Fairness in Online Social Network Timelines: Measurements, Models and Mechanism Design

Eduardo Hargreaves[a,*], Claudio Agosti[b], Daniel Menasché[a], Giovanni Neglia[c], Alexandre Reiffers-Masson[d], Eitan Altman[c]

[a] Federal University of Rio de Janeiro,Rio de Janeiro, Brazil
[b] University of Amsterdam (Amsterdam,Netherlands)
[c] INRIA (Sophia Antipolis, France)
[d] Indian Institute of Science (Bangalore, India)

Abstract

Facebook News Feed personalization algorithm has a significant impact, on a daily basis, on the lifestyle, mood and opinion of millions of Internet users. Nonetheless, the behavior of such algorithm lacks transparency, motivating measurements, modeling and analysis in order to understand and improve its properties. In this paper, we propose a reproducible methodology encompassing measurements, an analytical model and a fairness-based News Feed design. The model leverages the versatility and analytical tractability of time-to-live (TTL) counters to capture the visibility and occupancy of publishers over a News Feed. Measurements are used to parameterize and to validate the expressive power of the proposed model. Then, we conduct a what-if analysis to assess the visibility and occupancy bias incurred by users against a baseline derived from the model. Our results indicate that a significant bias exists and it is more prominent at the top position of the News Feed. In addition, we find that the bias is non-negligible even for users that are deliberately set as neutral with respect to their political views, motivating the proposal of a novel and more transparent fairness-based News Feed design.

Keywords: Facebook, measurements, social networks, timelines, bias, fairness

1. Introduction

Background. Online social networks (OSNs) have an increasingly important influence in the life of millions of Internet users, shaping their mood, tastes and political views [13, 47]. In essence, the goal of OSNs is to allow users to connect

*Corresponding author
Email addresses: eduardo@hargreaves.tech (Eduardo Hargreaves), claudio.agosti@tracking.exposed (Claudio Agosti), sadoc@dcc.ufrj.br (Daniel Menasché), giovanni.neglia@inria.fr (Giovanni Neglia), reiffers.alexandre@gmail.com (Alexandre Reiffers-Masson), eitan.altman@inria.fr (Eitan Altman)

and to efficiently share meaningful information. To this aim, one of the key building blocks of OSNs is their filtering algorithm, which personalizes content made available to each user of the system. Facebook, for instance, developed the *News Feed* algorithm for that purpose [19, 32].

Challenges. The News Feed algorithm is a recommendation system that shows posts to users based on inferred users' preferences, trading among possibly conflicting factors while prioritizing posts [58, 61, 51]. Hence, the News Feed algorithm shares common features with traditional recommender systems, such as those used by Netflix and Spotify, to recommend movies and music. For instance, in all such systems users typically do not provide explicit feedback about recommendations. Nonetheless, due to the nature of OSNs, the News Feed algorithm also poses its own set of challenges, related to the measurement, modeling and control of publishers' visibilities.

Facebook users may be unaware of the influence of the filtering they are subject to [31]. Such lack of awareness favors the creation of a *filter bubble* that reinforces users' opinions by selecting the users information diets [66, 53, 72]. While researchers are willing to understand the influence of Facebook through users' News Feed [30], the News Feed algorithm uses sensitive data about preferences, which precludes the sharing of datasets. Public datasets, in turn, are needed in order to parameterize models to reason about how the News Feed is populated.

Models are instrumental to perform what-if analysis, e.g., to understand how the News Feed would behave in the presence of different filtering algorithms. In addition, analytical models can also serve as building blocks towards novel mechanisms to design News Feed algorithms. Such foundational development of principled mechanisms to populate the News Feed is key to build transparency into the system.

Prior art. The literature on the News Feed algorithm includes measurements [16, 10, 11, 14], models [4, 23] and user awareness surveys [29]. Nonetheless, most of the prior work that quantifies the effect of OSNs on information diffusion with large datasets [7, 6, 2, 78, 13] relies on measurements obtained through restrictive non-disclosure agreements that are not made publicly available to other researchers and practitioners. As the data analyzed in such studies is very sensitive, and their sources are not audited, there are multiple potential factors and confounding variables that are unreachable to the general public.

Goals. Our goal is to provide insights on the filtering that occurs in OSNs through a reproducible methodology and a dataset made publicly available.[1] Given such measurements, we pose the following questions:

1. what would be the occupancies of the various sources under alternative scenarios wherein different filtering algorithms are in place?
2. how to design mechanisms to populate timelines in a principled fashion, accounting for users preferences and providing content diversity, e.g., under a fairness-based framework?

[1] https://github.com/EduardoHargreaves/Effect-of-the-OSN-on-the-elections

To address the first question above, we propose an analytical model for the News Feed. The model allows us to derive the occupancy and visibility of each publisher at users' timelines, as a function of the considered filtering process. Using the model, we conduct what-if analysis, e.g., to assess publishers' visibilities in a scenario without filters.

Then, we use the proposed model to build fairness-driven mechanisms to populate timelines. Utilities are used to capture the preferences of users with respect to the exposure to posts from different publishers. The mechanism leverages results on utility-based cache design [21], and accounts for fairness among publishers through utility functions.

Contributions. In this paper we take important steps towards measuring, modeling, auditing and designing timelines. Our key contributions are summarized as follows.

A measurement methodology is introduced to publicly and transparently audit the OSN ecosystem, focusing on the Facebook News Feed algorithm. The methodology encompasses an Internet browser extension to autonomously and independently collect information on the posts presented to users by the News Feed algorithm (Section 2). Such information is not available through the Facebook API.

Empirical findings are reported using data collected from a measurement campaign conducted during the 2018 Italian elections. We observed that *a*) the filtering algorithm is impacted by the profile of pages that a user "likes", *b*) this effect is more prominent at the topmost News Feed position and *c*) neutral users are also exposed to non-uniform filtering (Section 3).

An analytical model is proposed to quantify the visibility and occupancy of publishers in the users' News Feeds. The model allows us to conduct a what-if analysis, to assess the metrics of interest under different filtering mechanisms and is validated using data from the Italian election experiment (Sections 4 and 5).

A fairness-driven mechanism design is proposed, leveraging the proposed model and measurements (Section 6). Given a user profile, that "likes" a certain subset of publishers, the measurements are used to parameterize a simple instance of the model. Then, a family of α-fair utility functions are used to allocate resources to publishers subject to users preferences under a utility maximization framework.

A model-based bias assessment is conducted where the News Feed occupancies are contrasted against an unfiltered resource allocation baseline to quantify the *bias*, i.e., how publishers' occupancies are affected by the News Feed algorithm (Section 6.5).

2. Measurement methodology

The goal of our experiments is to assess the bias experienced by OSN users through a reproducible method. To this aim, we created controlled virtual users that have no social ties and that follow the same set of publishers. By considering minimalistic user profiles, we can assess how preferences towards publishers affect posts presented to users removing, for instance, the effect of social connections.

2.1. Terminology

Next, we introduce some basic terminology. *Publishers* produce *posts* that are fed into users' News Feeds. Each user consumes posts from his/her News Feed. A *News Feed* is an ordered list of posts, also referred to as a timeline, presented to a given user. News Feed may refer to the algorithm used by Facebook to fill the timeline, or to the timeline itself.

Users *follow* publishers that they are interested in. The News Feed of a user is filled up with posts from publishers that they follow. A user may follow a publisher to have posts from that publisher in the user's News Feed. A user who *likes* a publisher automatically follows that publisher. A user likes a publisher to show general support for its posts. In our work, users orientations are established by letting them *like* a subset of the preselected publishers.

2.2. Data collection methodology

Next, we present our measurement methodology. Although this methodology is general, for concreteness our description is based on the 2018 Italian Parliament election, which constitutes the key case study considered in this paper. The Italian election was held on March 4, 2018, and our experiment was conducted between January 10, 2018 and March 6, 2018, encompassing the preparation for the election campaign and the reactions to its outcome.

We asked some Italian voters to select a set of thirty representative public Facebook pages, six for each of the following five political orientations: center-left, far-right, left, five-star movement (M5S) and right. Appendix Appendix A contains the selection of representative pages and their respective political orientations mapping. The classification of publishers into political categories is debatable, but our focus in this paper is on the methodology rather than on specific political conclusions. Moreover, most of our results are detailed on a per-publisher basis, as a measurement-based political orientation classification is out of scope of this paper(see, e.g [70]). Then, we created six virtual Facebook users, henceforth also referred to as *bots*. Each bot followed *all* the thirty selected *pages*. We gave to five bots a specific political orientation, by making each of them "like" pages from the corresponding publishers. The sixth bot does not "like" any page, i.e., it has no political orientation. We call it *undecided*.

Each bot kept open an Internet browser window (Firefox or Chrome) accessing the Facebook page. The bots were instrumented to collect data on the posts to which they were exposed. To that aim, a browser extension named Facebook Tracking Exposed [3] was developed. The extension auto-scrolls the Facebook window at pre-established instants of the day. Every auto-scroll produces a set of posts which are stored at a local database. Each set of posts is referred to as a *snapshot*. Each bot was scheduled to collect thirteen snapshots per day. Snapshots were collected once every hour, from 7 am to 7 pm (Italian local time). Each post appearing in a snapshot counts as a post *impression*. At each bot, the developed browser extension collects all impressions and records their corresponding publisher, publication time, impression time, content, number of "likes" and number of shares. We also have a second dataset which contains the set of *all* posts published by the thirty pages during the interval of interest, as provided by the Facebook API. This dataset is used to study what users would experience in the absence of filters, or in the presence of alternative filters.

Information about impressions used to be available in 2015, in a deprecated version of the Facebook API. In any case, that information was not necessarily reliable as recognized by Facebook itself [34]. For such reasons, we believe that the developed browser extension and the methodology described in this section constitute important building blocks to promote transparency in the Facebook ecosystem.

2.3. Measurement challenges

Gaps in measurements: During our measurement campaign, we experienced measurement gaps due to two reasons: 1) the computer running a bot went down, due to unforeseen glitches such as lack of power and 2) at random points in time, either Facebook or related applications, such as the browser itself, solicit human interaction (e.g., by showing a pop-up requiring users to answer simple questions before proceeding). We denote by S_i the number of snapshots collected by the i-th bot. In our experiments, the bots are indexed from 1 to 6, denoting center-left, far-right, left, M5S, right and undecided orientations. The values of S_i equal 577, 504, 623, 674, 655, 576, for $i = 1, \ldots, 6$. To account for the different number of snapshots, all the reported results rely on values averaged across snapshots rather than quantities that depend on the absolute number of snapshots.

Small number of bots: we use six bots to capture different perspectives on the Facebook dynamics. Each bot provides a *personal perspective* on the system, which is well aligned with our goals. Although we considered a small population size, we believe that the limited points of view provided by the six bots already shed important insights on the biases introduced by Facebook. In particular, the consistent biases observed in our dataset, reported in the sections that follow, indicates that the collected sample is representative.

2.4. Metrics of interest

We define our key metrics of interest that will be obtained from the dataset generated by the experiment. We consider the top K positions of the News Feed of each user.

Definition 1 (visibility)**.** *Let π_{ij} be the fraction of snapshots from user i that contain at least one post from publisher j.*

Definition 2 (occupancy)**.** *Let N_{ij} be the average number of posts of publisher j in the News Feed of user i.*

We refer to π_{ij} and N_{ij} as the *visibility* and the *occupancy* of publisher j at News Feed i, respectively. The *normalized occupancy* is given by N_{ij}/K. The visibility and the (normalized) occupancy are two metrics of exposure of publishers to users [6].

Definition 3 (hit probability)**.** *Let h_{ij} be the probability that user i sees (or clicks) on a post of publisher j.*

We refer to h_{ij} as the hit probability of publisher j at user i. Then, $h_{ij} = N_{ij}/K$ if user i goes through all the top posts in the News Feed, and $h_{ij} = \pi_{ij}$ if he/she picks uniformly at random a single post in the News Feed.

3. Empirical findings

In the following two sections, we report our empirical findings from the perspective of publishers and users.

3.1. The effects of filtering on publishers

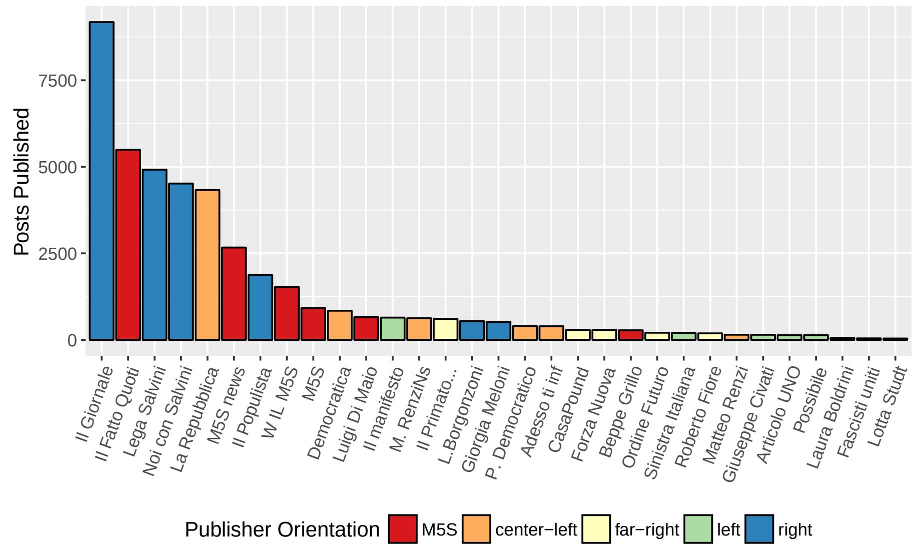
Next, we report findings on the behavior of the publishers and their general effect on users' News Feeds. Figure 1(a) shows the number of unique posts per publisher. We denote by C_j the number of posts of publisher j. This information was collected directly from the Facebook API. A few publishers generated thousands of posts during the considered time frame, whereas the majority generated tens of posts.

Figure 1(b) shows the number of impressions per publisher. This information was collected from our Facebook extension. Publishers are ordered based on the number of posts generated and seen in Figures 1(a) and 1(b), respectively. It is worth noting that the distinct order at which publishers appear in those figures is fruit of the filtering experienced by the users. In what follows, such filtering is further analyzed through measurements (Section 3.2), models (Section 4) and a combination of the two (Sections 5 and 6).

