

# **Modelando, Auditando e Prevendo a Visibilidade de Fontes no Facebook: Um Estudo de Caso nas Eleições Italianas**

## **Title: Modeling, auditing, and forecasting Facebook publisher visibilities: A case study in Italian elections**

**Eduardo Hargreaves<sup>1</sup>,  
Daniel Menasché<sup>1</sup>, Giovanni Neglia<sup>2</sup>, Claudio Agosti<sup>3</sup>**

<sup>1</sup>Dept. Ciência da Computação, UFRJ, Rio de Janeiro, Brasil

<sup>2</sup>Inria, Université Cote D'Azur, Sophia Antipolis, França

<sup>3</sup>Universidade de Amsterdã, Amsterdã, Holanda

eduardo@hargreaves.tech, sadoc@dcc.ufrj.br, giovanni.neglia@inria.fr

claudio.agosti@logioshermes.org

**Resumo.** *O Facebook tem um impacto significativo na vida de bilhões de usuários da Internet, todos os dias. Entretanto, os mecanismos usados pelo Facebook para filtrar mensagens apresentadas aos usuários não são de domínio público. Neste trabalho, propomos modelos, experimentos e análises de séries temporais para preencher esta lacuna. Como estudo de caso, realizamos experimentos reais com a News Feed do Facebook durante as eleições presidenciais italianas de 2018. Dentre as implicações dos nossos estudos, indicamos o potencial do modelo proposto para inferir diferentes métricas de visibilidade, a sua capacidade em auditar a News Feed do Facebook e a sua habilidade em prever a próxima ocupação da timeline dado o seu histórico de ocupação.*

**Abstract.** *Facebook has a significant impact on the lives of billions of Internet users, every day. However, the algorithms used by Facebook to filter messages presented to users are not in the public domain. In this work, we propose models, experiments and time series analysis to fill this gap. As a case study, we conducted experiments with the Facebook News Feed during the 2018 Italian presidential elections. Among the implications of our studies, we indicated the potential of the proposed model to accurately infer different visibility metrics, the ability to audit News Facebook feed and the power to forecast timeline occupancies given their previously observed values.*

## **1. Introdução**

O Facebook tem um impacto significativo na vida de milhões de usuários da Internet, todos os dias [Tsaparas 2017]. Como qualquer plataforma de compartilhamento de mídia, uma das principais funções do Facebook é organizar as informações disponíveis para melhorar a experiência do usuário. Para isso, o Facebook desenvolveu um *produto* de

filtragem, recomendação e personalização chamado *News Feed* que organiza e seleciona as publicações que aparecerão nas *timelines* dos usuários [Facebook 2017].

Entretanto, os mecanismos usados pelo Facebook para filtrar mensagens apresentadas aos usuários não são de domínio público, motivando pesquisas que envolvem desde estatísticas básicas até engenharia reversa de algoritmos. O escândalo da Cambridge Analytica [TheGuardian 2018], a possibilidade da influência do Facebook tanto nas eleições presidenciais americanas de 2016<sup>1</sup> quanto na campanha do *Brexit* [Cadwalladr 2019] ilustram a necessidade fundamental de garantir certo nível de fiscalização do Facebook pelos seus usuários e pela sociedade como um todo [Garfinkel et al. 2017, Diakopoulos 2016]. Essa fiscalização se torna ainda mais essencial porque os usuários preferem receber as notícias através de *gateways* como redes sociais do que diretamente dos portais de notícias [Newman et al. 2018].

Compreender um sistema complexo como o Facebook não é trivial. A visibilidade de uma certa fonte, por exemplo, depende da taxa de publicação de postagens por parte da fonte e dos interesses dos usuários. No artigo publicado em [TechCruch 2016] é dito que o Facebook utiliza aproximadamente 100.000 variáveis para escolher as publicações exibidas na suas *News Feeds*. Uma abordagem para avaliar o impacto de diferentes fatores na visibilidade consiste tanto na criação de modelos analíticos quanto no planejamento e realização de experimentos. O desafio é derivar métricas de interesse a partir de dados colhidos diretamente pelos usuários, sem depender exclusivamente das informações enviadas pelas plataformas de redes sociais online que estão sendo auditadas .

