertices with a non-positive index that is reached from  $n$ . In other words,  $\mathcal{L}_n$  is the set of leaves in the subgraph of vertices reached from  $n$ . Note that

$$C_n = \operatorname{argmin}_{i \in \mathcal{L}_n} U_i \quad \text{and} \quad V_n = \min_{i \in \mathcal{L}_n} U_i .$$

One can assess the system's long-term performance by analyzing the asymptotic behavior of the sequence  $(V_n)_{n \geq 1}$ . If  $V_n$  approaches  $\inf_{i \leq 0} U_i$  as  $n \rightarrow \infty$ , then this means that after waiting a sufficient amount of time, the topics recommended to the user tend to align more closely with their preferences. This gives us a consistency criterion for our model. We consider the three following configurations of the preference values:

- (i)  $U_0 < U_{-1} < U_{-2} < U_{-3} \dots$

---

(ii)  $U_0 > U_{-1} > U_{-2} > U_{-3} \dots$

(iii)  $(U_n)_{n \leq 0}$  i.i.d. and uniformly distributed in  $[0, 1]$ .

These assumptions correspond to natural scenarios and allow us to study the long-term consistency of the recommendation system. In the first case, topics have a monotone preference with the most recent (corresponding to  $i = 0$ ) being the most preferred one. In the second case, older topics are preferred, while in the third case, topics have a random order of preference. We say that the system is consistent if, in the long run, it offers near-optimal recommendations to the user in terms of their preference. In Section 1.3, we rigorously define various notions of consistency, corresponding to the scenarios above. These definitions lead us to study the sequence  $(V_n)_{n \geq 1}$  by studying the set of leaves  $\mathcal{L}_n$ . We will pay particular attention to the three following random variables:

- $R_n = \max\{i : i \in \mathcal{L}_n\}$  (the rightmost leaf in  $\mathcal{L}_n$ )
- $L_n = \min\{i : i \in \mathcal{L}_n\}$  (the leftmost leaf in  $\mathcal{L}_n$ )
- $\Lambda_n = |\mathcal{L}_n|$  (the number of leaves reached from  $n$ ).

Each random variable helps us understand the long-term behavior of  $V_n$  in the three initial configurations described above. Ideally, one would hope that  $V_n \rightarrow \inf_{i \leq 0} U_i$ . In configuration (i) this is equivalent to  $R_n \rightarrow 0$ , in configuration (ii) it is equivalent to  $L_n \rightarrow -\infty$ , while in case (iii) to  $\Lambda_n \rightarrow \infty$ . However, one cannot expect to have  $R_n \rightarrow 0$  in a general case. To introduce a more reasonable consistency criterion for case (i), we require boundedness of the sequence  $(R_n)_{n \geq 1}$ .

Let us define the subsets  $T_n^Z = T_n^W = S_n = \{n\}$  for any  $n \leq 0$ , and let

$$T_n^Z = \{n\} \cup T_{n-Z_n}^Z,$$

$$T_n^W = \{n\} \cup T_{n-W_n}^W,$$

$$\text{and } S_n = \{n\} \cup S_{n-\min(Z_n, W_n)}$$

for any  $n \geq 1$ , so that  $T_n^Z \subset T_n$  (resp.  $T_n^W \subset T_n$ ) is the set of vertices that can be reached from  $n$  by following only the  $Z$ -edges (resp.  $W$ -edges), and  $S_n \subset T_n$  is the set of vertices that can be reached from  $n$  by following only the shortest edges.

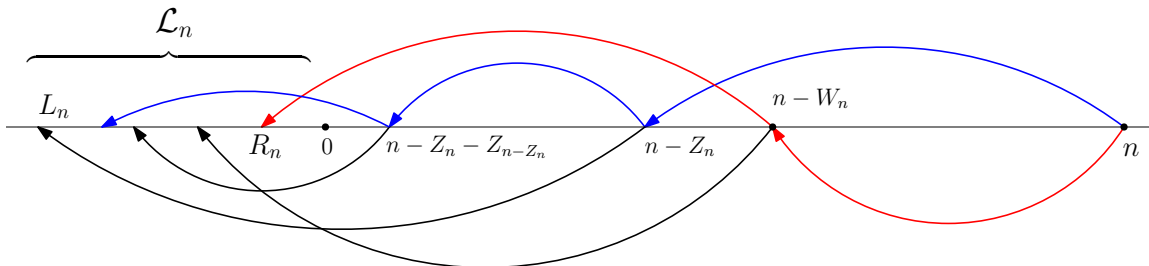


Figure 1: Illustration of the subgraph induced by  $T_n$ . Shown are the definitions of  $L_n$ ,  $R_n$ , and  $\mathcal{L}_n$ . Blue edges indicate the  $Z$ -chain of vertices  $T_n^Z$ , while red edges connect all vertices in the chain  $S_n$ .

Finally, for any  $n \geq 1$ , let  $r_n = \mathbb{P}(0 \in T_n^Z) = \mathbb{P}(0 \in T_n^W)$  denote the probability that vertex  $0$  belongs to the  $Z$ -chain starting at vertex  $n$ . We set  $r_0 = 1$  so that  $r_n$  satisfies the recursion

$$r_n = \sum_{i=1}^n q_i r_{n-i}. \quad (1.1)$$

For convenience, a table summarizing the main notation used throughout the paper is provided in the Appendix.

**Remark 1.** Observe that the chains  $T_n^Z$ ,  $T_n^W$ , and  $S_n$  can be naturally interpreted as a renewal process, where each step along the chain corresponds to a renewal event, and the edge lengths play the role of renewal intervals. Within this framework, the quantity  $r_n$  coincides with the renewal measure of the process corresponding to  $T_n^Z$ .

## 1.2 Related work

This paper builds on the work of Broutin, Devroye, Lugosi and Oliveira [6], who studied a one-choice version of this model by examining the properties of a graph whose vertex set is  $\mathbb{Z}$ , and whose edges are  $\{n, n - Z_n\}$  for every  $n \geq 1$ , given some sequence  $(Z_n)_{n \geq 1}$  of i.i.d. random variables distributed in  $\{1, 2, 3, \dots\}$ . This defines a random forest where each non-positive integer is the root of a tree. They also

---

define a random coloring by setting  $C_i = i$  for all  $i \leq 0$ , and  $C_n = C_{n-Z_n}$  for all  $n \geq 1$ . Each tree in the forest corresponds to a colour (i.e., a topic). The graph that we consider in our paper is nothing else than the superposition of two independent copies of this forest, and the sets  $T_n^Z$  and  $T_n^W$  defined above are the chains that connect the vertex  $n$  to the root of its tree in each of these two copies.

One of the main results in [6] shows that the process has two distinct behaviors: if  $\mathbb{E}Z = \infty$ , then, almost surely, all trees in the forest are finite, meaning that no topic will be recommended infinitely often, while if  $\mathbb{E}Z < \infty$ , after some random amount of time, all vertices connect to the same tree almost surely, meaning that all topics recommended to the user become the same. In the two-choice model however, there are three distinct regimes:  $\mathbb{E}Z < \infty$ ;  $\mathbb{E}Z = \infty$  yet  $\mathbb{E}\min(Z, W) < \infty$ ; and  $\mathbb{E}\min(Z, W) = \infty$ . We call these the light-, moderate-, and heavy-tailed regimes.

The random forest in [6] appears as a subgraph of a random graph model previously studied by Hammond and Sheffield in [10]. Given a sequence  $(Z_n)_{n \in \mathbb{Z}}$  of independent random variables, identically distributed in  $\{1, 2, 3, \dots\}$ , they consider a graph with vertex set  $\mathbb{Z}$  such that vertices  $n, m \in \mathbb{Z}$ , with  $m < n$ , are connected by an edge if and only if  $m = n - Z_n$ . One can obtain the random forest in [6] by removing all edges between any two vertices with a non-positive index. They show that the random graph has almost surely a unique component when the sum  $\sum_{n=1}^{\infty} r_n^2$  converges, and the graph almost surely has infinitely many connected components when the sum diverges.

Other variants of the (single-choice) random subtractive process have been analyzed in related probabilistic models. Blath, González, Kurt, and Spano [4] studied long-range seed bank models, where the genealogical structure of populations is governed by random time jumps analogous to the subtractive mechanism. Chierichetti, Kumar, and Tomkins [8] examined sequence models whose asymptotic behavior shares similarities with renewal-type processes underlying our setting. Baccelli, Haji-Mirsadeghi, Khezeli, and Sodre [3, 2, 1] explored subtractive dynamics in the context of unimodular random graphs and population processes. Finally, Igelbrink and Wakolbinger [12] investigated reinforced urn schemes where coalescence probabilities play a role closely related to recurrence properties of sub-

---

tractive chains.