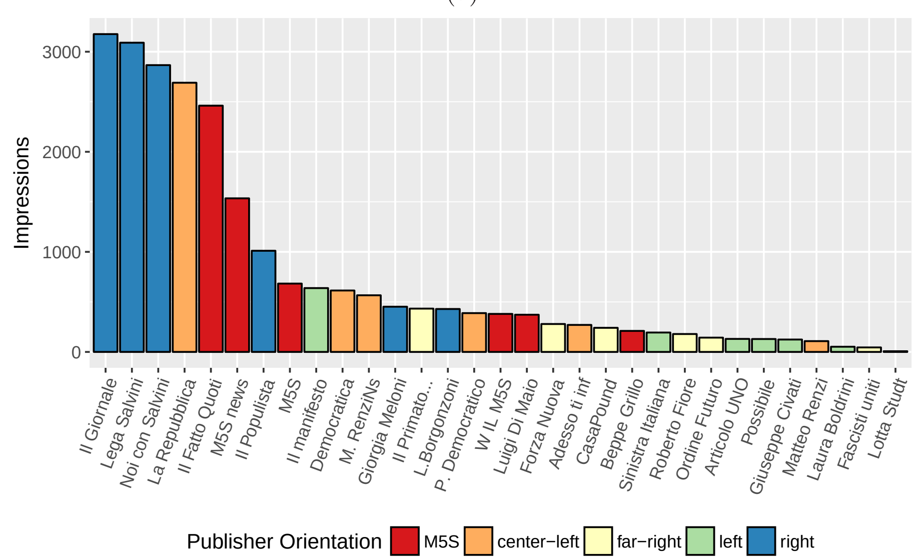
3.2. The effects of filtering on users

The effect of filtering is stronger at the topmost News Feed position. Figures 2(a) and 2(b) show the normalized publisher occupancy, as a function of the News Feed size (the corresponding visibilities are reported in Appendix Appendix B). The publishers are colored according to their political orientation (Fig. 2(a)) and user preferences (Fig. 2(b)). Figure 2(a) shows that the occupancy distribution over the five orientations changes with the considered News Feed size.

In Figure 2(b), a publisher is colored in blue (resp., red) at a given bot if the bot "likes" (resp., does not "like") the corresponding publisher. Note that, except for the right and far-right bots, the normalized occupancy of publishers that users "like" is maximum at the topmost position, achieving more than 70% at the center-left oriented bot. The right-oriented bot achieved a similar normalized occupancy when $K = 10$. The noteworthy bias on the topmost position must be placed under scrutiny, as there is a strong correlation between



(a)

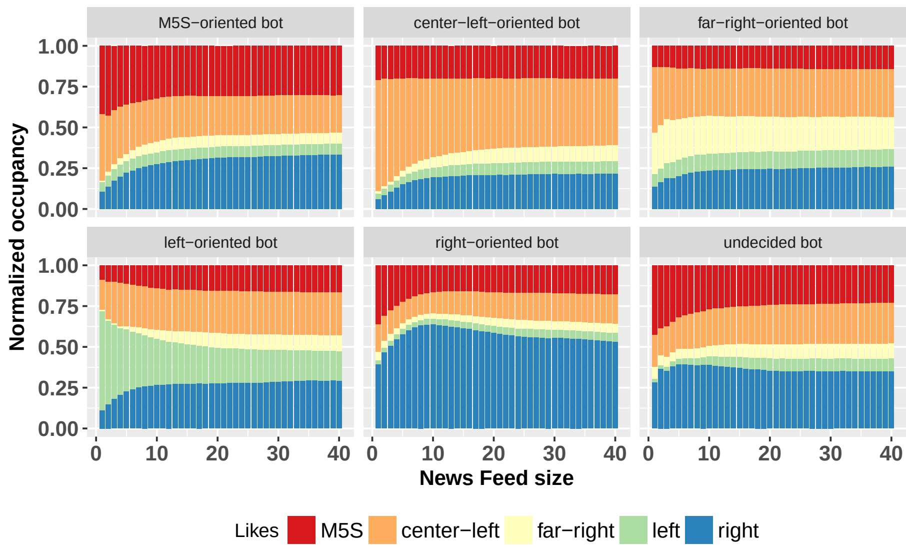


(b)

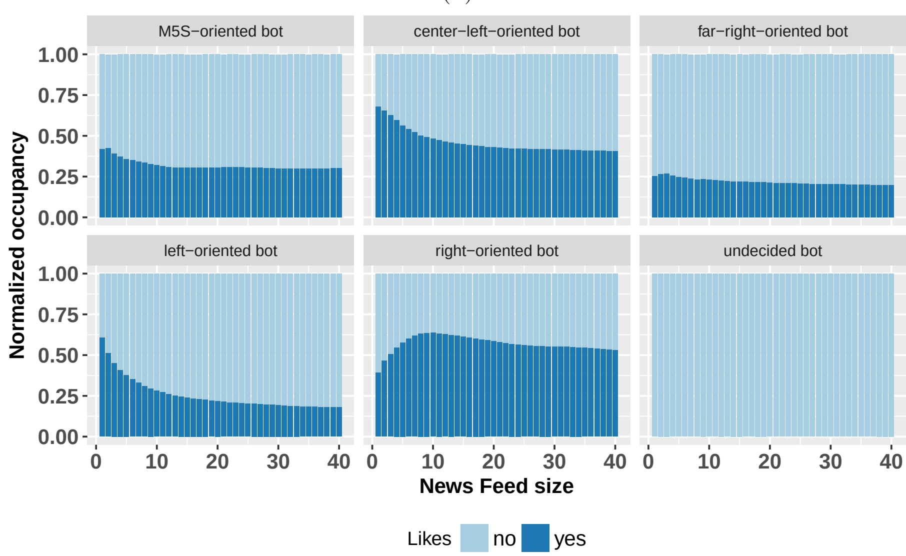
Figure 1: (a) Posts per publisher, (b) posts per publisher that were seen by at least one of our bots.

post positions and click rates [6, 82]. These figures also reveal that the amount of exposure to cross-cutting contents depends on the size of the News Feed.

Occupancy is impacted by orientation. Figure 2(a) also shows that the oc-



(a)



(b)

Figure 2: Normalized occupancy as a function of K, classified by (a) publisher orientations and (b) user preferences.

cupancies are impacted by the orientation of the bots. For instance, the News Feed of the bot with a center-left orientation was occupied mostly by center-left (red) publishers. As a notable exception, center-left posts were prevalent in the News Feed of the bot with a far-right orientation, where far-right posts are responsible for roughly 25% of the normalized occupancy. Nonetheless, the occupancy of far-right posts in that bot was still the highest among all bots.

Noticeable publishers selection. The bars in Figure 3 show the total number of impressions per publisher in the topmost position of the News Feed of each bot (the color of the bars indicates orientation). For the sake of readability, only publishers that achieved a normalized occupancy larger than 5% are represented in this figure. The black dots correspond to the number of posts created by each publisher (the publishers are ordered by the number of posts generated). Figure 3 shows that only a small subset of publishers are represented in topmost positions. For example, the center-left bot sees primarily posts from two of the publishers that it "likes". Moreover, the number of impressions per publisher is not proportional to the number of posts the publisher generated, a further indication of a filtering effect from News Feed algorithm.

Neutral users are also exposed to non-uniform filtering. It is worth noting that filtering affects also the "undecided" bot, with some publishers over-represented in the News feed.

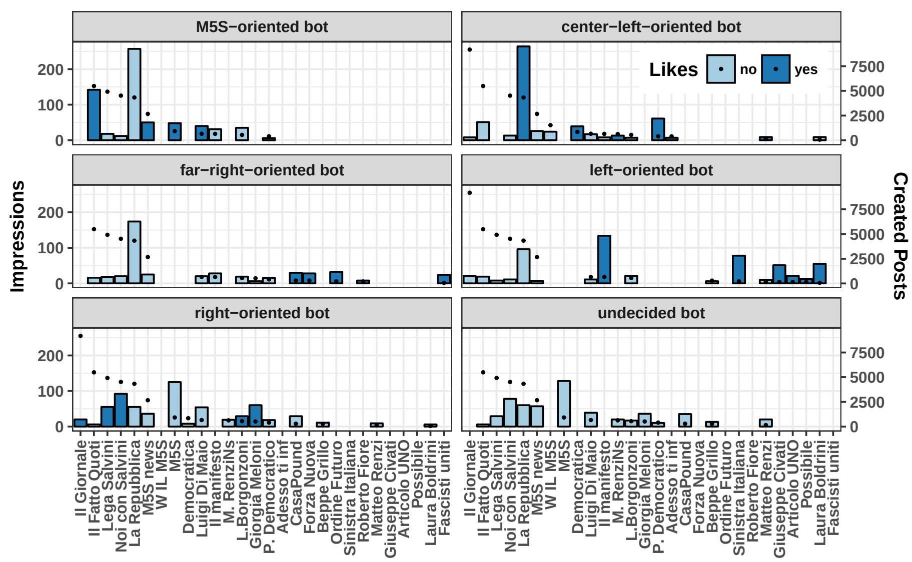


Figure 3: Publishers impressions at the six bots (bars colored by preferences) and number of created posts (black dots). The number of created posts is represented when at least one impression from the corresponding publisher was observed.

4. News Feed model

Next, we present the proposed News Feed model to derive occupancy and visibility metrics. We start by presenting the basic rationale behind the model in Section 4.1. Then, the model is introduced in Section 4.2.

4.1. Insights on News Feed modeling

Next, we introduce some of the key ideas that inspire the analytical model introduced in the following section.

4.1.1. Queues, caches and the News Feed

In the simplest setting, posts are organized at each News Feed in a *first in, first out* (FIFO) fashion. Then, the personalization algorithm at the News Feed of user i filters posts from each of the publishers. Given the rate Λ_j at which publisher j creates posts, we denote by λ_{ij} the corresponding *effective arrival rate* at which posts from publisher j are published at the News Feed of user i.

We assume that a News Feed has K slots. Under the FIFO approximation described above, new posts are inserted at the top of the News Feed and each new arrival shifts older posts one position lower. A post is evicted from the News Feed when it is shifted from position K. Although this is a preliminary step to capture the real News Feed operation, it cannot capture the stronger filtering at the topmost News Feed positions that we observed in Section 3.

In the remainder of this paper, we consider a generalization of the FIFO model, which accommodates different residence times for different posts using time-to-live (TTL) counters [21, 49, 68]. Under the TTL model, every time a post is inserted into a News Feed, it is associated to a timer (TTL), and the content remains in the News Feed until its timer expires. In what follows, we further detail the similarities and differences between TTL caches and the News Feed.

4.1.2. The News Feed is a publisher-driven cache

Next, we leverage a recently established connection between timelines in OSNs and caches [68], as summarized in Table 1.

Triggering events In the News Feed, the insertion of new posts is triggered by their creation. In caches, in contrast, user requests typically lead to content insertion and eviction. When proactive caching and prefetching of content is considered, the proactive caching is still usually fruit of content requests [79, 9, 43].

Nature of contents and requests When News Feed users search for user-generated content, they are typically interested in a class of items related to a given category. The demand for News Feeds posts is elastic. Consider, for instance, a user interested in the latest headlines from his favorite newspaper, or a college student willing to learn about the latest developments of his football team. There may be multiple posts that satisfy the demands of such users. The literature of caches, in contrast, presumes that each of the cached items is uniquely identified and non-substitutable, and that demand is inelastic.

Classes of items The occupancy of caches is given by the items that it stores, where each item is uniquely identified. Caches do not store repeated items. For the purposes of this work, in contrast, posts stored in a News Feed are

Table 1: Comparison between News Feed and caches

	News Feed	Cache
Trigger event	post publishing	content request
Insertions and evictions	after a post creation or user engagement	after a miss
Nature of contents and requests		
Cache stores	multiple items of content from given class	at most one copy of a specific item
Requests for	general content classes	specific content items
Classes of items		
Control occupancy	of items by given publisher	of specific content items
Capacity		
Capacity	infinite (average K topmost positions more relevant)	finite

distinguished solely by their publisher, and the News Feed may store multiple posts from the same publisher. Whereas the News Feed behaves as a cache that can store multiple copies of items of the same class, in traditional caches each item corresponds to its own class.

Capacity For all practical purposes, the News Feed can be assumed to be infinite in size, i.e., the News Feed can admit all the published posts. Nonetheless, it is well known that the topmost positions of the News Feed receive higher visibility. This, in turn, motivates our study of the (average) topmost K positions of the News Feed.

4.2. TTL News Feed model

Next, we introduce the proposed News Feed model. To each content posted in a News Feed we associate a time-to-live (TTL) timer. The timer is set to T when the content is inserted, and is decremented at fixed time intervals. Once the timer reaches zero, it expires and the associated content is removed from the News Feed.

4.2.1. Why TTL model?

It is well known that Facebook uses recency as a parameter to show posts to users [18, 33]. TTL counters are a natural way to capture the perishable nature of posts. Furthermore, TTL counters are a flexible way to extend FIFO schemes. In general, TTL-based models are well-suited to represent objects with expiration times [48].

While proposing the TTL model of a News Feed, our aim is not to argue whether Facebook deploys TTL counters, which is out of the scope of this paper. Instead, our goal is to show that a simple model can already capture the dynamics of Facebook News Feed. Then, we leverage the flexibility of TTL counters to propose novel News Feed algorithms.

4.2.2. Model description

Let $\mathcal{I}$ be the set of I users, and let $\mathcal{J}$ be the set of J publishers: $\mathcal{J}_i$ denotes the set of publishers followed by user $i \in \mathcal{I}$. Publisher $j \in \mathcal{J}$ publishes posts according to a Poisson process with rate Λ_j. The total publishing rate is $\Lambda = \sum_{j=1}^{J} \Lambda_j$. Whenever a content is generated, it is immediately sent to the News Feed of all users. In what follows, we provide further details about how user i reacts to the content arrival.

Content and publisher classes We consider C content classes. Each content class corresponds to a set of posts published at a given user News Feed. In the most general case, each user-publisher pair is associated to a given content class. In that case, the class associated to the i-th user and j-publisher is denoted by the ordered pair (i, j).

Alternatively, we associate each user-publisher pair to one of two classes. Class l_i (resp., $\bar{l}_i$) is the class of contents generated by publishers that the i-th user "likes" (resp., does not "like"). We denote by $L(i, j)$ the indicator variable which characterizes the set of publishers that a user likes,

$$L(i,j) = \begin{cases} 1, & \text{if user } i \text{ "likes" publisher } j \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

Order of posts The simplest instance of the proposed model corresponds to a FIFO queue, wherein contents are ordered in the News Feed based on the instant at which they are posted, and new arrivals shift older posts (Section 4.1.1). The TTL model, in contrast, does not presume any pre-established ordering of posts in the News Feed. In particular, it is flexible to account for eventual rearrangements of posts. Note that the TTL model can be parameterized to capture the behavior illustrated in Figure 2, wherein the bias present in the topmost positions is different from that seen in the remainder of the News Feed. This occurs, for instance, if posts with larger TTL are placed on the top of the News Feed, producing different biases at different News Feed positions.