Este artigo é uma versão ampliada de [Hargreaves et al. 2018], onde tivemos por objetivo modelar, auditar e prever o comportamento da *News Feed* do Facebook e de *timelines* em geral. O modelo proposto é baseado em modelos de filas e de caches, para os quais existem uma vasta literatura [Martina et al. 2014, Harchol-Balter 2013]. Em particular, parametrizamos e validamos o modelo proposto usando dados do Facebook, indicando o poder expressivo do mesmo. Após a validação do modelo, análises contrafactuais são realizadas para auditar e identificar o viés da *News Feed*, utilizando-se uma *timeline* não filtrada como referência. Finalmente, propomos um método preciso de previsão da ocupação das próximas *News Feeds* dado o histórico das ocupações passadas.

**Contribuições** Dentre as principais contribuições deste trabalho, destacamos as seguintes.

**Modelo de visibilidade** Propomos um modelo para estimar a visibilidade das publicações em função das taxas de criação das fontes e do algoritmo de filtragem da rede social. Usando dados reais do Facebook, validamos os modelos propostos.

**Estudo de caso** Usando usuários virtuais, colhemos visões distintas sobre as postagens no Facebook nas eleições da Itália em 2018. Tais medições motivam a criação de modelos para capturar a essência da dinâmica das publicações em *timelines*.

**Quantificação do viés do Facebook** Usando as medições e os modelos propostos, avaliamos qual teria sido a visibilidade das diferentes fontes se a *timelines* não estivessem sob a influência dos filtros. O viés do algoritmo foi então mensurado utilizando as *timelines* não filtradas como referência.

<sup>1</sup><https://intelligence.house.gov/social-media-content/>

**Método de previsões** Usando o histórico da distribuição de conteúdo na *timeline* de um usuário, apresentamos um método para prever a ocupação dentro de uma determinada janela de tempo futura. Tal predição serve, por exemplo, para planejar acessos ao Facebook evitando acessos ao sistema com muita frequência minimizando o risco de perder informações desejáveis.

O restante deste artigo está organizado da seguinte forma. A Seção 2 apresenta uma fundamentação teórica do trabalho, trazendo uma visão geral sobre o modelo proposto. A Seção 3 apresenta o modelo analítico proposto. Na Seção 4 apresentamos um estudo de caso contendo resultados empíricos obtidos durante a campanha pré-eleitoral italiana. O modelo é validado na Seção 5 com os dados do estudo de caso. A Seção 6 quantifica o viés do algoritmo do Facebook usando o modelo proposto. A Seção 7 apresenta o método de predição proposto bem como uma avaliação da sua precisão. A Seção 8 cobre os trabalhos relacionados. Uma seção de discussão contendo as limitações do trabalho é feita na Seção 9. A Seção 10 conclui o trabalho.

## 2. Fundamentação Teórica

Antes de apresentar o modelo proposto para representar o comportamento de *timelines*, iremos apresentar uma fundamentação teórica contendo definições importantes para a compreensão do trabalho e as ideias que influenciaram o modelo proposto.

### 2.1. Terminologia

*Timelines* consistem em uma sequência de publicações que são exibidas aos usuários quando estes acessam suas aplicações de redes sociais. No Twitter as *timelines* são chamadas de *timelines* mesmo, no Facebook elas são chamadas de *News Feed* e no LinkedIn são chamadas de *LinkedIn Feed*. Neste trabalho, apesar de apresentarmos um modelo para *timelines* em geral, utilizamos a *News Feed* do Facebook como estudo de caso. Por este motivo, os termos *News Feed* e *timelines* são empregados de forma intercambiável ao longo do texto.

Dividimos os membros das redes sociais virtuais entre fontes e usuários. Membros que geram conteúdos são fontes, e membros que consomem tais conteúdos são usuários. Ao administrar sua *timeline*, um usuário ou grupo de usuários pode subscrever a fontes, e um subconjunto das publicações destas fontes será exibida na *timeline*. Note que numa rede em que a maior parte dos conteúdos é gerada pelos próprios usuários (ex., Facebook), o papel de cada usuário dinamicamente muda entre fonte e usuário.

### 2.2. Uma visão geral sobre a modelagem de *timelines*

A seguir, apresentamos intuitivamente as ideias que suportam o modelo analítico introduzido na seção seguinte.