### 1.3 Results

The goal of this paper is to understand the behavior of the two-choice model by studying its consistency in each of the three configurations (i),(ii), and (iii), for three types of tails, light, moderate, and heavy:

1.  $\mathbb{E}Z < \infty$  (light tails)
2.  $\mathbb{E}Z = \infty$  and  $\mathbb{E}\min(Z, W) < \infty$  (moderate tails)
3.  $\mathbb{E}\min(Z, W) = \infty$  (heavy tails)

We now introduce the main definitions of consistency.

**Definition 1.** *We say that the system is consistent in configuration (ii) if  $L_n \rightarrow -\infty$ ; while it is consistent in configuration (iii) if  $\Lambda_n \rightarrow \infty$ . We differentiate between strong and weak consistency based on whether convergence occurs almost surely or in probability. In the case of configuration (i), we say that the system is weakly consistent if  $(R_n)_{n \geq 1}$  is a tight sequence. It is strongly consistent if the sequence  $(R_n)_{n \geq 1}$  is almost surely bounded.*

**Remark 2. PARETO TAILS.** *Consider a distribution with Pareto tails, that is,  $q_n = \Theta(1/n^{1+\alpha})$ , with  $\alpha > 0$ . This implies that  $p_n = \Theta(1/n^\alpha)$ . When  $\alpha > 1$ , we see that  $\mathbb{E}Z < \infty$ . The more interesting situation is when  $\alpha \in (0, 1]$ . For  $\alpha \in (1/2, 1]$ , we have  $\mathbb{E}Z = \infty$ , yet  $\mathbb{E}\min(Z, W) < \infty$ . Finally, for  $\alpha \in (0, 1/2]$ , we have  $\mathbb{E}\min(Z, W) = \infty$ . For light-tailed  $Z$ , we recall from Broutin, Devroye, Lugosi, Oliveira [6] that  $r_n \rightarrow 1/\mathbb{E}Z$ . For moderate and heavy tails, however, we have  $r_n \rightarrow 0$ . When  $\alpha \in (0, 1)$ , then  $Z$  is in the domain of attraction of the extremal stable law with parameter  $\alpha$ , which itself has a tail distribution function that decays at the rate  $1/n^\alpha$  (see, e.g., Ibragimov and Linnik [11], Zolotarev [15] or Malevich [13]). One can then deduce that  $r_n$  decays at the rate  $1/n^{1-\alpha}$ . In particular,  $\sum_n r_n^2 < \infty$  when  $\alpha \in (0, 1/2)$ , which roughly corresponds to the case of heavy tails.*

Next, we describe the main results of the paper. For a summary, see the table at the end of this section.

---

## Light tails

When  $\mathbb{E}Z < \infty$ , time jumps are so short that  $T_n$  can't reach distant leaves, thus forcing the set  $\mathcal{L}_n$  to be bounded. This case does not require much work, since one can quickly determine consistency or inconsistency in each of the three configurations just by looking at the sequence  $(L_n)_{n \geq 1}$ , which happens to be bounded almost surely.

**Theorem 2.** *Let  $L_\infty = \inf_{n \in \mathbb{N}} L_n$ . If  $\mathbb{E}Z < \infty$ , then  $L_\infty$  is finite almost surely. Thus, the sequences  $(R_n)_{n \in \mathbb{N}}$ ,  $(L_n)_{n \in \mathbb{N}}$  and  $(\Lambda_n)_{n \in \mathbb{N}}$  are almost surely bounded.*

**Proof.** Note that

$$\mathbb{P}(L_\infty \leq -x) \leq 2 \sum_{n=1}^{\infty} \mathbb{P}(n - Z_n \leq -x) = 2 \sum_{n=1}^{\infty} p_{n+x}.$$

Since  $\mathbb{E}Z < \infty$ , we know that  $\sum_{n=1}^{\infty} p_{n+x}$  is finite and goes to 0 as  $x \rightarrow +\infty$ . Thus, by continuity of measure,

$$\mathbb{P}(L_\infty = -\infty) = \lim_{x \rightarrow +\infty} \mathbb{P}(L_\infty \leq -x) = 0.$$

Since  $|L_n| \geq \Lambda_n$  and  $|L_n| \geq |R_n|$ , this also implies that  $(\Lambda_n)_{n \in \mathbb{N}}$  and  $(R_n)_{n \in \mathbb{N}}$  are almost surely bounded. ■

## Moderate and heavy tails

When  $\mathbb{E} \min(Z, W) = \infty$ , a different behavior emerges:

**Theorem 3.** *Let  $Z, W, (Z_n)_{n \geq 1}$  and  $(W_n)_{n \geq 1}$  be i.i.d. random variables. Assume furthermore that  $\mathbb{E} \min(Z, W) = \infty$ . Then, with probability one,*

$$\liminf_{n \rightarrow \infty} \Lambda_n < \infty.$$

*In other words, the system is not strongly consistent in the heavy-tailed regime in configuration (iii).*

---

**Proof.** Note that if  $\min(Z_n, W_n) \geq n$ , then  $\Lambda_n \in \{1, 2\}$ . Thus, to prove Theorem 3, it suffices to show that with probability one,  $\min(Z_n, W_n) \geq n$  happens infinitely often. This follows by the second Borel-Cantelli lemma since the events

$$(\{\min(Z_n, W_n) \geq n\})_{n \geq 0}$$

are independent and

$$\sum_{n \geq 0} \mathbb{P}(\min(Z_n, W_n) \geq n) = \mathbb{E}(\min(Z, W)) = \infty .$$

■

The behavior of the set  $\mathcal{L}_n$  for moderate and heavy tails is better understood by examining its extreme points  $L_n$  and  $R_n$ . The following two theorems are proved in Section 3.

**Theorem 4.** *If  $q_i > 0$  for all  $i \in \mathbb{N}$  and  $\mathbb{E}Z = \infty$ , then, with probability one,*

$$L_n \rightarrow -\infty .$$

*Thus, in configuration (ii) the system is strongly consistent in the moderate and heavy-tailed regimes.*

Strong (in-)consistency in configuration (i) is established in the next result:

**Theorem 5.** *Let  $Z, W, (Z_n)_{n \geq 1}$  and  $(W_n)_{n \geq 1}$  be i.i.d. random variables.*

1. *If  $\mathbb{E}\min(Z, W) < \infty$ , then  $\sup_{n \in \mathbb{N}} |R_n| < \infty$  almost surely.*
2. *If  $\mathbb{E}\min(Z, W) = \infty$ , then  $\sup_{n \in \mathbb{N}} |R_n| = \infty$  almost surely.*

Given that the leftmost leaf in  $\mathcal{L}_n$  goes to  $-\infty$  and the rightmost leaf remains bounded in the moderate heavy-tail regime (i.e., when  $\mathbb{E}Z = \infty$  and  $\mathbb{E}\min(Z, W) < \infty$ ), one can expect that the total number of leaves reached by the vertex  $n$  diverges under these two assumptions. Thus, strong consistency is observed in each of the three configurations (i), (ii), and (iii).

---

On the other hand, only configuration (i) leads to strong consistency when the distribution of  $Z$  has a heavy-tail (i.e., when  $\mathbb{E}\min(Z, W) = \infty$ ). However, when  $\sum_{n=1}^{\infty} r_n^2 < \infty$ , the model remains weakly consistent in configuration (iii), as established by the next two results. Theorems 6 and 7 are proved in Sections 4.1 and 4.2, respectively.

**Theorem 6.** *Assume that the distribution of  $Z$  is aperiodic, that is, the greatest common divisor of the support of  $Z$  is 1. If  $Z$  exhibits a moderate-sized tail (i.e.,  $\mathbb{E}\min(Z, W) < \infty$  and  $\mathbb{E}Z = \infty$ ), then*

$$\Lambda_n \rightarrow \infty \text{ almost surely.}$$

**Theorem 7.** *When  $Z$  has a heavy tail (i.e.,  $\mathbb{E}\min(Z, W) = \infty$ ), and  $\sum_{n \geq 0} r_n^2 < \infty$ , then*

$$\Lambda_n \rightarrow \infty \text{ in probability.}$$

We recall from Remark 2 that the summability of  $r_n^2$  is assured for nearly all heavy-tailed  $Z$ .

## 1.4 Summary

The following table summarizes our findings. Observe that weak and strong behavior coincide in most cases. They only differ in the extreme heavy tail case ( $\mathbb{E}\min(Z, W) = \infty$ ). For an optimal user experience, one would need  $\Lambda_n \rightarrow \infty$  almost surely, and the only case that assures this is when we have moderate tails. In addition, such moderate tails also guarantee strong consistency for the decreasing and increasing input scenarios (as is apparent from the strong consistency of  $L_n$  and  $R_n$ ).