Timer classes We consider per-class content dynamics. For concreteness, except otherwise noted we let $L(i, j)$ be the class of contents generated by publisher j at the News Feed of the i-th user.

Whenever a content from class $L(i, j)$ is generated, it is inserted in the News Feed of the i-th user and a timer with value $T_{L(i,j)}$ is associated to that content. Even though we assume, for simplicity, that the initial timer values are set to a fixed constant, our analysis also holds if the initial values of the timers are sampled from a probability distribution with mean $T_{L(i,j)}$.

We expect that $T_1 \geq T_0$, i.e., contents generated by publishers that the user "likes" remain longer in the News Feed, when compared against those that the user does not "like". Nonetheless, we do not explicitly assume any relationship between T_1 and T_0. Instead, we perform simple consistency checks using the collected measurement data (Section 5).

4.2.3. Metrics of interest

Next, we derive the metrics of interest corresponding to an infinite capacity News Feed. Recall that we assume that users scroll up to an average of K News Feed positions, i.e., we study the visibility and occupancy of an average of K topmost positions. To simplify notation we drop the explicit dependence of metrics and corresponding variables on the value of K, e.g., denoting $N_{ij}(K)$ and $T_{L(i,j)}(K)$ simply as N_{ij} and $T_{L(i,j)}$.

The occupancy of the j-th publisher at the i-th News Feed, N_{ij}, follows from Little's Law and is given by

$$N_{ij} = \Lambda_j T_{L(i,j)} \tag{2}$$

The expected number of slots occupied in a News Feed is given by the sum of N_{ij}, for all j,

$$\sum_{j \in \mathcal{J}} N_{ij} = K. \tag{3}$$

The visibility of publisher j at the News Feed of user i, π_{ij}, is given by

$$\pi_{ij} = 1 - e^{-N_{ij}}. \tag{4}$$

The equation above follows from the observation that the dynamics of posts by the j-th publisher at the News Feed of user i are given by an M/G/∞ queue. Arrivals of posts occur with rate Λ_j, and each arrival remains in the News Feed for an average of $T_{L(i,j)}$ time units. The probability that there is at least one customer in the M/G/∞ equals the probability that there is at least one post from publisher j at the News Feed of user i, and is given by (4).

4.2.4. Special case: FIFO News Feed

When all posts are associated to the same TTL T, and contents are inserted in the News Feed in the order that they are created, the TTL model behaves as a FIFO queue, as described in Section 4.1.1. If we further allow posts to be filtered, we refer to the resulting model as a *filtered FIFO* model. Let p_{ij} be the filtering probability. In a filtered FIFO model, we let $T = 0$ with probability $1 - p_{ij}$, and $T = \overline{T}$ otherwise, where $\overline{T}$ is a fixed and given constant, $\overline{T} > 0$.

As before, we assume that publisher $j \in \mathcal{J}$ publishes posts according to a Poisson process with rate Λ_j. Recall that the total publishing rate is given by $\Lambda = \sum_{j=1}^{J} \Lambda_j$. In the filtered model, let $\lambda_{ij} \leq \Lambda_j$ be the effective arrival rate of posts published by j in the News Feed of user i. Then, $\lambda_{ij} = p_{ij}\Lambda_j$.

Under the filtered FIFO model, (2) together with (3) imply that

$$\overline{T} = \frac{K}{\sum_k \lambda_{ik}}. \tag{5}$$

The visibility π_{ij} is given by (4), where

$$N_{ij} = \frac{\lambda_{ij} K}{\sum_l \lambda_{il}}. \tag{6}$$

If we further assume that $p_{ij} = p$ for all user-publisher pairs, we obtain the uniformly filtered FIFO model, where

$$N_j = \frac{\Lambda_j K}{\Lambda}. \tag{7}$$

As we assume that all users follow the same set of sources, under the uniformly filtered FIFO model the expected occupancy of publisher j at user i, N_{ij}, is the same for all users. Therefore, in this case we denote it simply as N_j.

4.2.5. Finite size FIFO News Feed

The analysis presented above assumes an infinite size News Feed, wherein users are interested, on average, at the topmost K positions. Alternatively, consider a finite size FIFO News Feed, which can accommodate up to K posts. We assume that a News Feed has K slots, new posts are inserted at the top of the News Feed and each new arrival shifts older posts one position lower. A post is evicted from the News Feed when it is shifted from position K.

We denote by λ_i the aggregate rate of posts published in the News Feed of user i, $\lambda_i = \sum_{j=1}^{J} \lambda_{ij}$. We further let $\lambda_{i,-j}$ be the arrival rate of posts in the News Feed of user i from all publishers other than j, $\lambda_{i,-j} = \lambda_i - \lambda_{ij}$.

The occupancy of contents of publisher j follows from Little's law and is given by

$$N_{ij} = \lambda_{ij} K / \lambda_i. \tag{8}$$

The visibility of publisher j is given by

$$\pi_{ij} = 1 - \left(\frac{\lambda_{i,-j}}{\lambda_i} \right)^K = 1 - \left(1 - \frac{N_{ij}}{K} \right)^K, \tag{9}$$

and the rationale goes as follows. After every new arrival, with probability $\lambda_{i,-j}/\lambda_i$ the topmost post of publisher j will be shifted down by one unit. After K consecutive shifts, which occur with probability $(\lambda_{i,-j}/\lambda_i)^K$, publisher j will not be visible at the News Feed of user i. When $K = 1$ we have $N_{ij} = \pi_{ij}$. For large values of K, (9) can be approximated by (4).

Under the FIFO instances of the model presented above, posts are filtered uniformly at random. For this reason, such instances yield simple baselines against which the filtering effects introduced by the News Feed algorithm can be compared. In Section 6 we revisit the FIFO model under this perspective.

5. A Model-based perspective at the measurement findings

In this section, we take a model-based perspective at the measurement findings. First, we implicitly account for user "likes" through their impact on the effective arrival rates in Section 5.1. Then, we explicitly account for "likes" in Section 5.2.

5.1. Indirectly accounting for "likes": a multi-class perspective on measurements

We validate the proposed model using data from the 2018 Italian elections, through a multi-class perspective on the measurements. Each user-publisher pair is associated to a class. Class (i, j) is associated to the i-th user and the j-th publisher, and corresponds to an effective arrival rate of λ_{ij} posts per snapshot.

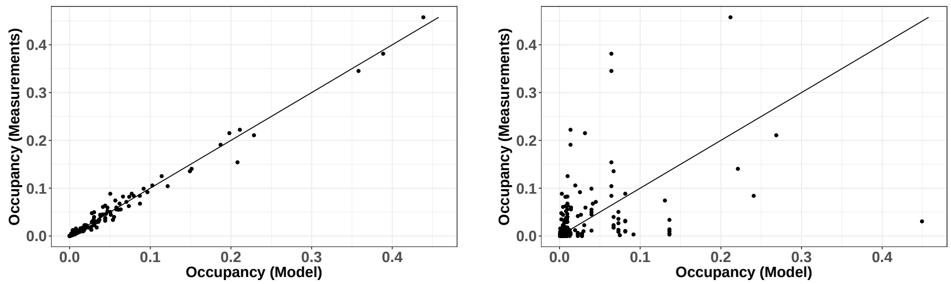
Variable	description
j	j-th publisher
i	i-th News Feed user
Λ_j	post creation rate by publisher j
λ_{ij}	arrival rate of posts from j at user i
λ_i	total arrival rate of posts at user i
$L(i,j)$	1 if user i likes publisher j, 0 otherwise
	TTL model variables
T_l	expected TTL for posts from class l, $l = L(i,j)$
$T_i^{(l)}$	expected TTL for posts from class l, with user discrimination, $l = L(i,j)$
T_{ij}	expected TTL for posts from publisher j at user i
$w_{L(i,j)}$	weight associated to class-$L(i,j)$
	Metrics of interest as estimated by the model
h_{ij}	hit probability of publisher j at user i
π_{ij}	visibility of publisher j at user i
N_{ij}	occupancy of publisher j at user i
	Metrics of interest as obtained from measurements
$\tilde{\pi}_{ij}$	measured visibility of j at i
$\tilde{N}_{ij}$	measured occupancy of j at i
	Measurements
I_{ij}	number of impressions from publisher j at user i (counting repeated posts multiple times)
Q_{ij}	number of unique posts from publisher j at user i
S_i	number of snapshots collected by user i
C_j	number of unique posts created by publisher j
G_l	number of user-publisher-post tuples corresponding to user-publisher pairs in class l
I_l	number of impressions with user-publisher in class l

Table 2: Table of notation

We start by introducing some additional notation regarding our measurements. The notation is summarized in Table 2. Let Q_{ij} be the measured numbered of unique posts from publisher j at user i. Let S_i be the number of snapshots taken by user i. Then, the measured effective arrival rate is given by

$$\tilde{\lambda}_{ij} = \frac{Q_{ij}}{S_i}. \tag{10}$$

Note that $\tilde{\lambda}_{ij}$ is modulated through "likes". In this section, at each bot i we assume that the same timer is used for all the publishers. Then, the TTL model is equivalent to FIFO (see Section 4.2.4). Publishers' presence in the News Feed is discriminated upstream by filtering their posts before they arrive to the News Feed, i.e., through rates λ_{ij}. For this reason, in this section we use model equations (4)-(6) with the arrival rate at a bot, λ_{ij}, set to the measured one, $\tilde{\lambda}_{ij}$. "Likes" are indirectly taken into account through rates.



(a) (b)

Figure 4: Model validation of occupancies for $K = 1$: (a) multi-class and (b) two-class.

Figure 4(a) compares measured occupancies against model predictions at the News Feed topmost position ($K = 1$). Each point corresponds to a user-publisher pair. A point ($x = N_{ij}, y = \tilde{N}_{ij}$) indicates that, for the given pair, an occupancy N_{ij} estimated by the proposed model using eq. (6) corresponds to a measured occupancy $\tilde{N}_{ij}$. Most of the points are close to the $\tilde{N}_{ij} = N_{ij}$ line, indicating the expressive power of the model. Appendix Appendix C.1 contains results for $K > 1$, accounting for visibility in addition to occupancy as the target metric.

Next, our goal is to quantitatively assess the expressive power of the multi-class model. To this aim, we conduct a linear regression followed by an hypothesis test on the coefficients produced by the linear regression. Let the measured occupancy be given as a function of the model-based occupancy as follows,

$$\tilde{N}_{ij} = \beta_1 N_{ij} + \beta_0. \tag{11}$$

The null and alternative hypotheses are given by

- H_0: there is no relationship between $\tilde{N}_{ij}$ and N_{ij}, i.e., $\beta_1 = 0$;
- H_a: there is relationship between $\tilde{N}_{ij}$ and N_{ij}, i.e., $\beta_1 \neq 0$

The *p-value* for $\beta_1 = 0$ is less than 2^{-16}, allowing us to reject the null hypothesis. We repeated the test for all values of K ranging from 1 to 30, and obtained similar results as indicated in Table 3.

5.2. Directly accounting for "likes": a two-class measurement analysis

In this section, we explicitly account for user "likes" in the News Feed occupancies. To this aim, we divide the publisher-user pairs into two classes, and show the expressive power of the model through a simple parameterization which involves only two parameters, T_1 and T_0, corresponding to the TTL of posts from publishers that users "like" and do not "like", respectively.

		Occupancy			Visibility		
Model	K	p-value	RMSE	R^2	p-value	RMSE	R^2
Multi-	1	$< 2^{-16}$	0.01	0.98	$< 2^{-16}$	0.02	0.98
-class	30	$< 2^{-16}$	0.28	0.93	$< 2^{-16}$	0.04	0.98
Two-	1	1.66^{-7}	0.07	0.13	4.56^{-8}	0.06	0.15
-class	30	$< 2^{-16}$	0.90	0.51	$< 2^{-16}$	0.25	0.49

Table 3: Summary of hypotheses test results

Let I_{ij} be the number of impressions from publisher j at user i. Let I_1 be the number of impressions at News Feeds of users who "like" the publishers of the corresponding impressions (I_0 is similarly defined). Then,

$$I_l = \sum_{\forall (i,j) | L(i,j)=l} I_{i,j}. \tag{12}$$

Correspondingly, let G_l (resp. G_0) be the number of posts generated, counted as many times as the number of users who "like" (do not "like") the publishers who generated these posts. Recall that C_j is the number of unique posts created by publisher j. Then,

$$G_l = \sum_{\forall (i,j) | L(i,j)=l} C_j. \tag{13}$$

The estimate of the TTL associated to class l is given by

$$\tilde{T}_l = I_l / G_l, \quad l \in \{0, 1\}. \tag{14}$$

In Figure 4(b) each point corresponds to a user-publisher pair. As in Figure 4(a), we let $K = 1$ (results for $K = 30$ are presented in Appendix Appendix C.1). In Figure 4(b), a point $(x = N_{ij}, y = \tilde{N}_{ij})$ indicates that, for the given pair, an occupancy N_{ij} estimated by the proposed model using eq. (2) corresponds to a measured occupancy of $\tilde{N}_{ij}$. We also resort to the same sort of hypothesis tests described in the previous section, and reject the null hypothesis according to which there is no relationship between the model and the measurements. A summary of the measurement results, for $K = 1$ and $K = 30$, is presented in Table 3.

The accuracy of the model can also be assessed through the R^2 score, which ranges between 1 and 0. An R^2 score of 1 indicates that the variance in the target variable is fully explained by the model. As expected, the R^2 scores (resp., *p-values*) of the two-class model are smaller (resp., larger) than those of the multi-class model. In addition, we observe that the predictive power of the two-class model when applied to the topmost position ($K = 1$) is significantly lower when compared against its application to the remainder of the News Feed ($K = 30$). This is partly explained by the fact that the bias is stronger at the topmost position (Section 3.2). We leave a more detailed analysis of simple models for the topmost position as subject for future work.

Summary In this section, we evaluated the explanatory power of the model in light of the measurements collected during the Italian 2018 elections. In particular, we have shown that a very simple instance of the model with two parameters is already able to capture the occupancies experienced by users (Section 5.2). In addition, we indicated that a multi-class instance of the model, wherein the number of parameter equals the number of publishers times the number of bots (180 in the experiment), produces occupancy estimates with higher accuracy, at the expense of additional complexity (Section 5.1). In the sections that follow, we leverage the proposed model for mechanism design purposes.