#### 2.2.1. Analogias entre filas, *caches* e *timelines*

Em redes sociais virtuais, as publicações criadas pelas fontes fluem através das conexões entre os seus respectivos membros e alcançam as *timelines* dos usuários interessados. Desta forma, *timelines* são um canal de comunicação entre fontes e usuários.

Na sua forma mais simples, as publicações são organizadas por ordem de chegada, de forma que estas entram e saem das *timelines* de acordo com uma ordem cronológica. Os algoritmos de personalização somente filtram as publicações da fonte  $j$  na *timeline* do usuário  $i$ , transformando uma taxa de criação de conteúdo  $\Lambda_j$  em uma taxa de exibição de mensagens  $\lambda_{ij}$ .

O comportamento recém descrito é o comportamento de uma fila do tipo primeiro a entrar, primeiro a sair, ou, como é mais conhecida, como uma fila FIFO (*first-in, first-out*). Desta forma, acreditamos ser natural utilizar a teoria de filas para modelar as métricas de interesses das *timelines*.

Também argumentamos que existem inúmeras similaridades entre *timelines* e *caches* de conteúdos. Tanto *timelines* quanto *caches* são utilizadas para armazenar conteúdos de interesse a usuários. Ambas podem ser encaradas como filtros, tendo em vista que têm tamanho limitado. Além disso, ambas em geral armazenam conteúdos mais recentes, e removem conteúdos que não são mais de interesse dos usuários. E, principalmente, ambas aumentam a eficiência da obtenção de informação do que está guardado nelas.

Existe uma ampla literatura sobre *caches* e teoria de filas. Estabelecendo a relação entre filas, *caches* e *timelines*, podemos nos aproveitar desta literatura para estudar *timelines* que, por serem mais recentes, receberam menos atenção da comunidade científica em comparação com *caches* e filas, que são melhor entendidas. No caminho oposto, acreditamos que resultados obtidos através dos estudos de *timelines* podem ser resultar em novas políticas de *caching* ou em novas formas de distribuição de conteúdo.

### 2.2.2. *Timelines* são *caches*?

Neste trabalho, um de nossos objetivos é compreender como projetar *timelines* (tendo em vista suas similaridades com *caches*) para melhor prover conteúdo de interesse dos usuários. Cabe destacar, entretanto, que existem importantes diferenças entre *timelines* e *caches*. Usuários de *timelines* estão tipicamente interessados numa *classe* de itens relacionados a uma certa categoria ou usuários. Esta é a principal diferença entre os dois. Por exemplo, um usuário deseja seguir as últimas notícias de seu jornal favorito. Esta trata-se de uma consulta flexível, em comparação com uma busca mais específica, por exemplo, a busca por um determinado episódio de uma série de televisão. A Tabela 1 ilustra algumas das diferenças entre *caches* e *timelines*.

Pelas razões acima, embora *timelines* e *caches* tenham muitas similaridades, elas necessitam de políticas de inserção e remoção de conteúdos distintas. Algoritmos de *caching* clássicos prestam-se a servir requisições por itens específicos. Algoritmos para *timelines*, em contrapartida, precisam lidar com a distribuição de conteúdo baseada em classes, como por exemplo, um tópico específico ou uma fonte preferida. Por estes motivos, consideramos que *timelines* são *caches* orientados a fontes ou classes e divergindo dos *caches* tradicionais que são orientados à requisição.

### 2.3. Métricas de interesse

Dentre as métricas de interesse, destacamos as seguintes:

**Tabela 1. Comparação entre timelines e caches**

	Timelines	Caches
Evento de interesse	publicação de conteúdo	chegada de requisições
Decisões de inserção e remoção	tomadas após a publicação	tomadas depois de um <i>miss</i>
Requisições	para classe de conteúdos	para conteúdos específicos
Controle de ocupação	de itens por classe	de itens específicos

**Repercussão (probabilidade de acerto)** de uma fonte é a probabilidade de um usuário efetivamente ler (e possivelmente clicar) em uma publicação. Nesse trabalho, não distinguimos entre probabilidade de acerto e probabilidade de *click*. A probabilidade de acerto de uma fonte pode ser dada pela visibilidade ou pela ocupação das postagens desta fonte.