---

	$\mathbb{E}Z < \infty$	$\mathbb{E}Z = \infty$ and $\mathbb{E}\min(Z, W) < \infty$	$\mathbb{E}\min(Z, W) = \infty$
	Strong consistency		
$L_n \rightarrow -\infty$ a.s.	no (Thm. 2)	yes (Thm. 4)	yes (Thm. 4)
$\Lambda_n \rightarrow \infty$ a.s.	no (Thm. 2)	yes (Thm. 6)	no (Thm. 3)
$R_n$ bounded a.s.	yes (Thm. 2)	yes (Thm. 5)	no (Thm. 5)
	Weak consistency		
$L_n \xrightarrow{\mathbb{P}} -\infty$	no (Thm. 2)	yes (Thm. 4)	yes (Thm. 4)
$\Lambda_n \xrightarrow{\mathbb{P}} \infty$	no (Thm. 2)	yes (Prop. 11)	yes if $\sum r_n^2 < \infty$ (Thm. 7)
$R_n$ is tight	yes (Thm. 2)	yes (Thm. 5)	

The rest of the paper contains the proofs of the results stated above. In Section 2 we recall some properties of the single-choice model that are useful in our analysis. In Section 3, Theorems 4 and 5 are proven. The main technical content of the paper is presented in Section 4 where the number of leaves is examined in both the moderate-, and heavy-tailed cases, culminating in the proofs of Theorems 6 and 7.

## 1.5 Simulations

To complement our theoretical analysis, we conducted a series of simulations to examine the empirical behavior of the model. We considered discrete Pareto distributions for the random variables  $Z$  and  $W$ , with probabilities  $q_n = c/n^{1+\alpha}$ , so that varying  $\alpha$  produces light-tailed, moderate-tailed, and heavy-tailed regimes. Our analysis focused on the evolution of the preferential value  $V_n$  of the topic recommended at time  $n$  in configuration (iii), where topics are assigned independent uniform preference values.

Figure 2 presents single realizations of  $V_n$  for each tail regime, illustrating the emergence (or absence) of strong consistency. Figure 3 shows the corresponding averages over 1000 independent simulations, which emphasize the form of weak consistency predicted in the heavy-tailed setting. The simulations reveal three qualitatively distinct behaviors:

- **Light-tailed regime ( $\alpha = 1.2$ ):** no evidence of consistency; recommendations stabilize around a single topic, and  $V_n$  does not converge to 0.
- **Moderate-tailed regime ( $\alpha = 0.6$ ):** strong consistency, with the process converging reliably to a dominant topic.
- **Heavy-tailed regime ( $\alpha = 0.3$ ):** strong inconsistency at the trajectory level, but weak consistency visible in the averaged dynamics, in agreement with Remark 2.

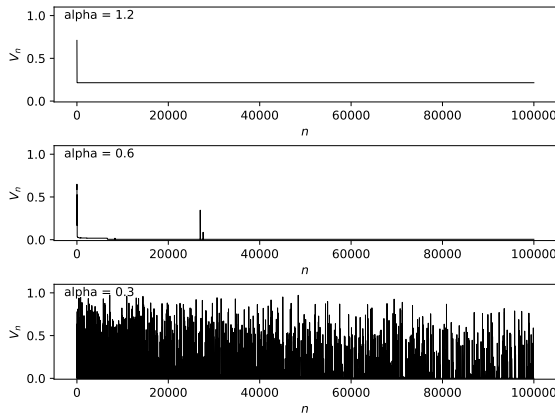


Figure 2: Simulation of the evolution of  $V_n$  in configuration (iii) for a single run across light-tailed, moderate-tailed, and heavy-tailed regimes.

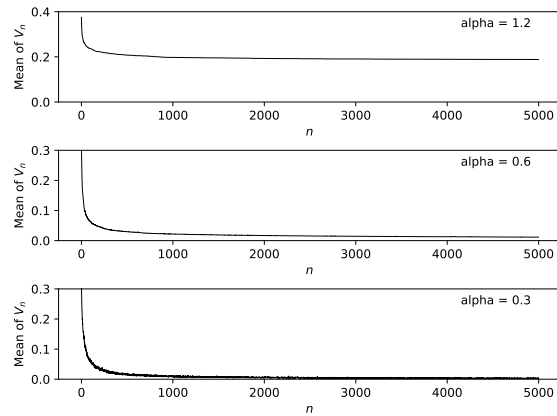


Figure 3: Evolution of the empirical mean of  $V_n$  in configuration (iii) over 1000 independent simulations for light-tailed, moderate-tailed, and heavy-tailed regimes.

Configurations (i) and (ii) were also examined, and the corresponding simulation outcomes are in full agreement with the theoretical results established for these regimes. Since their qualitative behavior is straightforward and consistent with the analytical predictions, the corresponding plots are omitted for brevity.

In addition to tracking the evolution of  $V_n$ , it is also instructive to visualize the subgraph induced by the vertices  $T_n$ , consisting of the vertices reached by vertex  $n$ . Coloring vertices with positive index in blue and vertices with non-positive index (leaves or topics) in red provides a structural perspective on the process. The resulting plots (see Figure 4) make the differences between the three regimes

---

strikingly clear: in the light-tailed case, the short edges yield a dense graph with many blue vertices but very few leaves; in the moderate-tailed case, the number of leaves grows substantially while blue vertices remain numerous; and in the heavy-tailed case, the dominance of long edges produces a sparse structure with very few blue vertices and a predominance of leaves. These observations are fully consistent with our theoretical analysis and complement the intuition provided by the  $V_n$  plots.

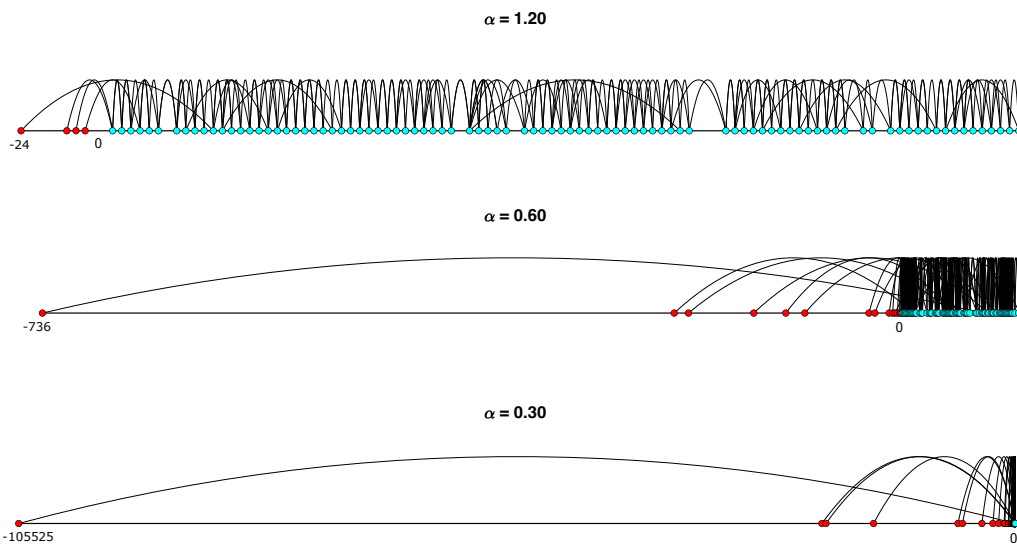


Figure 4: Visualization of the subgraph induced by  $T_n$  in the light-, moderate-, and heavy-tailed regimes, with  $n = 150$ .

## 2 The one-choice model

Some of our proofs use the observation that the graph contains several copies of the subtractive random forest studied in [6]. For example, Theorem 1 in [6] states that there is a unique infinite tree in the forest when  $\mathbb{E}Z < \infty$ :

**Theorem 8** (Broutin, Devroye, Lugosi, Oliveira [6]). *Let  $Z$  and  $(Z_n)_{n \geq 1}$  be i.i.d. random variables. Assume  $\mathbb{E}Z < \infty$  and  $q_1 > 0$ . Then there exists a positive random variable  $N$  with  $\mathbb{P}(N < \infty) = 1$  such that, with probability one,*

$$N \in T_n^Z \quad \text{for all } n \geq N .$$

---

**Remark 3.** Note that the actual statement of the theorem in [6] is slightly weaker, as it only asserts the existence of a random variable  $N$ , finite almost surely, such that for all  $n \geq N$ , the vertex  $n$  has the same colour as vertex  $N$ . This does not necessarily imply that  $N \in T_n^Z$ . Nevertheless, one can add this property of  $N$  without changing the proof provided in [6].

Define  $Z'_n = \min(Z_n, W_n)$  and  $Z' = \min(Z, W)$  and note that  $S_n = T_n^{Z'}$ . By Theorem 8, we obtain the following result for moderate and light tails.

**Corollary 9.** Let  $Z, W, (Z_n)_{n \geq 1}$  and  $(W_n)_{n \geq 1}$  be i.i.d. random variables. If  $\mathbb{E} \min(Z, W) < \infty$  and  $q_1 > 0$ , then there exists a positive random variable  $N$  with  $\mathbb{P}(N < \infty) = 1$  such that, with probability one,

$$N \in S_n \quad \text{for all } n \geq N .$$

When  $\mathbb{E}Z = \infty$ , Theorem 3 in [6] states that every tree in the forest is finite. This translates as follows.