6. A fairness-based News Feed mechanism

Next, we leverage the proposed News Feed model to derive a fairness-based mechanism to design News Feeds. We present the problem formulation (Section 6.1), followed by its general solution (Section 6.2) and by an analysis of α-fair utility functions (Section 6.3). Then, we bridge the utility maximization framework and measurements to illustrate the applicability of the mechanism (Section 6.5).

6.1. Utility maximization formulation

We associate to each user-publisher pair a utility function $U_{ij}(h_{ij})$ which is an increasing, strictly concave and continuously differentiable function of h_{ij} over the range $0 \leq h_{ij} \leq 1$. Recall from Section 2.4 that h_{ij} denotes the hit probability of publisher j at user i, where the hit probability captures the exposure of publishers to users.

We further assume that utilities are additive. Then, the goal is to maximize the sum of the utilities for each of the individual publishers [49, 21]. The optimization problem is posed as follows,

$$\max \quad \sum_{j=1}^{J} U_{ij}(h_{ij}) \tag{15}$$

$$\text{s.t.} \quad \sum_{j=1}^{J} N_{ij} = K, \qquad N_{ij} \geq 0 \tag{16}$$

where h_{ij} are concave and non-decreasing functions of N_{ij}. As discussed in Section 2.4, we consider two possible instantiations of the hit probability, corresponding to the normalized occupancy, given by N_{ij}/K, and the visibility π_{ij}, given by (4) and (9) under the TTL and finite capacity FIFO models, respectively. In summary,

$$h_{ij}(N_{ij}) = \begin{cases} N_{ij}/K, & \text{for normalized occupancy,} \\ 1 - e^{-N_{ij}}, & \text{for visibility (TTL model),} \\ 1 - (1 - N_{ij}/K)^K, & \text{for visibility (finite FIFO).} \end{cases}$$

In all cases above, $h_{ij}(N_{ij})$ is an increasing and concave function of N_{ij}. Let $\tilde{U}_{ij}(N_{ij}) = U_{ij}(h_{ij}(N_{ij}))$. It follows that

- $\tilde{U}_{ij}(N_{ij})$ is non-decreasiing, as U_{ij} and h_{ij} are so;
- $\tilde{U}_{ij}(N_{ij})$ is concave, as U_{ij} is concave and non-decreasing and h_{ij} is concave.

Therefore, the objective function (15) is equivalent to

$$\max \sum_{j=1}^{J} \tilde{U}_{ij}(N_{ij}). \tag{18}$$

The general solution to the unified problem formulation introduced above is presented in the sequel.

6.2. Problem solution

To solve the convex optimization problem posed above, we introduce its corresponding Lagrangian,

$$\mathcal{L}(\mathbf{N}, \beta) = \sum_{j} \tilde{U}_{ij}(N_{ij}) - \beta \sum_{j} (N_{ij} - K) \tag{19}$$

where $\mathbf{N}$ is the vector of N_{ij} values, and β is the Lagrange multiplier. Taking the derivative of the Lagrangian with respect to N_{ij},

$$\frac{\partial L}{\partial N_{ij}} = \tilde{U}'_{ij}(N_{ij}) - \beta. \tag{20}$$

An allocation $N^{\star}_{ij}$ is a global optimizer if and only if there exists $\beta^{\star}$ such that $\mathbf{N}^{\star}$ is feasible and

$$\begin{cases} \tilde{U}'_{ij}(N^{\star}_{ij}) - \beta^{\star} = 0, & \text{if} \quad N^{\star}_{ij} > 0, \\ \tilde{U}'_{ij}(N^{\star}_{ij}) - \beta^{\star} \le 0, & \text{if} \quad N^{\star}_{ij} = 0. \end{cases} \tag{21}$$

At the optimum, all the publishers that appear in the News Feed have assigned a space that equalize their marginal utilities (i.e. $N_{ij} > 0$ and $N_{ik} > 0$ imply that $\tilde{U}'_{ij}(N^{\star}_{ij}) = \tilde{U}'_{ik}(N^{\star}_{ik})$). The publishers that not appear have smaller marginal utility (i.e., $\tilde{U}'_{il}(N^{\star}_{il}) \le \tilde{U}'_{ij}(N^{\star}_{ij}) \quad \text{if} \quad N^{\star}_{ij} > 0$ and $N^{\star}_{il} = 0$). Remark that if $\tilde{U}'(0) = +\infty$ (e.g. if $U_{ij}(h) \propto \log(h)$), then necessarily $N^{\star}_{ij} > 0$ for each j.

The solution can be found by a water-filling type algorithm: we start from the null vector where no publisher appears in the timeline ($N_{ij} = 0$ for all j) and we gradually allocate space to the publisher(s) with largest marginal utility(ies).

6.2.1. Occupancy vs rate-based fairness

In the problem formulation and solution presented above, we accounted for occupancy-based rather than rate-based fairness [56]. To appreciate the distinction between the two sorts of fairness, consider the problem of setting the same average space to different publishers. Such occupancy-based allocation will penalize prolific publishers more strongly than a rate-based allocation where publisher rates are uniformly multiplied by a constant factor. In Section 6.5 we numerically contrast the effect of occupancy-based fairness against a baseline wherein occupancies are proportional to content generating rates.

6.3. News Feed fairness: α-fair utilities

The optimal allocation of News Feed space to publishers through problem (15) depends on the shape of the utility functions. The use of a given family of utility functions corresponds to the selection of a fairness criterion. In this section we characterize the optimal allocation under the usual concept of *α-fairness* as it is considered in communication networks [60, 49, 22, 21].

$$U_{ij}(h_{ij}) = \begin{cases} w_{ij}\frac{h_{ij}^{1-\alpha}}{1-\alpha}, & \alpha \leq 0, \alpha \neq 1. \\ w_{ij}\log(h_{ij}), & \alpha = 1. \end{cases} \tag{22}$$

In what follows we will consider the case where the hit probability coincides with the normalized occupancy.

6.3.1. Proportional fairness

Choosing $\alpha = 1$ yields proportional fairness. In this case $\tilde{U}_{ij}(h_{ij}) = w_{ij}\log(N_{ij}/K)$, implying that $N^\star_{ij} > 0$ for each publisher. From the first equation in (21), it follows that

$$\tilde{U}'_{ij}(N^\star_{ij}) = w_{ij}/N^\star_{ij} = \beta^\star. \tag{23}$$

Imposing the constraint (16) we obtain

$$\beta^\star = \sum_{j=1}^{J} w_{ij}/K.$$

Once the value of $\beta^\star$ is known, it can be substituted in (23) to yield

$$N^\star_{ij} = \frac{w_{ij}K}{\sum_k w_{ik}}, \qquad T^\star_{ij} = \frac{w_{ij}K}{\Lambda_j \sum_k w_{ik}}. \tag{24}$$

6.3.2. Potential delay fairness

If $\alpha = 2$, $U_{ij}(h_{ij}) = -w_{ij}/h_{ij}$, $U'(h_{ij}) = w_{ij}/h^2_{ij}$ and in a similar way we obtain

$$N^\star_{ij} = K\frac{\sqrt{w_{ij}}}{\sum_{j=1}^{J}\sqrt{w_{ij}}}, \qquad T^\star_{ij} = \frac{K}{\Lambda_j}\frac{\sqrt{w_{ij}}}{\sum_{j=1}^{J}\sqrt{w_{ij}}}. \tag{25}$$

6.3.3. Max-min fairness

Max-min fairness is the limiting case of α-fairness in the limit when α diverges [55]. In our case a max-min fair allocation corresponds to provide the same occupancy to each publisher. Then, we have:

$$N^\star_{ij} = K/J, \qquad T^\star_{ij} = K/(\Lambda_j J). \tag{26}$$

Note that the max-min fairness allocation is independent of the weights w_{ij}.

6.3.4. Summary

In this section, we presented expressions for publishers occupancy and visibility, under three fairness criteria. In the following section we show how to extend the obtained expressions to account for class-based metrics, and in Section 6.5 we compare the derived class-based metrics against baselines, illustrating a way to quantify biases from real measurements.

6.4. Class-based optimization

The framework introduced in the previous section for per-publisher optimization can be easily adapted to a per-class optimization. In this case we divide the user-publisher pairs into classes, and parameters are set in a per-class basis. To simplify presentation, we specialize the presentation to classes determined by user's "likes", where for each user i, we distinguish the class of publishers i likes (identified by $L(i,j) = 1$ in (1)), and the class of publishers that i follows without expressing likes (identified by $L(i,j) = 0$). We denote by $\lambda_i^{(l)}$ (resp., $N_{ij}^{(l)}$) the aggregate arrival rate (resp., occupancy) of posts of class l in the News Feed of user i,

$$\lambda_i^{(l)} = \sum_{j|L(i,j)=l} \Lambda_j, \qquad N_i^{(l)} = \sum_{j|L(i,j)=l} N_{ij}. \tag{27}$$

Let $T_i^{(l)}$ denote the TTL of posts of class l in the News Feed of user i. Then,

$$N_{ij} = \Lambda_j T_i^{(l)}, \quad \text{where } l = L(i,j). \tag{28}$$

The expression of $T_i^{\star(l)}$ for the three special fairness criteira considered in the previous section can be similarly derived, leading to:

$$T_i^{\star(l)} = \begin{cases} \dfrac{K}{\lambda_i^{(l)}} \dfrac{w_i^{(l)}}{\sum_{k=0}^{1} w_i^{(k)}}, & \text{proportional fairness,} \quad (29a) \\[2ex] \dfrac{K}{\lambda_i^{(l)}} \dfrac{\sqrt{w_i^{(l)}}}{\sum_{k=0}^{1} \sqrt{w_i^{(k)}}}, & \text{potential delay fairness,} \quad (29b) \\[2ex] \dfrac{K}{2\lambda_i^{(l)}}, & \text{max-min fairness.} \quad (29c) \end{cases}$$

6.4.1. Class-based vs publisher-based allocation

In a class-based fair allocation, space is allocated fairly across classes: class occupancies are determined by the specific fairness criteria, while, inside while inside a given class publisher occupancies are proportional to publishing rates as indicated by (28). Consider, for instance, a max-min fairness allocation. Then the class of "liked" publishers is posed to have the same average occupancy as the remaining publishers, but this does not translate into equal publisher occupancies. For example, in our experimental setting, each of the 6 publishers

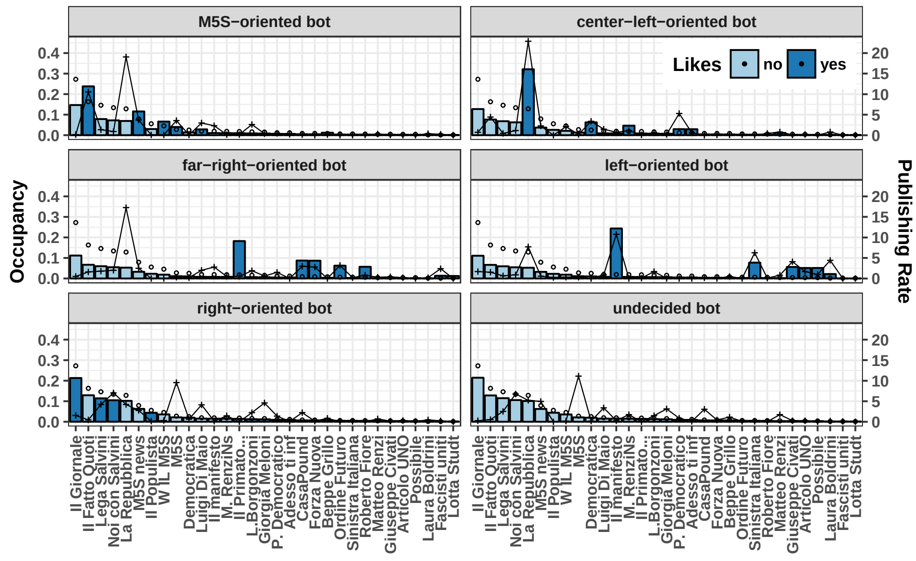


Figure 5: Publishers' occupancies under proportional fairness (bars colored by preferences), Facebook measured occupancies (crosses and lines) and publishing rates (circles), for the topmost position, at the six bots. At the undecided bot, proportional fairness occupancies decrease with respect to publishing rates, but measured occupancies deviate from baseline.

"liked" by bot i will on average occupy 1/12-th of bot-i timeline, while any of the other 24 publishers will on average get 1/48-th of it. Hence, the "liked publishers" are overall advantaged under max-min fairness allocation.

6.5. Italian election case study

Next, we illustrate how the proposed utility-based framework can shed further insights into the Italian election dataset. Throughout this section, we consider a proportional fairness allocation for the two-class model (Section 6.4). In particular here we show results for the topmost position and $w_i^{(l)} = 1$ for each i and l, while in Appendix Appendix E we report results for a broader set of weights and values of K.As discussed in Section 6.4.1, in this case the "liked" publishers get collectively as much timeline space as all the others.

Figure 5 shows different publishers' occupancies at the 6 bots when $K = 1$: occupancies measured at the bots (crosses and lines), occupancies computed by our model to maximize 2-class proportional fairness (bars) and publishers' posting rates (circles).

At the undecided bot, all publishers belong to the same class. Then, an allocation under proportional fairness yields occupancies proportional to publishing rates. In our measurements, in contrast, we observed that occupancies aren't proportional to rates (Sec 3.2), which indicates that the filtering effects are non-trivial even for neutral users. For all the other bots (except the right-oriented one), we observe that the "liked publishers" are favored under the

2-class fairness model. This is due to the fact that the 6 "liked" publishers get collectively as much space as all other 24 bots. Under each class, occupancies are split proportionally to publishing rates. Therefore, the "liked" publishers end up having less competition for space than other "non-liked" publishers with similar publishing rates. The situation is different at the right bot. Among the 6 right-oriented publishers "liked" by the right-oriented bot, we find some of the most prolific publishers in our dataset: their aggregate rate is almost equal to the aggregate rate of all the other 24 publishers. As a consequence, all the publishers get occupancies almost proportional to their publishing rate.

We observe that our simple 2-class model already qualitatively predicts some of the results observed in our traces: most of the "liked" publishers indeed also have larger measured occupancies (crosses and lines in Fig. 5). Nonetheless, there are still a number of exceptions. For example, "La Repubblica" exhibits a quite large occupancy independently of the orientation of the corresponding bot. This may be justified by the fact that "La Repubblica" is a newspaper, rather than a party or a candidate, and Facebook News Feed algorithm may be filtering less news. The same, however, does not hold for other newspapers appearing in the list. The occupancy of "Il Giornale" at the different News Feeds, for instance, is quite small (except at the right bot), even though it is the source with the largest publishing rate.