**Visibilidade** é a probabilidade de existir uma publicação de uma determinada fonte na *timeline*.

**Ocupação** é o número de publicações de uma determinada fonte na *timeline*.

Denotamos por  $\pi_{ij}$  a visibilidade da fonte  $j$  na *timeline* do usuário  $i$ . Denotamos por  $N_{ij}$  o número médio de publicações do *publisher*  $j$  na *timeline* de  $i$ . Neste trabalho, assumimos que a repercussão de uma fonte pode ser dada em função da visibilidade ou da ocupação. A repercussão igual a visibilidade condiz com usuários que eventualmente irão ler uma das publicações de cada uma das fontes presentes nas  $K$  primeiras posições de suas *timelines*. A repercussão igual a ocupação, por outro lado, captura o comportamento de usuários que irão influenciar-se mais por fontes que ocupam mais posições em suas *timelines*, ou seja, quanto mais posições ocupadas por uma fonte, maior a repercussão da mesma.

### 3. Um modelo analítico para *timelines*

A seguir, apresentamos o modelo analítico proposto, seguido por sua validação usando dados das eleições italianas de 2018.

#### 3.1. Descrição do modelo

Nesta seção, descrevemos um modelo analítico para capturar a dinâmica das publicações em uma *timeline*.

Seja  $\mathcal{I} := \{1, \dots, i, \dots, I\}$  o conjunto de usuários em estudo, cada usuário associado a uma *timeline*, e seja  $\mathcal{J} := \{1, \dots, j, \dots, J\}$  o conjunto de fontes. Seja  $i$  o usuário de interesse, cuja *timeline* desejamos modelar (a notação é sumarizada na Tabela 2).

A fonte  $j$  cria publicações segundo um processo Poisson com taxa  $\Lambda_j$ . Seja  $\lambda_{ij} \leq \Lambda_j$  a taxa efetiva com que a fonte  $j$  alimenta a *timeline*  $i$ . A taxa agregada de publicações chegando na *timeline* do usuário  $i$  é dada por

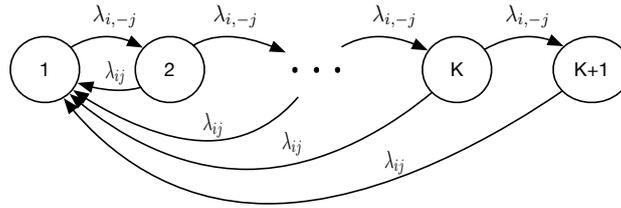
$$\lambda_i = \sum_{j=1}^J \lambda_{ij} \quad (1)$$

A taxa com que publicações de fontes, diferentes de  $j$ , chegam ao usuário  $i$  é

$$\lambda_{i,-j} = \lambda_i - \lambda_{ij} \quad (2)$$

**Tabela 2. Tabela de notação**

Variável	descrição
$K$	número de posições de interesse no topo da <i>timeline</i> ( <i>top K</i> )
$j$	$j$ -ésima fonte
$i$	$i$ -ésimo cliente (usuário)
$S_i$	número de fotografias tiradas no usuário $i$
$I_{ij}$	número de impressões da fonte $j$ no usuário $i$
$D_{ij}$	número de publicações distintas da fonte $j$ visualizadas pelo usuário $i$
$\Lambda_j$	taxa de criação de publicações da fonte $j$
$\lambda_{ij}$	taxa efetiva de chegada de publicações de $j$ na <i>timeline</i> de $i$
$\lambda_{i,-j}$	taxa efetiva de chegada de publicações de outras fontes (que não $j$ ) na <i>timeline</i> de $i$
$\lambda_i$	taxa total de chegada na <i>timeline</i> de $i$
$T_{ij}$	valor esperado para tempo de permanência das publicações da fonte $j$ na <i>timeline</i> de $i$
$\tilde{\pi}_{ij}$	visibilidade de $j$ medida na <i>timeline</i> de $i$
$\pi_{ij}$	visibilidade segundo modelo
$\hat{N}_{ij}$	ocupação média das publicações da fonte $j$ medida na <i>timeline</i> de $i$
$N_{ij}$	ocupação média das publicações de $j$ medida na <i>timeline</i> de $i$ segundo modelo proposto

**Figura 1. CTMC caracterizando a publicação da fonte  $j$  na posição mais no topo na *timeline* do usuário  $i$ .**

**Dinâmica FIFO** Assumimos que cada publicação entra no topo da *timeline*, na posição 1. Com taxa  $\lambda_i$ , as publicações tem sua posição incrementada em uma unidade. Estamos interessados nas primeiras  $K$  posições (slots) da *timeline*. De forma equivalente, assumimos que a *timeline* tem tamanho  $K$ . Uma publicação é removida da *timeline* quando ela passa da posição  $K$  para a posição fictícia  $K + 1$ .