**Theorem 10** (Broutin, Devroye, Lugosi, Oliveira [6]). Let  $Z$  and  $(Z_n)_{n \geq 1}$  be i.i.d. random variables. If  $\mathbb{E}Z = \infty$  and  $q_i > 0$  for all  $i \geq 1$ , then

$$\mathbb{P}(\cup_{i \leq 0} [\{n \geq 1 : i \in T_n^Z\} = \infty]) = 0 .$$

### 3 Rightmost and leftmost leaves

As stated in Theorem 4, the leftmost leaf  $L_n$  reached by vertex  $n$  diverges to  $-\infty$  whenever  $\mathbb{E}Z = \infty$ . This can be seen as a consequence of the behavior of the one-choice model, as one only needs to check the divergence to negative infinity of the leaf  $\min T_n^Z$  reached by vertex  $n$  by following only the  $Z$ -edges.

**Proof.** [Proof of Theorem 4] From Theorem 10 we know that with probability one, for all  $k \leq 0$ , there are at most a finite number of integers  $n \geq 1$  such that  $k \in T_n^Z$ . Thus, the sequence  $(\min T_n^Z)_{n \in \mathbb{N}}$ , taking values in  $\{0, -1, -2, \dots\}$ , cannot take the same value an infinite number of times, so it goes to  $-\infty$  as  $n \rightarrow \infty$  with probability one. This concludes the proof, since  $L_n \leq \min T_n^Z$  for all  $n \geq 1$ . ■

---

**Proof.** [Proof of Theorem 5] Assume that  $\mathbb{E} \min(Z, W) < \infty$ . From Corollary 9, we know that there exists some random variable  $N$  with  $\mathbb{P}(N < \infty) = 1$  such that, with probability one,  $N \in S_n$  for all  $n \geq N$ . In particular, for any  $n \geq N$  we have  $T_N \subset T_n$ , which implies that  $|R_N| \geq |R_n|$ . Thus, with probability one, we have

$$\sup_{n \in \mathbb{N}} |R_n| \leq \max_{1 \leq k \leq N} |R_k| < \infty .$$

When  $\mathbb{E} \min(Z, W) = \infty$ , we have

$$\sum_{n \geq 1} \mathbb{P}(\min(Z_n, W_n) \geq 2n) = \sum_{n \geq 1} p_{2n}^2 = \infty .$$

Hence, by the second Borel-Cantelli lemma, we know that with probability one we have  $\min(Z_n, W_n) \geq 2n$  infinitely often, which implies that  $R_n \leq -n$  infinitely often and proves the second statement of the theorem. ■

## 4 Number of leaves

### 4.1 Moderate tails

Consider the number  $\Lambda_n$  of leaves reached by vertex  $n$ , when the distribution of  $Z$  is such that  $\mathbb{E}Z = \infty$  and  $\mathbb{E} \min(Z, W) < \infty$ . We first show divergence in probability, implying weak consistency in configuration (iii).

**Proposition 11.** *Assume that the distribution of  $Z$  is aperiodic. If  $\mathbb{E}Z = \infty$  and  $\mathbb{E} \min(Z, W) < \infty$ , then*

$$\Lambda_n \rightarrow \infty \quad \text{in probability as } n \rightarrow \infty .$$

The proof uses the balance that arises in the moderate-tail regime. The distribution of delays is sufficiently heavy to ensure that each non-leaf vertex in  $T_n$  has a non-negligible chance of reaching a leaf, but light enough to guarantee that  $T_n$  contains a substantial number of non-leaf vertices. Consequently, the overall number of opportunities to generate new leaves grows significantly with  $n$ . This

---

balance is illustrated in Figure 4: only in the moderate tail regime ( $\alpha = 0.6$  in the figure) do we simultaneously observe a sizable number of blue vertices (non-leaves) and a large population of red vertices (leaves).

To formalize this heuristic, we focus on the chain  $S_n$ , obtained by repeatedly following the shortest outgoing edge starting from vertex  $n$ . In the moderate-tailed regime, this chain typically contains many vertices, and each of these vertices has an additional opportunity to connect to a leaf through its longest outgoing edge. We therefore define

$$J_n = \sum_{m=1}^n \mathbb{1}_{m \in S_n} \mathbb{1}_{\max(Z_m, W_m) \geq m},$$

so that  $J_n$  counts the number of vertices in  $S_n$  that are connected to a leaf via their longest outgoing edge. Controlling the growth of  $J_n$  makes it possible to turn the heuristic balance into a rigorous statement.

**Lemma 12.** *The expected number of vertices in the chain  $S_n$  that connect to a leaf through their longest outgoing edge grows to infinity with  $n$ :*

$$\mathbb{E}[J_n] \rightarrow \infty \quad \text{as } n \rightarrow \infty.$$

**Proof.** Note that for  $n \geq m \geq 1$ , the events  $\{m \in S_n\}$  and  $\{\max(Z_m, W_m) \geq m\}$  are independent. Define  $v_m := \mathbb{P}(0 \in S_m)$  for all  $m \geq 1$  and let  $v_0 = 1$ , so that

$$\mathbb{E}J_n = \sum_{m=1}^n \mathbb{P}(m \in S_n) \mathbb{P}(\max(Z_m, W_m) \geq m) = \sum_{m=1}^n v_{n-m} \mathbb{P}(\max(Z_m, W_m) \geq m).$$

The sequence  $(v_n)_{n \geq 0}$  can be better understood by considering the arithmetic renewal process  $M_n$  with holding times given by the i.i.d. random variables  $(\min(Z_i, W_i))_{i \geq 1}$ , and initialized with  $M_0 = 0$ . In other words,

$$M_n = \sum_{i=1}^n \min(Z_i, W_i).$$

By definition,  $v_n$  is the probability that  $n$  can be expressed as a sum of independent

---

random variables  $(\min(Z_i, W_i))_{i \geq 0}$ . In terms of the renewal process, this means that

$$v_n = \mathbb{P}(\exists k \geq 0 : M_k = n) .$$

Hence, the sequence  $(v_n)_{n \geq 0}$  corresponds to the renewal measure of the aperiodic renewal process  $(M_n)_{n \geq 0}$ . By the Erdős–Feller–Pollard renewal theorem [9], it follows that

$$\lim_{n \rightarrow \infty} v_n = \frac{1}{\mathbb{E} \min(Z, W)} > 0 . \quad (4.1)$$

Thus, for some constant  $c > 0$ , we have

$$\mathbb{E} J_n \geq c \sum_{m=1}^n p_m \rightarrow \infty \quad \text{as } n \rightarrow \infty . \quad (4.2)$$

■

We use the second-moment method to show that  $J_n \rightarrow \infty$  in probability.

**Lemma 13.** *Assume that  $\mathbb{E} Z = \infty$  and  $\mathbb{E} \min(Z, W) < \infty$ , and let  $\mathbb{V} J_n$  denote the variance of the random variable  $J_n$ . Then*

$$\frac{\mathbb{V} J_n}{(\mathbb{E} J_n)^2} \rightarrow 0 \quad \text{as } n \rightarrow \infty . \quad (4.3)$$

*In particular, we have:*

$$\mathbb{P}\left(J_n \leq \frac{\mathbb{E} J_n}{2}\right) = \mathbb{P}\left(J_n - \mathbb{E} J_n \leq -\frac{\mathbb{E} J_n}{2}\right) \leq 4 \frac{\mathbb{V} J_n}{(\mathbb{E} J_n)^2} \rightarrow 0 ,$$

*and therefore  $J_n \rightarrow \infty$  in probability as  $n \rightarrow \infty$ .*

**Proof.** [Proof of Lemma 13] To show (4.3), we prove that  $\mathbb{E}[J_n^2]/(\mathbb{E} J_n)^2 \rightarrow 1$  as  $n \rightarrow \infty$ . We first define

$$p_n^* = \mathbb{P}(\max(Z, W) \geq n) = 2p_n - p_n^2 .$$

---

Recall that  $v_n = \mathbb{P}(0 \in S_n)$ , as introduced earlier. Then, for  $n \geq 1$ ,

$$\begin{aligned}
(\mathbb{E}J_n)^2 &= 2 \sum_{1 \leq k < m \leq n} p_m^* p_k^* v_{n-m} v_{n-k} + \sum_{1 \leq m \leq n} (p_m^*)^2 v_{n-m}^2 \\
&= 2 \sum_{1 \leq k < m \leq n} p_m^* p_k^* v_{n-m} v_{n-k} + O(1)
\end{aligned} \tag{4.4}$$

since  $\sum_{1 \leq m \leq n} (p_m^*)^2 v_{n-m}^2$  is bounded by  $4\mathbb{E} \min(Z, W) < \infty$ .