In order to quantify the discrimination among different providers introduced by our utility maximization allocation or by Facebook News Feed filtering algorithm we introduce a new metric. We define the (occupancy) bias as the difference between a given normalized occupancy estimate and the normalized occupancy that would have been obtained if timelines were operated according to the FIFO model without content filtering. We denote by $b_{ij}^{(m)}$ the occupancy bias of publisher j at user i. The symbol m refers to the scenario against which the FIFO baseline model is compared. It equals Face, PropF, MaxMinF and PotentF respectively when referring to the occupancies derived from raw Facebook measurements, proportional fairness, max-min fairness and potential fairness models.

Definition 4 (bias)**.** *The bias incurred by posts of publisher j at the News Feed of user i is given by*

$$b_{ij}^{(m)} = \frac{N_{ij}^{(m)} - N_j}{K}. \tag{30}$$

In the definition above, N_j is the baseline occupancy under FIFO as given by (7) and does not depend on the specific bot. We observe that $\sum_j b_{ij}^{(m)} = 0$, as the sum of occupancies at a given bot equals K. Note that the definition of bias is general, and can be coupled with different baseline models for occupancy (see Appendix Appendix D).

Discussion Statistical bias is a systemic deviation, e.g., of an estimator, with respect to the true value of a parameter. In this paper we use the term bias with a different, albeit intuitively related, meaning. We assume that users explicitly express their preferences by selecting which publishers they follow and like. Any deviation from the unfiltered occupancy, as estimated by the proposed model, is attributed to bias. This kind of bias is called *social bias* [5]. Note that our definition of bias does not necessarily entail a negative connotation, as biasing users towards their interests might be a desired feature.

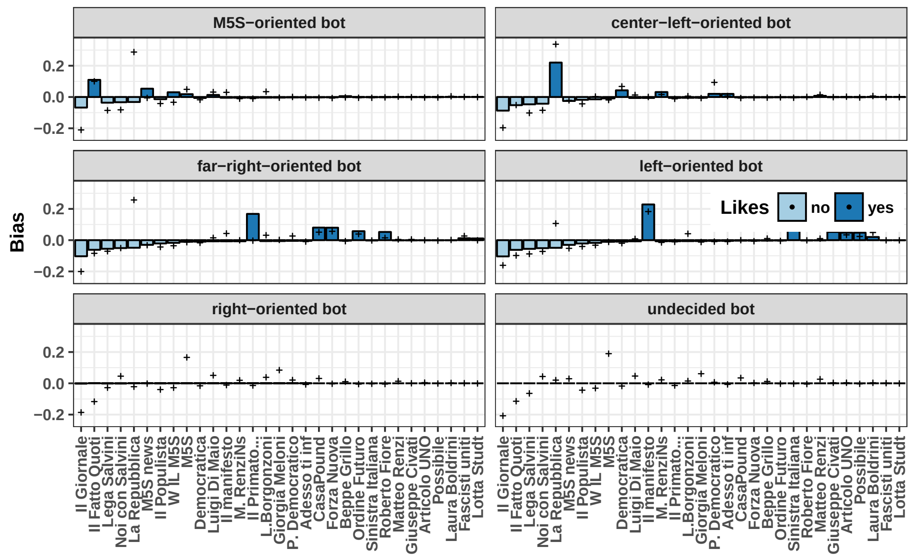


Figure 6: Proportional fairness bias (bars) and Facebook News Feed bias (crosses), for the six bots

Example Figure 6 numerically illustrate the behavior of bias, for the top publishers in the 2018 Italian election dataset, letting $K = 1$, respectively (recall that $b_{ij}^{(\mathrm{PropF})} = b_{ij}^{(\mathrm{MaxMinF})} = b_{ij}^{(\mathrm{PotentF})}$ as in this section we assume $w_i^{(l)} = 1$, and see Appendix Appendix D for additional results). Figure 6 supports the observations reported from Fig. 5. In particular we observe that $b_{ij}^{(\mathrm{PropF})}$ is null or almost null for the undecided bot and for the right bot, while for the others there are significant positive (resp., negative) biases for the "liked" (resp., "non-liked") contents. Facebook biases $b_{ij}^{(\mathrm{Face})}$ appear to be qualitatively aligned with those estimated through the 2-class proportional fairness model, , particularly at the center-left and left bots. Nonetheless, there are a number of exceptions, as discussed above.

We observe that the Facebook News Feed algorithm produces high biases even at the undecided bot. In addition, the bias at the topmost position does not reflect user preferences, specially at the far-right, right and undecided bots. Our analysis sheds light on some peculiar choices of Facebook that are difficult to explain even taking into account users' preferences as expressed through "likes."

6.6. Practical implications

The proposed mechanisms evidence the challenges involved in building a fair News Feed. Any fair solution must trade between conflicting goals such as 1) giving more exposure to favorite publishers, 2) reserving space for non-preferred publishers, 3) penalizing publishers that produce irrelevant content at high rates (e.g., spam) and 4) avoiding bias towards publishers that produce at low rates. We believe that the proposed mechanisms constitute a principled way to cope with such conflicting goals, through the use of utility functions.

Ultimately, users should be aware of the filtering that they are exposed to, and tune their utilities based on their needs. By promoting user awareness, the risk of amplifying filter bubbles should be mitigated. A brief discussion of the compliance of the proposed methodology to Facebook policies is presented in a short preliminary version of this work [45].

One way to implement the proposed mechanisms is through tools such as MIT Gobo [38] or FeedVis [31]. Such social media aggregators can be fed by posts from different sources, e.g., Facebook, Twitter and Youtube, providing filters that users can control. Therefore, users can explicitly set mechanisms to decide what is edited out of their News Feed. We envision that these tools can be coupled with variants of the News Feed control mechanisms proposed in this paper. Then, A/B tests may be used to select which mechanisms are best suited to different users based on their explicit and implicit feedback, e.g., obtained through questionnaires and click-rates.

7. Related work

The literature on Facebook News Feed includes topological aspects related to cascading structures and growth [42, 83, 16] and its effects on the creation of echo chambers and polarization [10, 11].

7.1. Social networks, TTL-counters and utility-based allocation

The proposed News Feed model relies on TTL counters. TTL-based caching mechanisms are versatile and flexible, and can be used to reproduce the behavior of traditional caching policies such as LRU and FIFO [37, 81, 21, 36, 64]. In this paper, we leverage the analytical tractability of TTL-based caches, showing how to adapt them for the purposes of modeling and optimization of the News Feed.

The implications of the limited budget of attention of users in OSNs have been previously studied by Reiffers-Masson *et al.* [69] and Jiang *et al.* [46]. In these two papers, the authors consider the problem of optimally deciding what to post and to read, respectively. Such works are complementary to ours. To the best of our knowledge, none of the previous works considered the problem of inferring the visibility of publishers from News Feed measurements, and using such measurements to parameterize models and propose utility-based mechanisms for News Feed design.

7.2. Fairness, accountability and transparency

The literature on fairness, accountability and transparency (FAT) is rapidly growing [35, 59, 87], accounting for its implications on social networks [84], risk score estimation [50, 8], recommender system [51, 89, 15, 75], resource allocation [54], individuals classification in order to prevent discrimination [26] and computational policy [77, 41], using tools such as causal analysis [90], quantitative input influence [20] and machine learning [88, 67].

Surveys and books on notions of fairness include those by Moulin [62], Zliobaite [91], Rmoei and Ruggieri [71], Narayanan [63] and Drosou *et al.* [25]. The later is a survey about diversity concepts from the information retrieval literature [86]. The mechanism design proposed in this paper is both a diversity-aware and fairness-aware allocation scheme, taking users' and publishers' perspectives, respectively.

ACM [39] introduces a set of principles intended to ensure fairness in the evolving policy and technology ecosystem: awareness, access and redness, accountability, explanation, data provenance, auditability, and validation and testing. We particularly focus on awareness, explanation and auditability, as we do not rely on the Facebook API to collect impressions. Algorithmic transparency is one of the cornerstones of the General Data Protection Regulation (GDPR), which stresses the importance of providing explanations for automatic recommendations [44]. The measurements, models and mechanisms proposed in this paper contribute to the development of GDPR-compliant policies, as the allocations derived from the proposed model-based mechanism are built on top of first principles.

Most of the previous literature on social fairness assumes that utility functions are non-parametric [90, 1, 85] or, in classification problems, that cost functions are linear combinations of false positive and false negative rates [67, 50, 17]. Convex utility functions, such as the α-fair family of utilities, are prevalent in the literature of networking and computer systems [60, 40, 49, 21, 74, 65, 12]. In this paper, we identify how parameterized utilities can be applicable to the analysis of social networks. We believe that such connection is a step towards promoting more dialogue between the networking and the online social network communities on the issue of fairness.

Algorithmic bias and forms to audit it were investigated in [5, 24, 73]. In [27, 52] it was shown that search engine rank manipulation can influence the vote of undecided citizens. The models proposed in this paper further foster accountability, by quantifying bias in the Facebook News Feed.

7.2.1. Public datasets and reproducible methodologies

The behavior of users searching for visibility was studied in [30, 14, 76]. Such studies are primarily based on small datasets. A notable exception is [7, 6, 2, 78, 13], who considered a massive dataset provided by Facebook through restrictive non-disclosure agreements. Datasets to assess Facebook publishers' visibilities are usually not made publicly available. Our work aims to contribute by filling that gap.

It is out of the scope of this paper to present a nuts-and-bolts perspective on how Facebook News Feed works. Instead, our goal is to provide a simple model that can explain the occupancy and visibility of different publishers, given a reproducible measurement framework. We profile Facebook, which is taken as a black box to be scientifically analyzed. This approach dates back to Skinner tests [28], and has been gaining significant attention in the literature of social networks [80].

8. Conclusion

We presented a framework encompassing reproducible measurements, analytical models and utility-based mechanisms for the analysis of Facebook News Feed algorithm. The analytical model enables quantitative what-if analysis to assess the bias introduced by the News Feed algorithm. The utility-based mechanisms shed light into novel directions towards the control of the News Feed.

Our measurements indicate that the News Feed algorithm currently tends to reinforce the orientation indicated by users about the pages they "like", by filtering posts and creating biases among the set of followed publishers. The effects of filtering are stronger at the topmost position where only a fraction of the set of publishers followed by the users was represented. We observed that a neutral user that did not "like" any page was also exposed to a noticeable bias.

Facebook mission is to "give people the power to build community." We believe that the measurements, model and tools presented in this work are one step further towards that goal, as they help evaluating algorithms' transparency and promote user awareness about the filtering process they are submitted to. Ultimately, such awareness is key to protect and empower Facebook users, communities, society and democracy as a whole [57].

Acknowledgements

This work was partially funded by the Brazilian-French project THANES (co-sponsored by FAPERJ and INRIA). D. Menasché was sponsored by a JCNE fellowship (FAPERJ), CNPq and in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. This work has also been supported by the French government, through the UCA JEDI "Investments in the Future" project managed by the National Research Agency (ANR) with the reference number ANR-15-IDEX-01.

References

[1] Rediet Abebe, Jon Kleinberg, and David C Parkes. 2017. Fair division via social comparison. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 281–289.

[2] Jamie E. Guillory Adam D. I. Kramer and Jeffrey T. Hancock. 2014. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences* (2014). https://doi.org/10.1073/pnas.1412469111

[3] Claudio Agosti. 2018. Facebook tracking exposed. https://facebook.tracking.exposed.

[4] Eitan Altman, Parmod Kumar, Srinivasan Venkatramanan, and Anurag Kumar. 2013. Competition over timeline in social networks. *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining - ASONAM '13* (2013).

[5] Ricardo Baeza-Yates. 2018. Bias on the web. *Commun. ACM* 61, 6 (may 2018), 54–61. https://doi.org/10.1145/3209581

[6] Eytan Bakshy, Solomon Messing, and Lada A Adamic. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science* 348, 6239 (2015), 1130–1132.

[7] Eytan Bakshy, Itamar Rosenn, Cameron Marlow, and Lada Adamic. 2012. The role of social networks in information diffusion. In *Proceedings of the 21st international conference on World Wide Web - WWW '12*. ACM Press, New York, NY, USA, 519. arXiv:1201.4145

[8] Chelsea Barabas, Karthik Dinakar, Joichi Ito Virza, Jonathan Zittrain, et al. 2017. Interventions over Predictions: Reframing the Ethical Debate for Actuarial Risk Assessment. *arXiv preprint arXiv:1712.08238* (2017).

[9] Ejder Bastug, Mehdi Bennis, and Mérouane Debbah. 2014. Living on the edge: The role of proactive caching in 5G wireless networks. *IEEE Communications Magazine* 52, 8 (2014), 82–89.

[10] Alessandro Bessi. 2016. Personality traits and echo chambers on Facebook. *Computers in Human Behavior* 65 (2016), 319 – 324. https://doi.org/10.1016/j.chb.2016.08.016

[11] Alessandro Bessi, Fabiana Zollo, Michela Del Vicario, Michelangelo Puliga, Antonio Scala, Guido Caldarelli, Brian Uzzi, and Walter Quattrociocchi. 2016. Users polarization on Facebook and Youtube. *PloS one* 11, 8 (2016), e0159641.

[12] Thomas Bonald and Laurent Massoulié. 2001. Impact of fairness on Internet performance. In *ACM SIGMETRICS Performance Evaluation Review*, Vol. 29. ACM, 82–91.

[13] Robert M Bond, Christopher J Fariss, Jason J Jones, Adam DI Kramer, Cameron Marlow, Jaime E Settle, and James H Fowler. 2012. A 61-million-person experiment in social influence and political mobilization. *Nature* 489, 7415 (2012), 295.

[14] Taina Bucher. 2012. Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. *New Media & Society* (2012). https://doi.org/10.1177/1461444812440159

[15] L Elisa Celis, Damian Straszak, and Nisheeth K Vishnoi. 2017. Ranking with Fairness Constraints. (2017), 1–32. https://doi.org/10.4230/LIPIcs.ICALP.2018.28 arXiv:1704.06840

[16] Justin Cheng, Jon Kleinberg, Jure Leskovec, David Liben-Nowell, Bogdan State, Karthik Subbian, and Lada Adamic. 2018. Do Diffusion Protocols Govern Cascade Growth? *Icwsm* (2018). https://www-cs.stanford.edu/people/jure/pubs/diffusion-icwsm18.pdf

[17] Sam Corbett-Davies, Emma Pierson, Avi Feller, Sharad Goel, and Aziz Huq. 2017. Algorithmic decision making and the cost of fairness. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 797–806.