**Cadeia de Markov** A seguir, consideramos uma cadeia de Markov para capturar a evolução das publicações da fonte  $j$ , na *timeline* do usuário  $i$ . O objetivo é calcular a visibilidade e a ocupação de  $j$  na *timeline* do usuário  $i$ . Por isso, não levamos em conta a posição de cada publicação de  $j$  na *timeline* de  $i$ , mas apenas a posição da publicação mais no topo. Enquanto houver um *post* de  $j$  na *timeline* de  $i$ , a fonte  $j$  estará visível.

A Figura 1 mostra uma cadeia de Markov tempo contínuo (CMTTC) cuja variável de estado  $X$  representa a posição mais no topo ocupada por uma publicação da fonte  $j$ . Seja  $\tilde{\pi}_{ij}(x) = P(X = x)$ . Com taxa  $\lambda_{ij}$ , uma publicação de  $j$  chega no topo *timeline*. Com taxa  $\lambda_{i,-j}$ , uma publicação de outra fonte faz com que as publicações de  $j$  sejam movidas para a posição  $x + 1$ . A fonte  $j$  estará fora da *timeline* se a sua publicação mais no topo encontrar-se na posição fictícia  $K + 1$ . Logo, a visibilidade  $\pi_{ij}$  é dada por

$$\pi_{ij} = 1 - \tilde{\pi}_{ij}(K + 1). \quad (3)$$

**Proposição 1.** *Em uma FIFO timeline, a visibilidade da fonte  $j$  na timeline do usuário  $i$  é*

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

*Demonstração.* O resultado segue imediatamente a partir das equações de balanço do sistema, que podem ser derivadas diretamente a partir da Figura 1.  $\square$

**Proposição 2.** *Em uma FIFO timeline, a ocupação média da fonte  $j$  na timeline do usuário  $i$  é*

$$N_{ij} = \frac{\lambda_{ij}K}{\lambda_i}. \quad (4)$$

*Demonstração.* Quando a  $m$ -ésima publicação entra na *timeline* ela fica visível até a chegada da  $(m + K + 1)$ -ésima publicação. Já que o tempo médio entre chegadas é  $1/\lambda_i$ , o tempo médio de residência da  $m$ -ésima publicação é  $K/\lambda_i$ . Em estado estacionário, seja  $T_{ij}$  o tempo médio que publicações de  $j$  ficam na *timeline* de  $i$ ,  $T_{ij} = K/\lambda_i$ . Pela lei de Little,  $N_{ij} = \lambda_{ij}T_{ij} = \frac{\lambda_{ij}K}{\lambda_i}$ .  $\square$

**Resumo modelo FIFO** O modelo FIFO é um dos modelos analiticamente tratáveis mais simples que se possa conceber para uma *timeline*. Nesta seção, apresentamos o modelo FIFO e indicamos como ele pode ser usado para derivar a visibilidade e ocupação média de cada fonte em uma *timeline*. Em trabalhos futuros, pretendemos considerar modelos alternativos, mais flexíveis, como aqueles baseados em *caches* do tipo TTL.

## 4. Um estudo de caso sobre as eleições italianas

### 4.1. Coleta dos dados

As eleições italianas foram no dia 04 de março de 2018 e o experimento com o Facebook ocorreu entre os dias 10 de janeiro de 2018 e 06 de março de 2018 de forma que procuramos analisar o período antecedente às eleições e os dois dias seguintes para capturarmos as reações ao resultado das eleições.

Foram selecionadas 30 fontes italianas que posteriormente foram classificadas de acordo com uma das cinco orientações políticas: esquerda, centro-esquerda, direita, ultra-direita e movimento 5 estrelas. Também foram criados 6 usuários virtuais no Facebook que seguem todo o conjunto composto pelas 30 fontes escolhidas. Entretanto, os usuários foram polarizados de forma que cada usuário curtiava publicações de apenas uma orientação política. O sexto usuário foi caracterizado como indeciso uma vez que não curtiu publicações de numa página. A cada hora, colhemos as publicações apresentadas na *timeline* de cada usuários. Tais fotografias constituem nossa base de dados.