Now, define for any  $m \geq 1$  the random variable  $X_m = \max(Z_m, W_m)$  and observe that for any  $1 \leq k < m \leq n$ , the events  $\{X_k \geq k\}$  and  $\{X_m \geq m, m \in S_n, k \in S_n\}$  are independent. Thus, for any fixed  $n \geq 1$ , we may write

$$\begin{aligned}
\mathbb{E}[J_n^2] &= 2 \sum_{1 \leq k < m \leq n} \mathbb{P}(m \in S_n, k \in S_n, X_m \geq m, X_k \geq k) + \mathbb{E}J_n \\
&= 2 \sum_{1 \leq k < m \leq n} p_k^* \mathbb{P}(m \in S_n, k \in S_n, X_m \geq m) + \mathbb{E}J_n \\
&= 2 \sum_{1 \leq k < m \leq n} p_k^* \mathbb{P}[k \in S_n, X_m \geq m | m \in S_n] \mathbb{P}(m \in S_n) + \mathbb{E}J_n
\end{aligned}$$

If  $m \in S_n$ , the event  $\{k \in S_n, X_m \geq m\}$  can only happen if one of the two random variables  $Z_m$  and  $W_m$  is greater or equal to  $m$  and the other one is smaller than  $m - k$ , for otherwise  $k$  could not belong to  $S_n$  when  $1 \leq k < m \leq n$ . Thus, using the independence of  $Z_m$  with respect to  $W_m$ ,  $\{k \in S_n\}$  and  $\{m \in S_n\}$ , and using the fact that  $p_m^* + p_m^2 = 2p_m$ ,

$$\begin{aligned}
\mathbb{E}[J_n^2] & \tag{4.5} \\
&= 2 \sum_{1 \leq k < m \leq n} p_k^* \times 2\mathbb{P}[Z_m \geq m, W_m \leq m - k, k \in S_n | m \in S_n] \mathbb{P}(m \in S_n) + \mathbb{E}J_n \\
&= 2 \sum_{1 \leq k < m \leq n} p_k^* (p_m^* + p_m^2) v_{n-m} \mathbb{P}[W_m \leq m - k, k \in S_n | m \in S_n] + \mathbb{E}J_n \\
&= 4 \sum_{1 \leq k < m \leq n} p_k^* (p_m^* + p_m^2) v_{n-m} \sum_{i=1}^{k-m} \mathbb{P}(k \in S_{m-i}) \mathbb{P}(W_m = i) + \mathbb{E}J_n \\
&= 2 \sum_{1 \leq k < m \leq n} p_k^* (p_m^* + p_m^2) v_{n-m} \sum_{i=1}^{m-k} v_{m-k-i} q_i + \mathbb{E}J_n. \tag{4.6}
\end{aligned}$$

---

Observe that

$$\begin{aligned}
\sum_{1 \leq k < m \leq n} p_k^* p_m^2 v_{n-m} \sum_{i=1}^{m-k} q_i v_{m-k-i} &\leq \sum_{1 \leq k < m \leq n} p_k^* p_m^2 v_{n-m} \\
&\leq \sum_{1 \leq k < m \leq n} p_k^* p_m^2 \\
&\leq \left( \sum_{m=1}^{\infty} p_m^2 \right) \left( \sum_{k=1}^n p_k^* \right) \\
&= O(\mathbb{E}J_n)
\end{aligned}$$

by using the inequality in (4.2), which leads to

$$\mathbb{E}[J_n^2] = 2 \sum_{1 \leq k < m \leq n} p_k^* p_m^* v_{n-m} \sum_{i=1}^{m-k} v_{m-k-i} q_i + O(\mathbb{E}J_n). \quad (4.7)$$

Fix some  $\epsilon > 0$ . By (4.1), we know that there exists some constant  $x > 0$  such that for all  $m \geq x$ , we have  $|v_m - \lambda| < \epsilon$ , where  $\lambda = 1/(\mathbb{E} \min(Z, W))$ .

If  $m \geq x$ , see that

$$\begin{aligned}
\sum_{i=1}^m q_i v_{m-i} &\leq \sum_{i=1}^{m-x} q_i (\lambda + \epsilon) + \sum_{m-x < i \leq m} q_i \\
&\leq (\lambda + \epsilon) + p_{m-x}.
\end{aligned}$$

Thus, there exists some constant  $y > x$  such that for all  $m \geq y$ ,

$$\sum_{i=1}^m q_i v_{m-i} \leq \lambda + 2\epsilon.$$

Moreover,

$$\begin{aligned}
\sum_{\substack{1 \leq k < m \leq n \\ \text{s.t. } m-k < y}} p_m^* v_{n-m} p_k^* \underbrace{\sum_{i=1}^{m-k} v_{m-k-i} q_i}_{\leq 1} &\leq \sum_{1 \leq m \leq n} p_m^* v_{n-m} \sum_{k=m-y}^m 1 \\
&= (y+1) \mathbb{E}J_n = O(\mathbb{E}J_n),
\end{aligned}$$

---

and

$$\begin{aligned} \sum_{\substack{1 \leq k < m \leq n \\ \text{s.t. } n-m < y}} p_k^* p_m^* v_{n-m} \underbrace{\sum_{i=1}^{m-k} v_{m-k-i} q_i}_{\leq 1} &\leq \sum_{n-y \leq m \leq n} 1 \sum_{k=1}^n p_k^* \\ &\leq (y+1) \sum_{k=1}^n p_k^* = O(\mathbb{E}J_n). \end{aligned}$$

Putting everything together, we have that

$$\mathbb{E}[J_n^2] = \sum_{m=1}^{n-y} \sum_{k=1}^{m-y} p_m^* p_k^* v_{n-k} \sum_{i=1}^{m-k} q_i v_{m-k-i} + O(\mathbb{E}J_n). \quad (4.8)$$

Similarly, we have

$$(\mathbb{E}J_n)^2 = \sum_{s=1}^{n-y} \sum_{k=1}^{m-y} p_m^* p_k^* v_{n-m} v_{n-k} + O(\mathbb{E}J_n). \quad (4.9)$$

Finally, observe that for any  $n > y$  we have:

$$\sum_{m=1}^{n-y} \sum_{k=1}^{m-y} p_m^* p_k^* v_{n-m} \underbrace{\sum_{i=1}^{m-k} q_i v_{m-k-i}}_{\leq (\lambda+2\epsilon)^2} \leq (\lambda+2\epsilon)^2 \sum_{m=1}^{n-y} \sum_{k=1}^{m-y} p_m^* p_k^* \quad (4.10)$$

and

$$\sum_{m=1}^{n-y} \sum_{k=1}^{m-y} p_m^* p_k^* \underbrace{v_{n-m} v_{n-k}}_{\geq (\lambda-\epsilon)^2} \geq (\lambda-\epsilon)^2 \sum_{m=1}^{n-y} \sum_{k=1}^{m-y} p_m^* p_k^*. \quad (4.11)$$

From (4.8), (4.9), (4.10) and (4.11) it follows that

$$\limsup_{n \rightarrow \infty} \frac{\mathbb{E}[J_n^2]}{(\mathbb{E}J_n)^2} \leq \left( \frac{\lambda+2\epsilon}{\lambda-\epsilon} \right)^2.$$

---

Since this is true for any  $\epsilon > 0$  small enough, and  $\mathbb{E}[J_n^2] \geq (\mathbb{E}J_n)^2$ , we conclude that

$$\frac{\mathbb{E}[J_n^2]}{(\mathbb{E}J_n)^2} \rightarrow 1$$

as  $n \rightarrow \infty$ . ■

We now turn to the proof of Proposition 11, which asserts that the number of leaves in the medium-tail scenario diverges in probability. We have already shown that  $J_n \rightarrow \infty$  in probability, where  $J_n$  counts the number of edges connecting vertices in the chain  $S_n$  to leaves via their longest outgoing edge. Proposition 11 follows from this result, but some care is needed:  $J_n$  counts edges, not distinct leaves, and multiple edges may connect to the same leaf. The key idea of the proof is to show that the number of edges of the type  $\max(Z_n, W_n)$  connecting to any single leaf cannot grow to infinity. This ensures that the divergence of  $J_n$  indeed implies that the number of distinct leaves connected to  $T_n$  also diverges to infinity in probability.

**Proof.** [Proof of Proposition 11] Let us define

$$\mathcal{L}_n^* = \{k \leq 0 : \text{there exists an } m \in S_n \text{ such that } \max(Z_m, W_m) = m - k\}.$$

In other words,  $\mathcal{L}_n^*$  is the subset of vertices in  $\mathcal{L}_n$  that are connected to some vertex in  $S_n$  with positive index through the longest edge (i.e., given by  $\max(Z, W)$ ).