[18] Tech Cruch. 2016. How Facebook News Feed Works. https://techcrunch.com/2016/09/06/ultimate-guide-to-the-news-feed/.

[19] Tech Cruch. 2017. Facebook now has 2 billion monthly users. https://techcrunch.com.

[20] Anupam Datta, Shayak Sen, and Yair Zick. 2016. Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems. In *Security and Privacy (SP), 2016 IEEE Symposium on*. IEEE, 598–617.

[21] Mostafa Dehghan, Laurent Massoulie, Don Towsley, Daniel Menasche, and Y. C. Tay. 2016. A utility optimization approach to network cache design. In *Proceedings of IEEE INFOCOM 2016*.

[22] Apostolos Destounis, Mari Kobayashi, Georgios Paschos, and Asma Ghorbel. 2017. Alpha fair coded caching. In *Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), 2017 15th International Symposium on*. IEEE, 1–8.

[23] Ranbir Dhounchak, Veeraruna Kavitha, and Eitan Altman. 2017. A Viral Timeline Branching Process to study a Social Network. In *29th International Teletraffic Congress PhD Workshop, {ITC} PhD Workshop 2017.*

[24] Nicholas Diakopoulos. 2013. Algorithmic accountability reporting: On the investigation of black boxes. *Tow Center for Digital Journalism A Tow/Knight Brief* (2013), 1–33. https://doi.org/10.1002/ejoc.201200111 arXiv:arXiv:1011.1669v3

[25] Marina Drosou, HV Jagadish, Evaggelia Pitoura, and Julia Stoyanovich. 2017. Diversity in big data: A review. *Big data* 5, 2 (2017), 73–84.

[26] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. 2012. Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference*. ACM, 214–226.

[27] Robert Epstein and Ronald E Robertson. 2015. The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences of the United States of America* 112, 33 (2015), E4512–21. https://doi.org/10.1073/pnas.1419828112

[28] Ziv Epstein, Blakeley H. Payne, Judy Hanwen Shen, Abhimanyu Dubey, Bjarke Felbo, Matthew Groh, Nick Obradovich, Manuel Cebrián, and Iyad Rahwan. 2018. Closing the AI Knowledge Gap. *CoRR* abs/1803.07233 (2018). arXiv:1803.07233 http://arxiv.org/abs/1803.07233

[29] Motahhare Eslami, Amirhossein Aleyasen, Karrie Karahalios, Kevin Hamilton, and Christian Sandvig. 2015. Feedvis: A path for exploring news feed curation algorithms. In *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing*. ACM, 65–68.

[30] Motahhare Eslami, Karrie Karahalios, Christian Sandvig, Kristen Vaccaro, Aimee Rickman, Kevin Hamilton, and Alex Kirlik. 2016. First I like it, then I hide it: Folk theories of social feeds. In *Proceedings of the 2016 cHI conference on human factors in computing systems*. ACM, 2371–2382.

[31] Motahhare Eslami, Aimee Rickman, Kristen Vaccaro, Amirhossein Aleyasen, Andy Vuong, Karrie Karahalios, Kevin Hamilton, and Christian Sandvig. 2015. "I always assumed that I wasn't really that close to [her]". *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15* (2015), 153–162. https://doi.org/10.1145/2702123.2702556

[32] Facebook. 2017. Controlling What You See in News Feed. https://www.facebook.com/help/335291769884272. (https://developers.facebook.com/docs/graph-api/).

[33] Facebook. 2018. https://newsroom.fb.com/news/2018/05/inside-feed-news-feed-ranking (2:30).

[34] Facebook. 2018. Graph API Reference. http://bit.ly/2HB1vzq.

[35] FAT conference. 2018. Conference on Fairness, Accountability, and Transparency (FAT*). https://fatconference.org/.

[36] Andrés Ferragut, Ismael Rodríguez, and Fernando Paganini. 2016. Optimizing TTL caches under heavy-tailed demands. In *ACM SIGMETRICS Performance Evaluation Review*, Vol. 44. ACM, 101–112.

[37] Nicaise Choungmo Fofack, Philippe Nain, Giovanni Neglia, and Don Towsley. 2012. Analysis of TTL-based cache networks. In *Performance Evaluation Methodologies and Tools (VALUETOOLS), 2012 6th International Conference on.* 1–10. https://doi.org/10.1016/j.comnet.2014.03.006

[38] MIT Media Lab's Center for Civic Media. 2018. GOBO. https://gobo.social/.

[39] Simson Garfinkel, Jeanna Matthews, Stuart S. Shapiro, and Jonathan M. Smith. 2017. Toward algorithmic transparency and accountability. *Commun. ACM* 60, 9 (2017), 5–5. https://doi.org/10.1145/3125780

[40] Ali Ghodsi, Matei Zaharia, Benjamin Hindman, Andy Konwinski, Scott Shenker, and Ion Stoica. 2011. Dominant Resource Fairness: Fair Allocation of Multiple Resource Types.. In *Nsdi*, Vol. 11. 24–24.

[41] Geoffrey Gilbert, Petra Ahrweiler, Peter Barbrook-Johnson, Kavin Narasimhan, and Helen Wilkinson. 2018. Computational Modelling of Public Policy: Reflections on Practice. *Journal of Artificial Societies and Social Simulation* 21, 1 (2018), 1–14.

[42] Minas Gjoka, Maciej Kurant, Carter T. Butts, and Athina Markopoulou. 2010. Walking in facebook: A case study of unbiased sampling of OSNs. *Proceedings - IEEE INFOCOM* (2010). https://doi.org/10.1109/INFCOM.2010.5462078 arXiv:0906.0060

[43] Negin Golrezaei, Andreas F Molisch, Alexandros G Dimakis, and Giuseppe Caire. 2013. Femtocaching and device-to-device collaboration: A new architecture for wireless video distribution. *IEEE Communications Magazine* 51, 4 (2013), 142–149.

[44] Bryce Goodman and Seth Flaxman. 2016. EU regulations on algorithmic decision-making and a "right to explanation". In *ICML workshop on human interpretability in machine learning (WHI 2016), New York, NY. http://arxiv. org/abs/1606.08813 v1.*

[45] Eduardo Hargreaves, Claudio Agosti, Daniel Menasché, Giovanni Neglia, Alexandre Reiffers-Masson, and Eitan Altman. 2018. Biases in the Facebook News Feed: a Case Study on the Italian Elections. *Proceedings of the IEEE/ACM International Conference on Social Networks Analysis and Mining (ASONAM 2018).* https://doi.org/10.1109/ASONAM.2018.8508659 arXiv:1807.08346

[46] Bo Jiang, Nidhi Hegde, Laurent Massoulié, and Don Towsley. 2013. How to optimally allocate your budget of attention in social networks. In *INFOCOM, 2013 Proceedings IEEE*. IEEE, 2373–2381.

[47] Jason J Jones, Robert M Bond, Eytan Bakshy, Dean Eckles, and James H Fowler. 2017. Social influence and political mobilization: Further evidence from a randomized experiment in the 2012 US presidential election. *PloS one* 12, 4 (2017), e0173851.

[48] Jaeyeon Jung, A.W. Berger, and Hari Balakrishnan. 2003. Modeling TTL-based Internet caches. In *Proceedings of IEEE INFOCOM 2003*, Vol. 1. IEEE, 417–426. http://ieeexplore.ieee.org/document/1208693/

[49] Frank Kelly. 1997. Charging and rate control for elastic traffic. *European Transactions on Telecommunications* 8, 1 (1997), 33–37. https://doi.org/10.1002/ett.4460080106

[50] Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. 2017. Inherent Trade-Offs in the Fair Determination of Risk Scores. In *Proceeding of the 8th Innovations in Theoretical Computer Science Conference, ITCS 2017*.

[51] Subhashini Krishnasamy, Rajat Sen, Sanjay Shakkottai, and Sewoong Oh. 2016. Detecting Sponsored Recommendations. *ACM Transactions on Modeling and Performance Evaluation of Computing Systems* 2, 1 (2016), 1–29. https://doi.org/10.1145/2988543 arXiv:arXiv:1504.03713v1

[52] Juhi Kulshrestha, Motahhare Eslami, Johnnatan Messias, Muhammad Bilal Zafar, Saptarshi Ghosh, Krishna P Gummadi, and Karrie Karahalios. 2017. Quantifying Search Bias: Investigating Sources of Bias for Political Searches in Social Media. *20th ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW 2017)* (2017), 417–432.

[53] Juhi Kulshrestha, Muhammad Bilal Zafar, Lisette Espin Noboa, Krishna P Gummadi, and Saptarshi Ghosh. 2015. Characterizing Information Diets of Social Media Users. In *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM'15)*. 218–227.

[54] Tian Lan, David Kao, Mung Chiang, and Ashutosh Sabharwal. 2010. An Axiomatic Theory of Fairness in Network Resource Allocation. In *Proceedings of IEEE INFOCOM*. 1–9. https://doi.org/10.1109/INFCOM.2010.5461911 arXiv:arXiv:0906.0557v4

[55] Jean-Yves Le Boudec. 2016. *Rate adaptation, congestion control and fairness: A tutorial.* Technical Report. http://moodle.epfl.ch/file.php/523/CC{_}Tutorial/cc.pdf

[56] Richard T. B. Ma Y. C. Tay Lianjie Shi, Xin Wang. 2018. Weighted Fair Caching: Occupancy-Centric Allocation for Space-Shared Resources. In *IFIP WG Performance 2018*.

[57] Gideon Lichfield. 2018. Technology is threatening our democracy. How do we save it? *MIT Technology Review* (September/October 2018).

[58] Alfred Lua. 2018. Decoding the Facebook Algorithm: A Fully Up-to-Date List of the Algorithm Factors and Changes. https://blog.bufferapp.com/facebook-news-feed-algorithm.

[59] Microsoft Research. 2018. FATE: Fairness, Accountability, Transparency, and Ethics in AI. https://www.microsoft.com/en-us/research/group/fate/.

[60] Jeonghoon Mo and Jean Walrand. 2000. Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on networking* 8, 5 (2000), 556–567.

[61] Adam Mosseri. 2018. News Feed Ranking in Three Minutes Flat. https://newsroom.fb.com/news/2018/04/inside-feed-news-feed-ranking/.

[62] Hervé Moulin. 2004. *Fair Division and Collective Welfare.* MIT Press.

[63] Arvind Narayanan. 2018. Tutorial: 21 Fairness Definitions and Their Politics. https://fatconference.org/2018/livestream_vh220.html.

[64] Giovanni Neglia, Damiano Carra, and Pietro Michiardi. 2017. Cache policies for linear utility maximization. In *Proceedings of IEEE INFOCOM 2017*.

[65] Daniel Pérez Palomar and Mung Chiang. 2006. A tutorial on decomposition methods for network utility maximization. *IEEE Journal on Selected Areas in Communications* 24, 8 (2006), 1439–1451.

[66] Eli Pariser. 2011. *The Filter Bubble: What the Internet Is Hiding from You.* The Penguin Group.

[67] Geoff Pleiss, Manish Raghavan, Felix Wu, Jon Kleinberg, and Kilian Q Weinberger. 2017. On fairness and calibration. In *Advances in Neural Information Processing Systems*. 5684–5693.

[68] Alexandre Reiffers-Masson, Eduardo Hargreaves, Eitan Altman, Wouter Caarls, and Daniel S. Menasché. 2017. Timelines Are Publisher-Driven Caches: Analyzing and Shaping Timeline Networks. *SIGMETRICS Perform. Eval. Rev.* 44, 3 (Jan. 2017), 26–29.

[69] Alexandre Reiffers-Masson, Yezekael Hayel, and Eitan Altman. 2017. Posting Behavior Dynamics and Active Filtering for Content Diversity in Social Networks. *IEEE transactions on Signal and Information Processing over Networks* 3, 2 (2017), 376–387.

[70] Filipe Ribeiro, Lucas Henrique, Fabrício Benevenuto, Abhijnan Chakraborty, Juhi Kulshrestha, Mahmoudreza Babaei, and Krishna P. Gummadi. 2018. Media Bias Monitor: Quantifying Biases of Social Media News Outlets at Large-Scale. In *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM'18)*.

[71] Andrea Romei and Salvatore Ruggieri. 2014. A multidisciplinary survey on discrimination analysis. *The Knowledge Engineering Review* 29, 5 (2014), 582–638.

[72] Wilbert Samuel Rossi, Jan Willem Polderman, and Paolo Frasca. 2018. The closed loop between opinion formation and personalised recommendations. (2018), 1–24. arXiv:arXiv:1809.04644v1

[73] Christian Sandvig, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. 2014. Auditing Algorithms : Research Methods for Detecting Discrimination on Internet Platforms. *International Communication Association* (2014), 1–20. https://doi.org/10.1109/DEXA.2009.55

[74] Srinivas Shakkottai and Rayadurgam Srikant. 2008. Network optimization and control. *Foundations and Trends in Networking* 2, 3 (2008), 271–379.

[75] Ashudeep Singh and Thorsten Joachims. 2018. Fairness of Exposure in Rankings. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining - KDD '18*. ACM Press, New York, New York, USA, 2219–2228. https://doi.org/10.1145/3219819.3220088 arXiv:arXiv:1802.07281v2

[76] Manya Sleeper, Rebecca Balebako, Sauvik Das, Amber Lynn McConahy, Jason Wiese, and Lorrie Faith Cranor. 2013. The post that wasn't: exploring self-censorship on facebook. *Proceedings of the 2013 conference on Computer supported cooperative work* (2013), 793. https://doi.org/10.1145/2441776.2441865

[77] Stanford computational policy lab. 2018. Stanford computational policy lab. https://policylab.stanford.edu/.

[78] Eric Sun, Itamar Rosenn, Cameron A. Marlow, and Thomas M. Lento. 2009. Gesundheit ! Modeling Contagion through Facebook News Feed Mechanics of Facebook Page Diffusion. *Proceedings of the 3rd International ICWSM Conference* 2000 (2009), 146–153. https://doi.org/noDOI.URLcorrect

[79] John Tadrous and Atilla Eryilmaz. 2016. On optimal proactive caching for mobile networks with demand uncertainties. *IEEE/ACM Transactions on Networking* 24, 5 (2016), 2715–2727.

[80] Sarah Tan, Rich Caruana, Giles Hooker, and Yin Lou. 2017. Detecting Bias in Black-Box Models Using Transparent Model Distillation. 1 (2017). arXiv:1710.06169 http://arxiv.org/abs/1710.06169

[81] Michele Tortelli, Dario Rossi, and Emilio Leonardi. 2017. ModelGraft: Accurate, scalable, and flexible performance evaluation of general cache networks. In *Proceedings of the 28th International Teletraffic Congress, ITC 2016*, Vol. 1. 304–312.