As “fotografias” das *timelines* foram “tiradas” através de uma extensão dos navegadores Chrome e Firefox chamada *facebook.tracking.exposed*<sup>2</sup>. Essa extensão coleta os dados públicos, e retorna a data de criação da publicação, o momento da visualização, o

<sup>2</sup>Facebook tracking exposed: <https://facebook.tracking.exposed/>

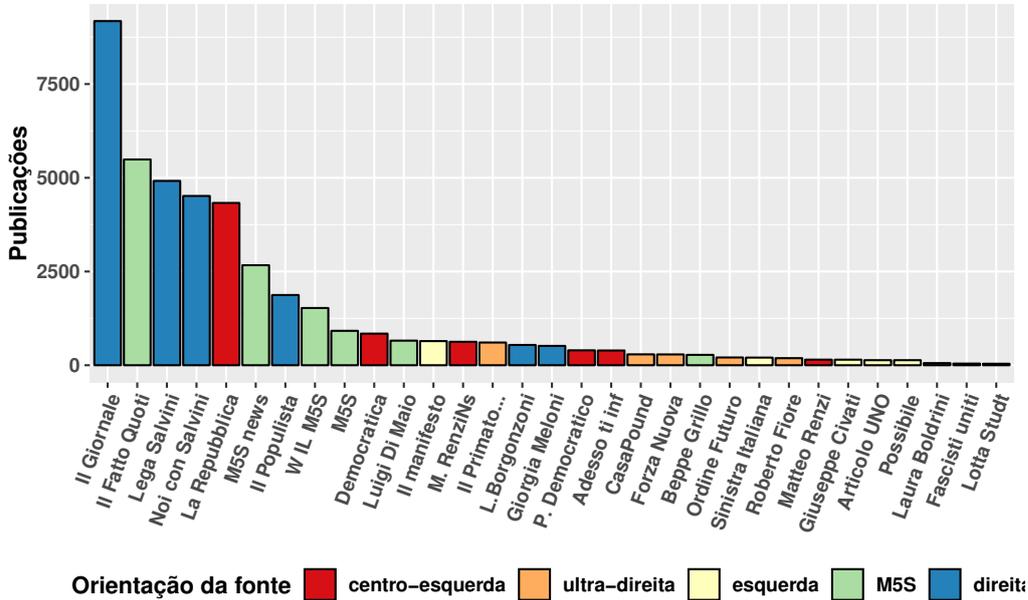


Figura 2. Total de publicações por fonte

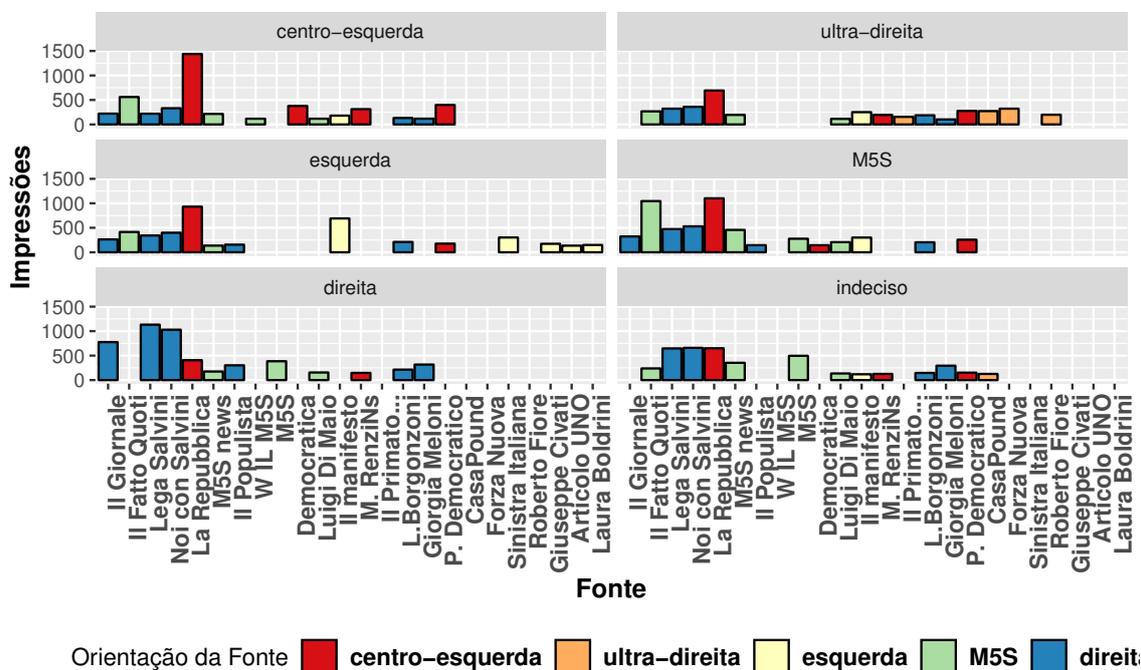
usuário que visualizou, a fonte, o conteúdo, a quantidade de reações, o número de compartilhamentos, e a ordem de aparição da publicação. Em paralelo, a API do Facebook foi utilizada para a obtenção de todas as publicações das fontes selecionadas. Chamamos de  $S_i$  o número de fotografias tiradas no  $i$ -ésimo usuário. Nos nossos experimentos, os usuários foram indexados de 1 a 6, denotando as orientações de centro-esquerda, ultra-direita, esquerda, M5S, direita e a indecisa. Os valores de  $S_i$  obtidos foram: 577, 504, 623, 674, 655, 576, com  $i = 1, \dots, 6$ . A diferença entre o número de fotografias advém de falhas de medição, e.g., máquinas travarem, falta de luz e queda de conexão. Embora existam tais falhas, acreditamos que estatisticamente elas não afetam os resultados, tendo em vista o grande volume de fotografias colhidas.