Given  $n \geq 1$  and  $k \in \mathbb{Z}$ , we introduce the random variable

$$D_{n,k} = \sum_{m=1}^n \mathbb{1}_{\max(Z_m, W_m) = m - k}.$$

Note that for any fixed  $x > 0$ ,

$$\mathbb{P}(|\mathcal{L}_n^*| < x) \tag{4.12}$$

$$\begin{aligned} &\leq \mathbb{P}\left(\left[\max_{k \in \mathcal{L}_n^*} \sum_{m \in S_n} \mathbb{1}_{\max(Z_m, W_m) = m - k} < \frac{\mathbb{E}J_n}{2x}\right] \cap [|\mathcal{L}_n^*| < x]\right) \\ &\quad + \mathbb{P}\left(\bigcup_{k \leq 0} \left[D_{n,k} \geq \frac{\mathbb{E}J_n}{2x}\right]\right). \end{aligned} \tag{4.13}$$

---

Since  $J_n$  can be written as

$$J_n = \sum_{k \in \mathcal{L}_n^*} \sum_{m \in \mathcal{S}_n} \mathbb{1}_{\max(Z_m, W_m) = m-k},$$

we have that

$$\begin{aligned} & \mathbb{P} \left( \left[ \max_{k \in \mathcal{L}_n^*} \sum_{m \in \mathcal{S}_n} \mathbb{1}_{\max(Z_m, W_m) = m-k} < \frac{\mathbb{E}J_n}{2x} \right] \cap [|\mathcal{L}_n^*| < x] \right) \\ & \leq \mathbb{P} \left( J_n \leq \frac{\mathbb{E}J_n}{2} \right) \\ & \leq 4 \frac{\mathbb{V}J_n}{(\mathbb{E}J_n)^2} \rightarrow 0. \end{aligned} \tag{4.14}$$

Moreover, for any  $k \leq 0$  and  $y > 0$ , by Chernoff's bound [7], [5], we have that

$$\begin{aligned} \mathbb{P}(D_{n,k} > y) &= \mathbb{P} \left( \sum_{m=1}^n \mathbb{1}_{\max(Z_m, W_m) = m-k} > y \right) \\ &\leq \exp \left( y - p_{1-k}^* - y \log \left( \frac{y}{p_{1-k}^*} \right) \right) \\ &\leq \left( \frac{e}{y} \right)^y (p_{1-k}^*)^y. \end{aligned}$$

Let  $y_n = \mathbb{E}J_n/(2x)$  and assume that  $n$  is large enough so we have  $y_n \geq 2$  and

$$\sum_{m=1}^{\infty} (p_m^*)^{y_n} \leq \sum_{m=1}^{\infty} (p_m^*)^2 < \infty \text{ (since } \mathbb{E} \min(Z, W) < \infty \text{)}.$$

Thus, by the union bound,

$$\begin{aligned} \mathbb{P} \left( \bigcup_{k \leq 0} \left[ D_{n,k} \geq \frac{\mathbb{E}J_n}{2x} \right] \right) &\leq \sum_{k \leq 0} \left( \frac{e}{y_n} \right)^{y_n} (p_{1-k}^*)^{y_n} \\ &\leq \left( \frac{e}{y_n} \right)^{y_n} \sum_{k=1}^{\infty} (p_k^*)^2 \rightarrow 0. \end{aligned} \tag{4.15}$$

By putting (4.13), (4.14), and (4.15) together we have that  $\mathbb{P}(|\mathcal{L}_n^*| < x) \rightarrow 0$  as  $n \rightarrow \infty$  for every fixed  $x > 0$ , which proves Proposition 11. ■

---

Having established divergence in probability, we are now ready to prove almost sure divergence.

**Proof.** [Proof of Theorem 6] For any  $n \geq 1$ , define  $Y_n = \min\{m - \min(Z_m, W_m) : m \geq n\}$ . A simple union bound gives, for arbitrary  $k$ :

$$\mathbb{P}(Y_n \leq k) \leq \sum_{m \geq n} \mathbb{P}(\min(Z_m, W_m) > m - k).$$

Since  $\mathbb{E}\min(Z, W) < \infty$ , the sum above is finite and goes to 0 as  $n$  goes to infinity. Hence,  $Y_n \rightarrow \infty$  in probability.

Let us fix  $n \geq 0$  and apply Corollary 9 to the model obtained by shifting by  $n$  the set of integers so that leaves are indexed by the set  $\{k : k \leq n\}$  and the edges are given by the sequences of random variables  $(Z_k)_{k \geq n+1}$  and  $(W_k)_{k \geq n+1}$ . This gives us a random variable  $N_n > n$ , finite almost surely, such that  $N_n \in S_k$  for all  $k \geq N_n$ . Let  $I_n \leq n\}$  be the root of this tree, and note that

$$Y_n \leq I_n \leq n < N_n.$$

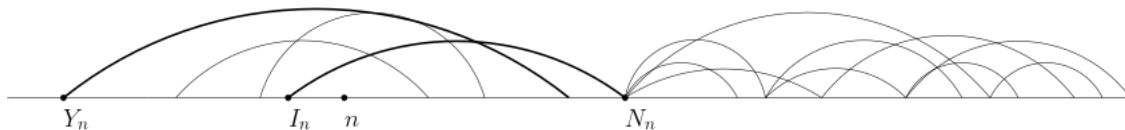


Figure 5: Illustration of the proof of Theorem 6

Since  $Y_n \rightarrow \infty$  in probability,  $I_n \rightarrow \infty$  in probability as well. Moreover, for all  $n \geq 1$ , one can choose  $N_n$  such that  $N_n - n$  is distributed as  $N_0$ . The fact that all vertices  $k \geq N_n$  belong to the tree rooted at  $I_n$  in this single-choice model implies that  $I_n \in T_k$  for all  $k \geq N_n$ , and in particular  $\Lambda_{I_n} \leq \Lambda_k$  for all  $k \geq N_n$ . From Proposition 11 we know that  $\Lambda_n \rightarrow \infty$  in probability. Hence,  $\Lambda_{I_n} \rightarrow \infty$  in probability. Moreover, for any fixed  $m \geq 1$  we have

$$\begin{aligned} \mathbb{P}(\cup_{k \geq 2n} [\Lambda_k \leq m]) &\leq \mathbb{P}(N_n > 2n) + \mathbb{P}(N_n \leq 2n, \Lambda_{I_n} \leq m) \\ &\leq \mathbb{P}(N_0 > n) + \mathbb{P}(\Lambda_{I_n} \leq m). \end{aligned}$$

---

Thus, for any  $m \geq 1$ ,  $\mathbb{P}(\cup_{k \geq 2n} [\Lambda_k \leq m]) \rightarrow 0$  as  $n \rightarrow \infty$ . By continuity of measure, this implies that for any  $m \geq 1$ ,

$$\mathbb{P}(\cap_{n \geq 1} \cup_{k \geq 2n} [\Lambda_k \leq m]) = 0.$$

By the union bound,

$$\mathbb{P}(\cup_{m \geq 1} \cap_{n \geq 1} \cup_{k \geq 2n} [\Lambda_k \leq m]) = 0,$$

which implies that  $\Lambda_n \rightarrow \infty$  almost surely. ■

## 4.2 Heavy tails

In this section, we study the heavy-tailed case, that is,  $\mathbb{E} \min(Z, W) = \infty$ . From Theorem 3 we already know that  $\Lambda_n$  cannot go to infinity with probability one. Here we show that, in contrast to this,  $\Lambda_n \rightarrow \infty$  in probability when  $\sum_{n=1}^{\infty} r_n^2 < \infty$ .

**Lemma 14.** *For any integers  $k \in \mathbb{Z}$  and  $n \geq \max(1, k)$ ,*

$$\mathbb{P}(k \in T_n^Z) \leq r_{n-k}.$$

**Proof.** Clearly, we have  $\mathbb{P}(k \in T_n^Z) = r_{n-k}$  whenever  $k \geq 0$ . Moreover,  $r_n$  satisfies the recursion

$$r_n = \sum_{i=1}^n q_i r_{n-i}.$$

Thus, for any  $k > 0$ , we have

$$\begin{aligned} \mathbb{P}(-k \in T_n^Z) &= \sum_{i=1}^n \mathbb{P}(Z_i = i+k) \mathbb{P}(i \in T_n) \\ &= \sum_{i=1}^n q_{k+i} r_{n-i} \\ &= \sum_{i=k+1}^{n+k} q_i r_{n+k-i} \end{aligned}$$

$$\leq \sum_{i=1}^{n+k} q_i r_{n+k-i} = r_{n+k}.$$

■

**Proof.** [Proof of Theorem 7] Assume that  $\sum_{n=1}^{\infty} r_n^2 < \infty$ . Fix a constant  $w > 0$  and let  $I_n^w$  denote the number of intervals of the form  $(kw, (k+1)w]$ , with  $k \in \mathbb{Z}_{\geq 0}$ , that are intersected by the chain  $T_n^Z$ . Note that, for any  $x > 0$ , if  $I_n^w < x$  then  $T_n^Z < wx$ , thus

$$\begin{aligned} \mathbb{P}(I_n^w < x) &\leq \mathbb{P}(T_n^Z < wx) \\ &\leq \mathbb{P}(Z_1 + \dots + Z_{\lfloor wx \rfloor} > n) \rightarrow 0. \end{aligned} \quad (4.16)$$

Let  $\ell \geq 1$  be an arbitrary integer and observe that, if the event  $I_n^w \geq 2\ell$  occurs, then one can define a random set of vertices  $A_n$  such that

1.  $A_n$  is independent of  $(W_k)_{k \geq 1}$ ,
2.  $A_n \subset T_n^Z$ ,
3.  $|A_n| = \ell$ ,
4. for all  $k \neq m \in A_n$ ,  $|k - m| \geq w$ .