[82] Zeynep Tufekci. 2015. How Facebook's Algorithm Suppresses Content Diversity (Modestly) and How the Newsfeed Rules Your Clicks. https://medium.com/message.

[83] Johan Ugander, Brian Karrer, Lars Backstrom, and Cameron Marlow. 2011. The Anatomy of the Facebook Social Graph. *CoRR* (2011). https://doi.org/10.1145/2532508.2532512 arXiv:1111.4503v1

[84] Sebastián Valenzuela, Namsu Park, and Kerk F Kee. 2009. Is there social capital in a social network site?: Facebook use and college students' life satisfaction, trust, and participation. *Journal of computer-mediated communication* 14, 4 (2009), 875–901.

[85] Hal R Varian. 1983. Non-parametric tests of consumer behaviour. *The review of economic studies* 50, 1 (1983), 99–110.

[86] Liang Wang, Gareth Tyson, Jussi Kangasharju, and Jon Crowcroft. 2017. Milking the Cache Cow With Fairness in Mind. *IEEE/ACM Transactions on Networking* 25, 5 (2017). https://doi.org/10.1109/TNET.2017.2707131

[87] Jeannette Wing. 2018. Data for Good: Scary AI and the Dangers of Big Data. https://www.youtube.com/watch?v=YL6zXHa5-RY.

[88] Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez Rodriguez, and Krishna P Gummadi. 2015. Learning fair classifiers. *arXiv preprint arXiv:1507.05259* (2015).

[89] Meike Zehlike, Francesco Bonchi, Carlos Castillo, Sara Hajian, Mohamed Megahed, and Ricardo Baeza-Yates. 2017. FA*IR: A Fair Top-k Ranking Algorithm. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management - CIKM '17*. ACM Press, New York, NY, USA, 1569–1578. https://doi.org/10.1145/3132847.3132938 arXiv:1706.06368

[90] Junzhe Zhang and Elias Bareinboim. 2017. Fairness in Decision-Making: The Causal Explanation Formula. *Purdue technical report R-30-L* (2017).

[91] Indrė Žliobaitė. 2017. Measuring discrimination in algorithmic decision making. *Data Mining and Knowledge Discovery* 31, 4 (2017), 1060–1089.

Appendix A. List of publishers used in the experiment

Table A.4 contains a list of publishers followed in the Italian experiment with their respective orientations. During the experiments, the page entitled *Fasciti uniti per L'italia* was shutdown by Facebook and it was replaced by *Lotta Studentesca*.

Appendix B. Publishers' visibilities

In what follows we complement results presented in Section 3.2. Whereas in Section 3.2 we showed how *occupancies* varied as a function of the rate at which publishers create posts, in this appendix we focus on *visibilities*. Figure B.7 shows the visibility for the top publishers, at the topmost News Feed position ($K = 1$), under the six considered bots. It indicates, for instance, that even at the undecided bot, visibilities do not vary monotonically with respect to publishers' post creation rates. Figure B.8 considers the case $K = 30$. As expected, the visibilities increase as K grows from 1 to 30. Nonetheless, when $K = 30$ we still find some top publishers that have almost negligible visibility at a number of bots. In particular, eight out of the thirty top publishers have negligible visibility at the undecided bot.

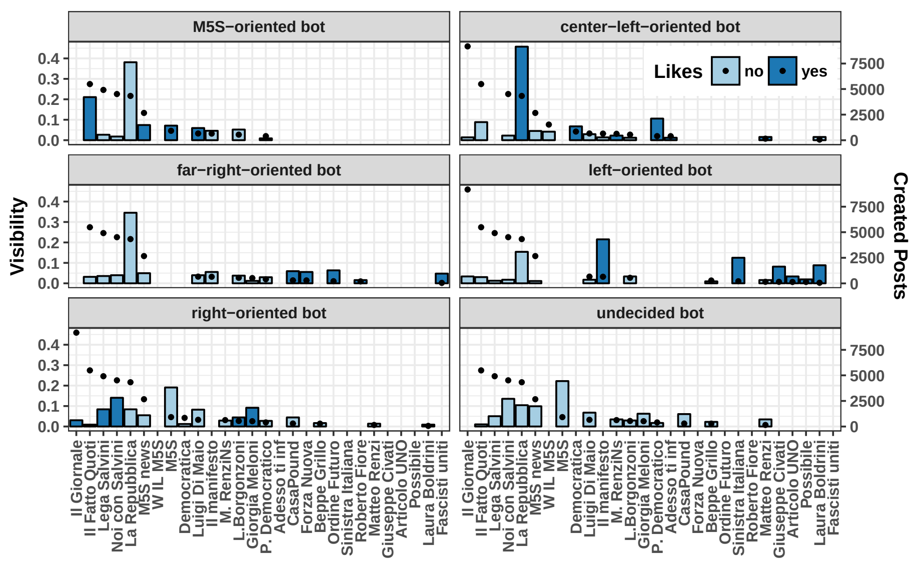


Figure B.7: Publishers' visibilities at the six bots (bars colored by preferences) and number of created posts (black dots), for $K = 1$.

Table A.4: Lists of the publishers followed in the Italian experiment with their respective labels.

Orientation	Page URL	Publisher Name
Right	NoiconSalviniUfficiale	Noi con Salvini
Right	ilpopulista.it	Il Populista
Right	ilGiornale	Il Giornale
Right	legasalvinipremier	Lega - Salvini Premier
Right	rivogliobologna	Lucia Borgonzoni
Right	giorgiameloni.paginaufficiale	Giorgia Meloni
Far-right	Fascisti-uniti-per-Litalia-411675765615435	Fascisti uniti per L'italia
Far-right	Lotta-Studentesca-257153365332	Lotta Studentesca
Far-right	OrdineFuturo	Ordine Futuro
Far-right	ilprimatonatsionale	Il Primato Nazionale
Far-right	ForzaNuovaPaginaUfficiale	Forza Nuova
Far-right	casapounditalia	CasaPound Italia
Far-right	RobertoFiorePaginaUfficiale	Roberto Fiore
Left	Articolo1Modempro	Articolo UNO
Left	sinistraitalianaSI	Sinistra Italiana
Left	ilmanifesto	il manifesto
Left	Possibile.it	Possibile
Left	giuseppecivati	Giuseppe Civati
Left	Laura-Boldrini-325228170920721	Laura Boldrini
Center-left	Adessotiinformo	Adesso ti informo
Center-left	matteorenzinews	Matteo Renzi News
Center-left	Repubblica	la Repubblica
Center-left	democratica	Democratica
Center-left	matteorenziufficiale	Matteo Renzi
Center-left	partitodemocratico.it	Partito Democratico
M5S	news.m5s	M5S news
M5S	WIlM5s	W IL M5S
M5S	ilFattoQuotidiano	Il Fatto Quotidiano
M5S	movimentocinquestelle	MoVimento 5 Stelle
M5S	LuigiDiMaio	Luigi Di Maio
M5S	beppegrillo.it	Beppe Grillo

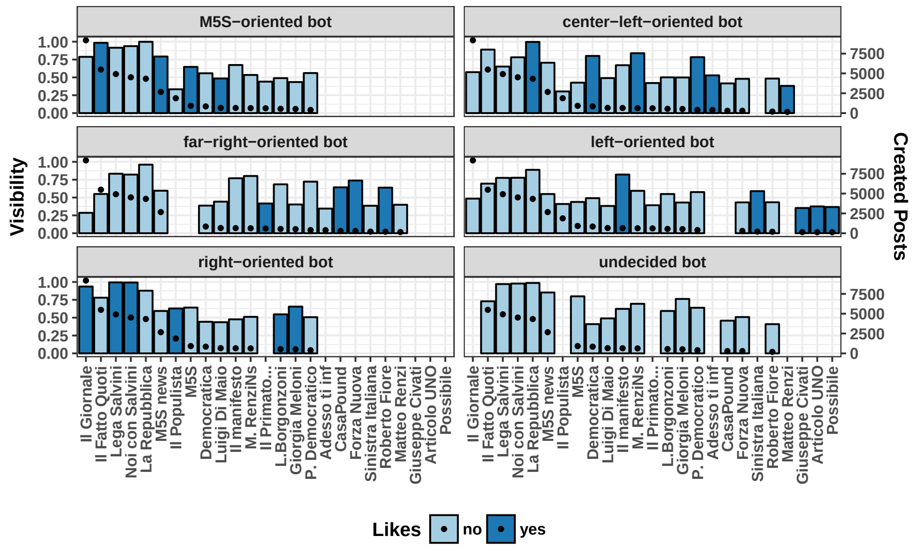


Figure B.8: Publishers visibilities at the six bots (bars colored by preferences) and number of created posts (black dots), for $K = 30$.

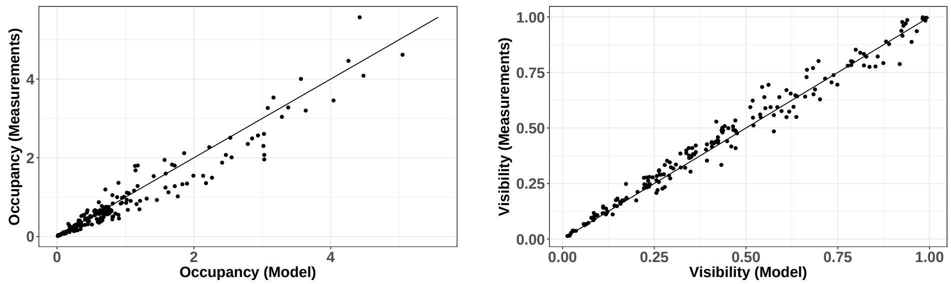
Appendix C. Model validation

Appendix C.1. Validation for $K = 30$

From Figures C.9(a) and C.9(b) to Figures C.10(a) and (b), each point corresponds to a user-publisher pair. We let $K = 30$ (results for $K = 1$ are presented in Section 5). In Figure C.9(a) (resp., Fig. C.9(b)), a point $(x = N_{ij}, y = \tilde{N}_{ij})$ (resp., $x = \pi_{ij}, y = \tilde{\pi}_{ij}$) indicates that, for the given pair, an occupancy N_{ij} (resp., visibility π_{ij}) estimated by the multi-class model using eq. (2) (resp., (4)) corresponds to a measured occupancy of $\tilde{N}_{ij}$ (resp., measured visibility of $\tilde{\pi}_{ij}$). Most of the points are close to the $\tilde{N}_{ij} = N_{ij}$ line, indicating the expressive power of the model. In Figure C.10(a) (resp., Fig. C.10(b)), a point $(x = N_{ij}, y = \tilde{N}_{ij})$ (resp., $x = \pi_{ij}, y = \tilde{\pi}_{ij}$) indicates that, for the given pair, an occupancy N_{ij} (resp., visibility π_{ij}) estimated by the two-class model using eq. (2) (resp., (4)) corresponds to a measured occupancy of $\tilde{N}_{ij}$ (resp., measured visibility of $\tilde{\pi}_{ij}$). The two-class model has two parameters, while the number of parameters in the multi-class model is equal to the number of publishers times the number of bots (180 in the experiment). For this reason, the accuracy of the former is significantly lower than the later.

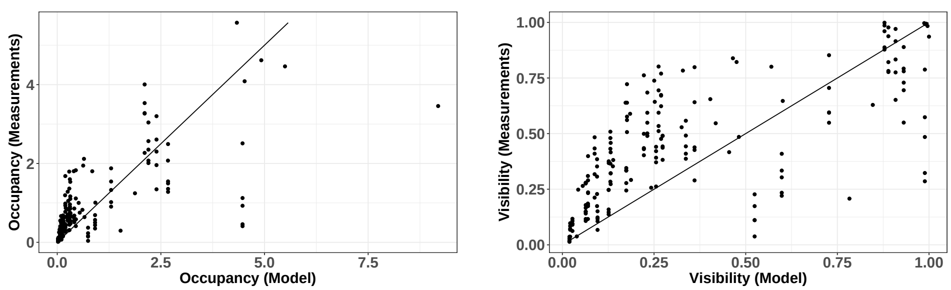
Appendix C.2. 2017 French presidential elections

Another experiment was conducted during the 2017 French presidential elections where four Facebook bots were created and monitored. The experiments started in April 28, 2017, and ended in May 08, 2017. Our profiles were kept



(a) (b)

Figure C.9: Multi-class model validation for the (*a*) occupancy and (*b*) visibility metrics, for $K = 30$.

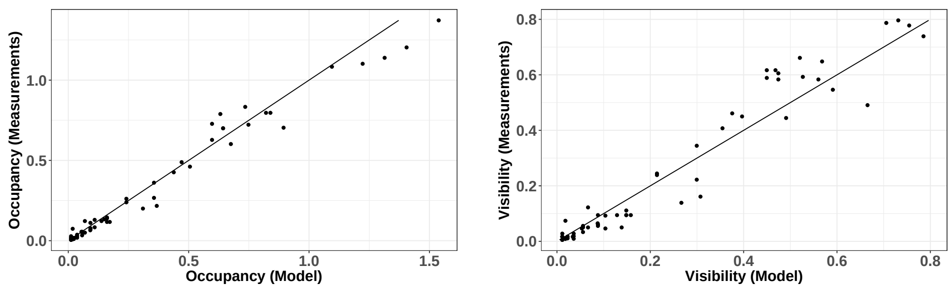


(a) (b)

Figure C.10: Two-class model validation for the (*a*) occupancy and (*b*) visibility metrics, for $K = 30$.

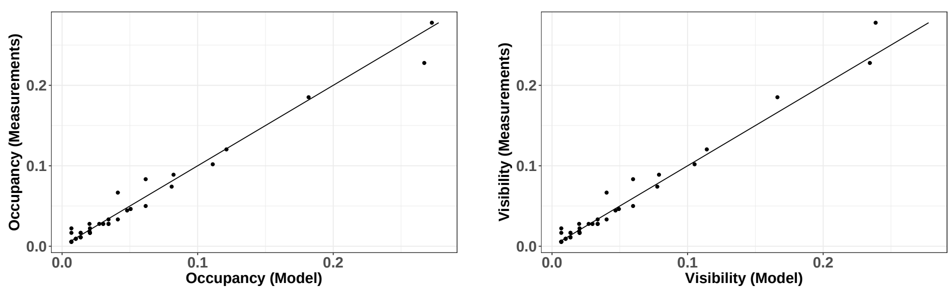
with no friends, and they all followed the same group of 13 pages in a addition to a number of random pages. We adopted the multi-class approach (Section 5.1) to parametrize and validate the model with such dataset. Figures C.11 and C.12 show our model validation for $K = 10$ and $K = 1$ (topmost position), respectively. The values predicted by the model are very close to the measured points indicating once again the expressive power of the model. The dataset corresponding to the French elections is publicly available.[2]

[2]https://github.com/EduardoHargreaves/Effect-of-the-OSN-on-the-elections



(a) (b)

Figure C.11: Model Validation for the (a) occupancy and (b) visibility metrics, for $K = 10$ using the 2017 French Elections Dataset.