Cabe destacar que a API do Facebook desde 2015 não oferece os dados acima [Facebook 2018b]. Mesmo quando a API era aberta, o Facebook dizia que a informação fornecida pela API não era fidedigna [Facebook 2018a]. Utilizamos esta API somente para colhermos informações sobre o número de publicações por fonte.

#### 4.2. Achados empíricos

A seguir, apresentamos uma visão geral dos dados colhidos. A Figura 2(a) ilustra o número de publicações por fonte. Esta informação foi colhida diretamente da API do Facebook. Algumas poucas fontes geraram milhares de publicações durante o período considerado, enquanto que a maioria gerou dezenas de publicações.

Chamamos cada publicação visualizada de impressão, e chamamos de  $I_{ij}$  o número total de impressões da fonte  $j$  no usuário  $i$ . Se uma publicação é visualizada duas vezes, ela conta com duas impressões. Esta informação foi colhida a partir de nosso aplicativo (Facebook Tracking Exposed). A Figura 3 mostra o total de impressões  $I_{ij}$  por usuário. Em nenhum usuário a distribuição das publicações vistas foi semelhante a distribuição das publicações criadas. Em todos os usuários, as fontes mais vistas refle-



**Figura 3. Total de impressões no topo da timeline por fonte em cada usuário (Apenas ocupações superiores a 0.2 estão representadas)**

tiram a polarização dos mesmos. Por exemplo, o usuário que curtiu fontes de esquerda (no topo a direita da Figura 3) visualizou mais postagens de fontes de esquerda do que os demais. Cabe lembrar que todos os usuários seguiram todas as fontes (e apenas distinguiram-se pelas curtidas) e que o viés fica claro independente da taxa com quem as fontes geraram conteúdos (ilustrada na Figura 2).

A Figura 4(a) mostra o número de impressões em cada um dos seis usuários, agrupadas por orientação política, e a Figura 4(b) mostra o número total de publicações por orientação política. As Figuras 4(a) e 4(b) corroboram as observações anteriores: o viés das publicações reflete o viés dos usuários.

Cabe destacar que a grande diferença entre a ordem de aparição das fontes nas Figuras 2 e 3 é fruto da filtragem realizada pelo Facebook. Um dos objetivos do presente trabalho é propor um modelo analítico que nos permita compreender os efeitos de tal filtragem nas métricas de visibilidade e ocupação das fontes nas *timelines*.

Embora as observações apresentadas acima em parte sejam esperadas, cabe destacar que analisamos também usuários neutros. Para usuários neutros, que não curtem nenhuma fonte, seria de esperar que a presença de publicações fosse semelhante a Figura 2, ou então que existisse uma uniformidade entre as fontes. A Figura 3 mostra que esse não é o caso. Podemos constatar o alto número de impressões da fonte M5S no eleitor indeciso. É importante observar que o partido M5S foi o partido que recebeu mais votos nas eleições italianas. O modelo analítico apresentado a seguir nos permite realizar estudos contrafatuais, para averiguar o que ocorre com a visibilidade das fontes caso diferentes filtros sejam adotados.