Conditionally on  $I_n^w \geq 2\ell$ , we have

$$\mathbb{P}\left(\bigcup_{k \neq m, k, m \in A_n} T_k^W \cap T_m^W \neq \emptyset\right) \quad (4.17)$$

$$\begin{aligned} &\leq \mathbb{E} \sum_{\substack{k, m \in A_n \\ \text{s.t. } k \neq m}} \mathbb{1}_{T_k^W \cap T_m^W \neq \emptyset} \\ &\leq \mathbb{E} \sum_{\substack{k, m \in A_n \\ \text{s.t. } k \neq m}} \mathbb{P}(T_k^W \cap T_m^W \neq \emptyset) \\ &\leq \binom{\ell}{2} \max_{\substack{k, m \leq n \\ \text{s.t. } |k-m| \geq w}} \mathbb{P}(T_k^W \cap T_m^W \neq \emptyset) \end{aligned} \quad (4.18)$$

by using the first moment method and the independence of  $A_n$  with respect to

---

$(W_k)_{k \geq 1}$ . Let  $k < m$  be a pair of integers such that  $|k - m| \geq w$  and observe that

$$\mathbb{P}(T_k^W \cap T_m^W \neq \emptyset) = \mathbb{P}(T_k^W \cap T_m^Z \neq \emptyset).$$

Thus, the union bound, Lemma 14 and the Cauchy-Schwarz inequality yield

$$\begin{aligned} \mathbb{P}(T_k^W \cap T_m^W \neq \emptyset) &\leq \sum_{i=-\infty}^k \mathbb{P}(i \in T_k^W) \mathbb{P}(i \in T_m^Z) \\ &\leq \sum_{i=-\infty}^k r_{k-i} r_{m-i} \\ &\leq \sqrt{\sum_{i=0}^{\infty} r_i^2} \times \sqrt{\sum_{i=m-k}^{\infty} r_i^2}. \end{aligned} \tag{4.19}$$

From (4.18) and (4.19), it follows that

$$\mathbb{P}\left(\bigcup_{k \neq m, k \in A_n} T_k^W \cap T_m^W \neq \emptyset\right) \leq \binom{\ell}{2} \sqrt{\sum_{i=0}^{\infty} r_i^2} \times \sqrt{\sum_{i=w}^{\infty} r_i^2}. \tag{4.20}$$

Fix  $\epsilon > 0$ . From (4.20) we know that we can choose  $w$  big enough so we have for  $n \geq 1$ ,

$$\mathbb{P}\left(\bigcup_{k \neq m, k \in A_n} T_k^W \cap T_m^W \neq \emptyset\right) \leq \epsilon. \tag{4.21}$$

Finally, observe that if  $\Lambda_n < \ell$  and  $I_n^w \geq 2\ell$ , then there have to be at least two distinct vertices  $k, m$  in  $A_n$  such that the chains  $T_k^W$  and  $T_m^W$  meet each other: otherwise, each chain  $T_k^W$  with  $k \in A_n$  would lead to a distinct leaf in  $T_n$ , contradicting the event  $\Lambda_n < \ell$ . Thus, by (4.20) we have

$$\mathbb{P}(\Lambda_n < \ell) \leq \mathbb{P}(I_n^w < 2\ell) + \epsilon.$$

From (4.16) it follows that for any  $\epsilon > 0$  and any integer  $\ell \geq 1$

$$\limsup_{n \rightarrow \infty} \mathbb{P}(\Lambda_n < \ell) \leq \epsilon,$$

which concludes the proof of Theorem 7. ■

---

## 5 Generalization to $k$ choices.

A natural generalization of the model allows for  $k \geq 2$  choices. Fix  $k \geq 2$  and consider  $k$  independent sequences of i.i.d. random variables  $(Z_i^{(1)})_{i \geq 1}, \dots, (Z_i^{(k)})_{i \geq 1}$ . Given a sequence of preference values  $(U_i)_{i \leq 0}$ , we define a sequential random coloring by setting  $V_n = U_n$  and  $C_n = n$  for all  $n \leq 0$ , and for  $n > 0$  letting

$$C_n = C_{n-Z_n^{(j)}} \quad \text{and} \quad V_n = V_{n-Z_n^{(j)}} \quad \text{where} \quad j = \operatorname{argmin}_{1 \leq i \leq k} V_{n-Z_n^{(i)}}$$

for  $n \geq 1$ . The set of vertices reached from vertex  $n$  is now defined recursively as

$$T_n = \{n\} \cup \left( \bigcup_{i=1}^k T_{n-Z_n^{(i)}} \right), \quad n \geq 1,$$

with  $T_n = \{n\}$  for all  $n \leq 0$ . As in the two-choice model, we may then define  $R_n$ ,  $\Lambda_n$ , and  $L_n$  accordingly. The graph associated with this model therefore, contains multiple embedded copies of the two-choice process. It follows from this observation that consistency in the two-choice model carries over to the  $k$ -choice model for any  $k \geq 2$ , and for each of the three configurations of the initial pool of topics considered in this paper. As a consequence, Theorem 4 extends automatically to the  $k$ -choice model:

**Theorem 15.** *Assume that  $q_i > 0$  for all  $i \in \mathbb{N}$  and  $\mathbb{E}Z = \infty$ . Then, in the  $k$ -choice model, with probability one,*

$$L_n \rightarrow -\infty.$$

*Thus, the system is strongly consistent in configuration (ii).*

However, the increase in the number of available recommendations allows one to work under weaker assumptions on the tail of  $Z$ , and in this setting the results can be sharpened. In particular, one can adapt the proof given in Section 4.1 by modifying the recursion defining  $S_n$  to

$$S_n = \{n\} \cup S_{n - \min_{1 \leq i \leq k} Z_n^{(i)}},$$

---

and by controlling the growth of

$$J_n = \sum_{m=1}^n \mathbb{1}_{m \in \mathcal{S}_n} \mathbb{1}_{\max_{1 \leq i \leq k} Z_m^{(i)} \geq m},$$

to establish the following result.

**Theorem 16.** *Assume that the distribution of  $Z$  is aperiodic, and that*

$$\mathbb{E}[Z] = \infty, \quad \mathbb{E}\left[\min_{1 \leq i \leq k} Z^{(i)}\right] < \infty,$$

where  $Z^{(1)}, \dots, Z^{(k)}$  are independent copies of  $Z$ . Then, in the  $k$ -choice model,

$$\Lambda_n \rightarrow \infty \quad \text{almost surely.}$$

The arguments used in the proofs of Theorems 2, 3 and Theorem 5 extend naturally to the present setting, leading to the following three results.

**Theorem 17.** *Let  $L_\infty = \inf_{n \in \mathbb{N}} L_n$ . If  $\mathbb{E}Z < \infty$ , then  $L_\infty$  is finite almost surely. Thus, in the  $k$ -choice model, the sequences  $(R_n)_{n \in \mathbb{N}}$ ,  $(L_n)_{n \in \mathbb{N}}$  and  $(\Lambda_n)_{n \in \mathbb{N}}$  are almost surely bounded.*

**Theorem 18.** *Let  $Z^{(1)}, \dots, Z^{(k)}$  be independent copies of  $Z$ . Assume that*

$$\mathbb{E} \min_{1 \leq i \leq k} Z^{(i)} = \infty.$$

Then, in the  $k$ -choice model, with probability one,

$$\liminf_{n \rightarrow \infty} \Lambda_n < \infty.$$

**Theorem 19.** *Let  $Z^{(1)}, \dots, Z^{(k)}$  be independent copies of  $Z$ .*

1. *If  $\mathbb{E} \min_{1 \leq i \leq k} Z^{(i)} < \infty$ , then  $\sup_{n \in \mathbb{N}} |R_n| < \infty$  almost surely in the  $k$ -choice model.*
2. *If  $\mathbb{E} \min_{1 \leq i \leq k} Z^{(i)} = \infty$ , then  $\sup_{n \in \mathbb{N}} |R_n| = \infty$  almost surely in the  $k$ -choice model.*

In summary, the consistency table from Section 1.4 remains unchanged in the  $k$ -choice model. The only difference is that the moderate-tail regime is

---

now characterized by the conditions  $\mathbb{E} \min_{1 \leq i \leq k} Z^{(i)} < \infty$  and  $\mathbb{E} Z^{(1)} = \infty$ , while the heavy-tail regime corresponds to  $\mathbb{E} \min_{1 \leq i \leq k} Z^{(i)} = \infty$ .