(a) (b)

Figure C.12: Model Validation for the (a) occupancy and (b) visibility metrics, for the topmost position using the 2017 French Elections Dataset.

Appendix D. Potential delay and max-min fairness

Next, we report results on the potential delay, max-min and proportional fairness. In Figure D.13 we set uniform filtering as our baseline, whereas in Figures D.14 and D.15 we set the Facebook measurements as the baseline in equation (30). We let $w_i^{(1)} = 2$ and $w_i^{(0)} = 1$. Figure D.13 shows the occupancy for the top publishers, for $K = 1$, under the Facebook measurements, potential fairness, max-min fairness and proportional fairness. Note that the general trends of proportional fairness, potential delay fairness and max-min fairness are similar. Potential delay fairness tends to favor publishers that bots "like" more than the other fairness functions, while max-min tends to favor less. Facebook occupancies, in contrast, do not reflect any of the two considered fairness criteria. At the undecided bot, all fairness criteria yields the same occupancies and biases, reflecting the lack of preferences of this bot.

Figures D.14 and D.15 show the biases of the considered fairness criteria using the Facebook as baseline, for $K = 30$ and $K = 1$. In both cases, we note

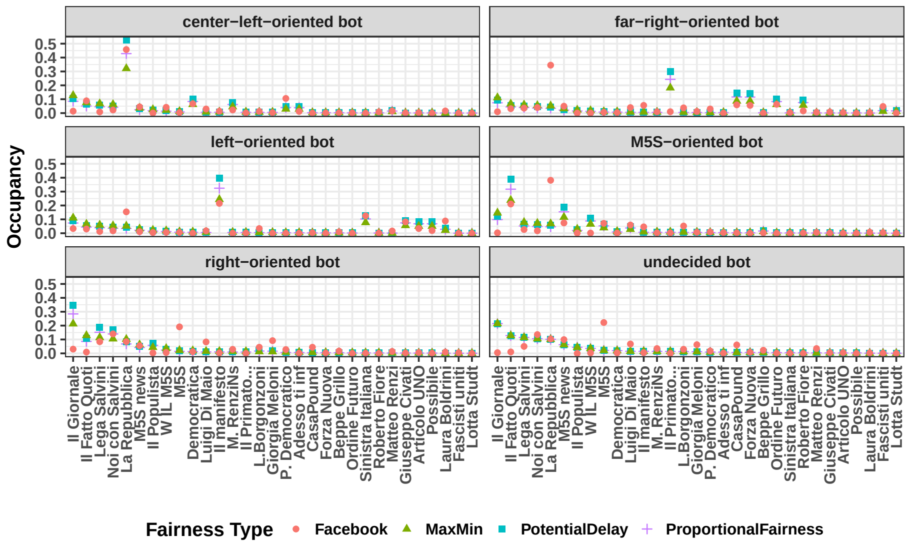


Figure D.13: Publishers' occupancies under the three types of fairness, in addition to measured Facebook occupancies, for $K = 1$, $w_i^{(1)} = 2$ and $w_i^{(0)} = 1$ for all users, at the six bots.

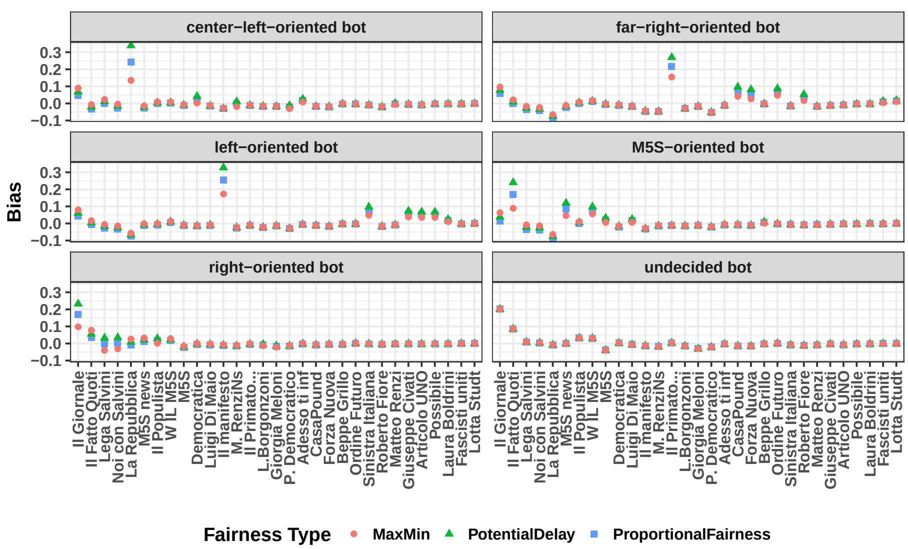


Figure D.14: Publishers' bias under the three types of fairness: Facebook occupancy set as baseline, for $K = 30$, $w_i^{(1)} = 2$ and $w_i^{(0)} = 1$ for all users, at the six bots. Potential delay fairness (green tringles) tends to favor "liked" publishers more than proportional fairness (blue squares).

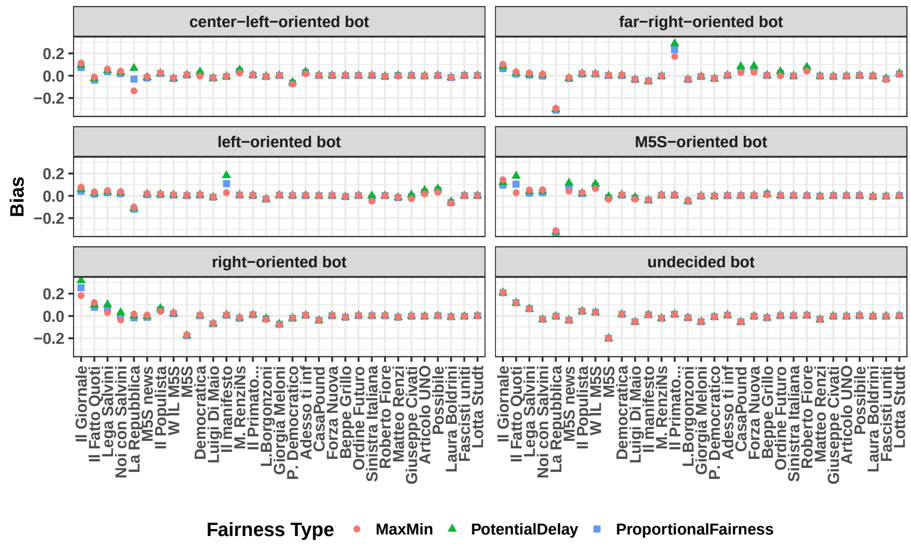


Figure D.15: Publishers' bias under the three types of fairness: Facebook occupancy set as baseline, for $K = 1$, $w_i^{(1)} = 2$ and $w_i^{(0)} = 1$ for all users, at the six bots. Potential delay fairness (green tringles) tends to favor "liked" publishers more than proportional fairness (blue squares).

that potential delay fairness (green triangles) tends to favor "liked" publishers more than proportional fairness (blue squares). When $K = 1$, Facebook tends to penalize the publishers that produced more posts at the undecided bot. Note also that there is a positive negative bias towards M5S posts at the undecided bot, meaning that Facebook allocated far more occupancy to M5S posts than the proposed methods would allocate.

Appendix E. Sensitivity analysis with respect of weights and News Feed size

In Section 6.5 we assumed $w_i^{(l)} = 1$ when considering the two-class model. Recall that $w_i^{(0)}$ (resp., $w_i^{(1)}$) correspond to publishers that a bot does not "like" (resp., "likes").

In Figure E.16, we keep $w_i^{(0)} = 1$ and vary $w_i^{(1)}$ from 1 to 10 to show the impact of the weights on the occupancies. We consider proportional fairness allocations, with $K = 30$. Figure E.16 shows that a ten-fold increase in $w_i^{(1)}$, from 1 to 10, may lead to an up to two-fold increase in the occupancies of publishers that bots "like". This is the case, for instance, with "Il Fatto Quotidiano", which was classified as a M5S source, and which significantly benefited from the increase of $w_i^{(1)}$ at the M5S-oriented bot.

Next, we consider the impact of K on our results. We observe that our utility optimization framework produces occupancies that are directly proportional to K (see, for example, (24)), and the corresponding biases are independent from K. For this reason $b_{ij}^{(\mathrm{PropF})}$ does not change between Figures E.17(a) and E.17(b). In contrast, the shape of the biases accounting for the Facebook measurements, $b_{ij}^{(\mathrm{Face})}$, are substantially different for $K = 1$ and $K = 30$, with stronger biases for $K = 1$ (in agreement with the empirical findings reported in Section 3.2).

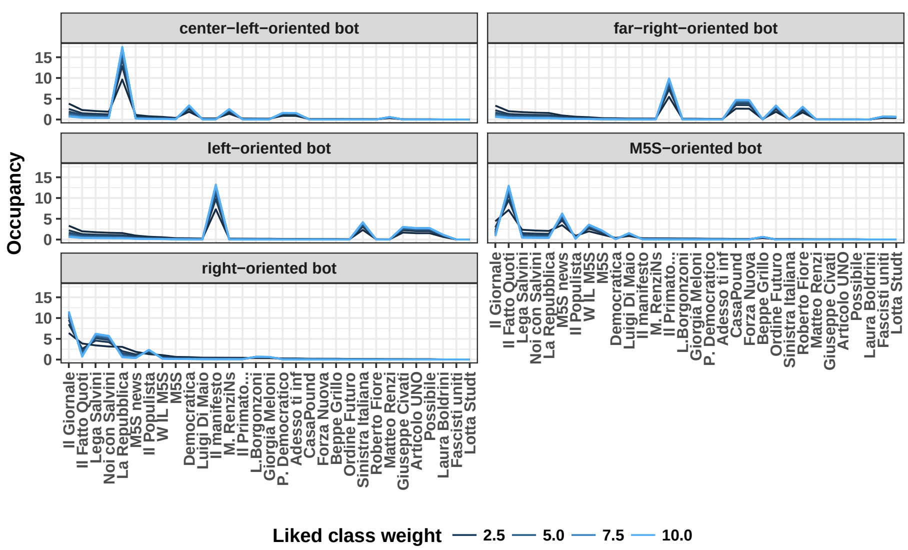
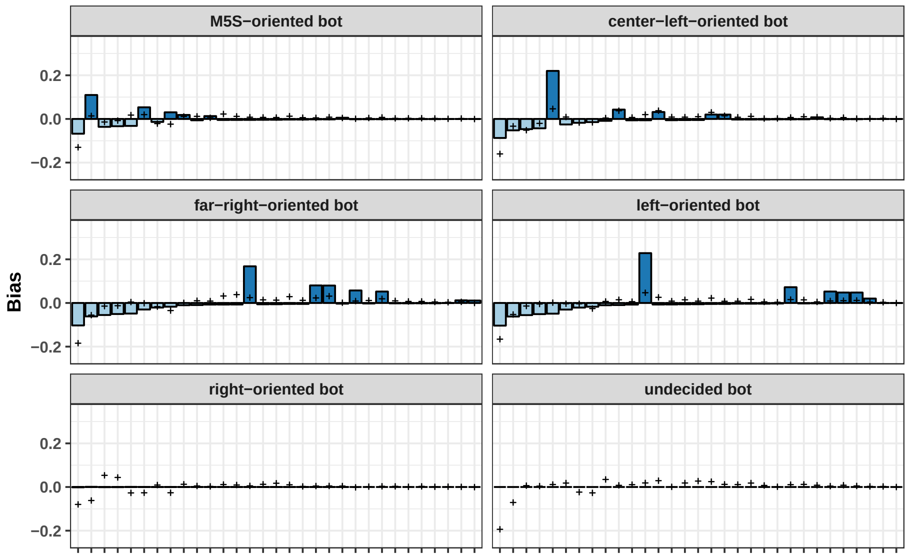
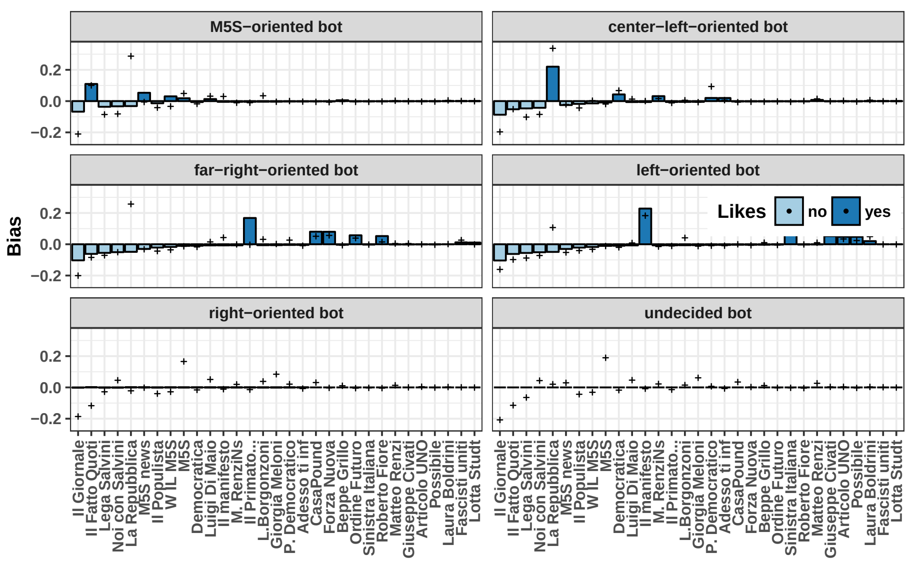


Figure E.16: Publishers' occupancies under proportional fairness for $K = 30$ at the six bots (bars colored by preferences) , $w_i^{(1)}$ ranging from 1 to 10 and $w_i^{(0)} = 1$.



(a) $K = 30$



(b) $K = 1$

Figure E.17: Proportional fairness bias (bars) and Facebook News Feed bias (crosses), for the six bots. Note that the bias experienced by the Facebook News Feed significantly differs between $K = 30$ and $K = 1$, and tends to be stronger at the topmost position.