## 6 Conclusion

We introduced and studied a simple mathematical model for recommendation systems based on giving each user two random choices, which include a mixture of past choices and untried options. In this setup, we identify three regimes based on the tail behavior of the sizes of the jumps in the past.

Although our model is primarily theoretical, its analysis offers qualitative insights that may inform the design of recommendation systems. In particular, it highlights how structural properties of the recommendation process can influence long-term behavior and diversity. Extending the model to incorporate personalization and adaptive feedback remains an interesting direction for future work.

### 6.1 Comparison with standard baselines

To situate our results within the broader landscape of recommendation methods, it is natural to consider connections to standard algorithmic baselines. Classical approaches, such as collaborative filtering, multi-armed bandits, and deep learning-based recommender systems, rely on adaptive strategies to cope with noisy or partially observed user preferences. While our model is highly stylized and focuses on the structural properties of the recommendation process rather than optimization or learning, comparing its behavior conceptually to these frameworks helps highlight both its scope and its limitations.

A natural point of comparison is with multi-armed bandit algorithms [16, 17], which also model sequential decision-making and recommendation dynamics. However, the two frameworks differ fundamentally: in our model, the “rewards” (preference values) are static and noise-free, with randomness arising solely from the distribution of edge lengths, while in bandit algorithms, the primary challenge is to manage uncertainty in the reward distributions through exploration and exploitation. Because of this distinction, standard bandit baselines are not di-

---

rectly applicable to our setting, and a meaningful comparison would require a reformulation of the bandit problem that departs significantly from its usual assumptions.

A closer comparison can be drawn with bandit algorithms in the noiseless setting, where the reward associated with each arm is fixed and known after a single pull. In that framework, the problem is essentially trivial: the optimal arm is identified immediately, and the learning phase vanishes. By contrast, our model remains nontrivial even in the absence of noise, since the randomness lies not in the evaluation of topics but in the structure of the recommendation process itself. The stochastic nature of the edge lengths forces the system to repeatedly revisit or discover new topics, giving rise to rich asymptotic behaviors such as strong or weak consistency.

## 6.2 Perspectives and open problems

One might naturally expect the  $k$ -choice model to exhibit distinct behavior in the  $k + 1$  regimes defined by the tail conditions

$$\mathbb{E} \min_{1 \leq i \leq \ell} Z^{(i)} = \infty \quad \text{and} \quad \mathbb{E} \min_{1 \leq i \leq \ell+1} Z^{(i)} < \infty,$$

for each  $\ell \in \{0, \dots, k\}$  (with the understanding that the left-hand condition is omitted when  $\ell = 0$  and the right-hand condition is omitted when  $\ell = k$ ). Although our results extend straightforwardly to the  $k$ -choice setting, the current consistency analysis does not distinguish between these individual regimes. A deeper understanding of the distinctions between these regimes and their implications for recommendation outcomes would be of significant interest.

The present model remains highly abstract and does not directly reflect real-world applications. For instance, the assumption of static preference values for each topic may not capture the evolution of actual user behavior. A natural direction for future work is to develop dynamic generalizations of the model that account for time-varying preferences for each topic.

---

## Appendix

Symbol	Description
$T_n$	Set of vertices reached by vertex $n$ .
$\mathcal{L}_n$	Set of vertices with non-positive index (or <i>leaves</i> ) that is reached from vertex $n$ . ( $\mathcal{L}_n = T_n \cap \{0, -1, -2, \dots\}$ )
$\Lambda_n$	Number of leaves reached from $n$ . ( $\Lambda_n =  \mathcal{L}_n $ )
$R_n$	Rightmost leaf in $T_n$ . ( $R_n = \max\{i : i \in \mathcal{L}_n\}$ )
$L_n$	Leftmost leaf in $T_n$ . ( $L_n = \min\{i : i \in \mathcal{L}_n\}$ )
$T_n^Z$ (resp. $T_n^W$ )	Chain in $T_n$ obtained by repeatedly taking the $Z$ -edges (resp. the $W$ -edges)
$S_n$	Chain in $T_n$ obtained by repeatedly taking shortest outgoing edges
$J_n$	Number of vertices in $S_n$ connected to a leaf via their longest outgoing edge
$C_n$	Topic recommended at time $n$
$Z_n, W_n$	Independent random delays for recommendations
$U_i$	Preference value of topic $i$ ( $i \leq 0$ )
$V_n$	Preference value of topic seen at time $n$
$p_i, q_i$	Probabilities associated with $Z$ ( $p_i = \mathbb{P}(Z \geq i)$ , $q_i = \mathbb{P}(Z = i)$ )
$p_i^*$	Upper tail probability associated with $\max(Z, W)$ ( $p_i^* = \mathbb{P}(\max(Z, W) \geq i) = 2p_i - p_i^2$ )
$r_n$	Probability that vertex 0 belongs to the $Z$ -chain starting at vertex $n$ ( $r_n = \mathbb{P}(0 \in T_n^Z) = \mathbb{P}(0 \in T_n^W)$ )
$v_n$	Probability that vertex 0 belongs to the chain $S_n$ ( $v_n = \mathbb{P}(0 \in S_n)$ )
$\mathbb{1}_A$	Indicator function of event $A$
$\mathbb{E}X$	Expected value of random variable $X$
$\mathbb{V}X$	Variance of random variable $X$

Notation Table

## Funding

Luc Devroye was supported by NSERC under grant number RGPIN-2024-04164. Gábor Lugosi acknowledges the support of Ayudas Fundación BBVA a Proyectos de Investigación Científica 2021 and the Spanish Ministry of Economy and Competitiveness grant PID2022-138268NB-I00, financed by MCIN AEI 10.13039

## References

- [1] Baccelli, F., Haji-Mirsadeghi, M.-O., and Khaniha, S. (2022). Coupling from the past for the null recurrent Markov chain. *The Annals of Applied Probability*, 34(4):3631–3664.
- [2] Baccelli, F., Haji-Mirsadeghi, M.-O., and Khezeli, A. (2018). Eternal family trees and dynamics on unimodular random graphs. In *Unimodularity in Randomly Generated Graphs*, Contemporary Mathematics, AMS-eBooks vol. 719, pages 85–127. American Mathematical Society, Providence, RI.
- [3] Baccelli, F. and Sodre, A. (2019). Renewal processes, population dynamics, and unimodular trees. *Journal of Applied Probability*, 56(2):339–357.
- [4] Blath, J., González Casanova, A., Kurt, N., and Spano, D. (2013). The ancestral process of long-range seed bank models. *Journal of Applied Probability*, 50(3):741–759.
- [5] Boucheron, S., Lugosi, G., and Massart, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, Oxford.
- [6] Broutin, N., Devroye, L., Lugosi, G., and Oliveira, R. I. (2024). Subtractive random forests. *ALEA Latin American Journal of Probability and Mathematical Statistics*, 21:575–591.
- [7] Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics*, 23(4):493–507.
- [8] Chierichetti, F., Kumar, R., and Tomkins, A. (2020). Asymptotic behavior of sequence models. In *Proceedings of The Web Conference 2020*, pages 2824–2830.
- [9] Erdős, P., Feller, W., and Pollard, H. (1949). A theorem on power series. *Bulletin of the American Mathematical Society*, 55:201–204.
- [10] Hammond, A. and Sheffield, S. (2013). Power law pólya’s urn and fractional brownian motion. *Probability Theory and Related Fields*, 157(3-4):691–719.
- [11] Ibragimov, I. and Linnik, Y. (1971). *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff Publishing, Groningen.
- [12] Igelbrink, J. L. and Wakolbinger, A. (2022). Asymptotic Gaussianity via coalescence probabilities in the Hammond–Sheffield urn. *ALEA Latin American Journal of Probability and Mathematical Statistics*, 20:53–74.

- 
- [13] Malevich, N. (2015). *Approximations and Asymptotic Expansions for Sums of Heavy-tailed Random Variables*. Ph.D. Dissertation, Otto-von-Guericke-Universität Magdeburg.
- [14] Zhang, S., Yao, L., Sun, A., and Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys*, 52(1):1–38.
- [15] Zolotarev, V. (1986). *One-Dimensional Stable Distributions*. American Mathematical Society, Providence.
- [16] Lattimore, T. and Szepesvári, C. (2020). *Bandit Algorithms*. Cambridge University Press, Cambridge.
- [17] Bubeck, S. and Cesa-Bianchi, N. (2012) *Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems*. Now Publishers, Inc., Norwell.