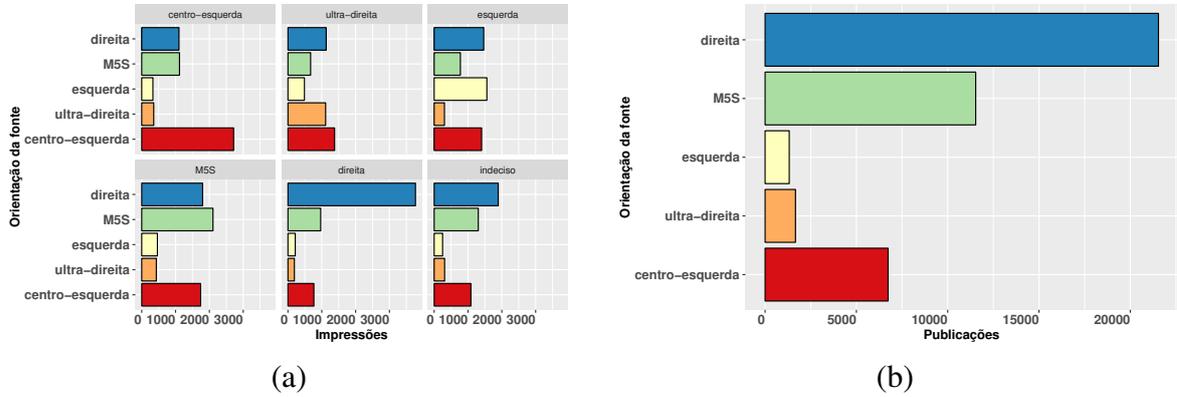


Figura 4. (a) Total de impressões no topo da timeline (por orientação política), (b) Total de publicações por orientação política

## 5. Validação

A seguir, apresentamos a validação do modelo proposto. Pretendemos verificar se o modelo tem capacidade expressiva para capturar os dados colhidos em ambiente real (eleições da Itália), e se ele é de fácil parametrização (requerendo apenas a taxa efetiva de publicações por fonte por usuário).

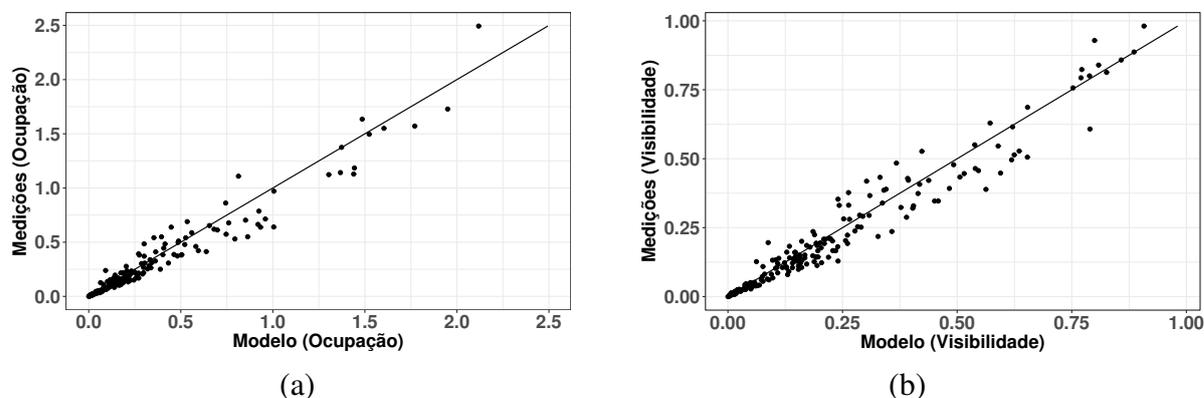
Chamamos de  $D_{ij}$  o número de publicações distantes da fonte  $j$  vistas pelo usuário  $i$ . Cabe ressaltar que esse número é menor ou igual a  $I_{ij}$ . A taxa  $\lambda_{ij}$  é dada pela razão  $D_{ij}/S_j$ . Cada taxa obtida  $\lambda_{ij}$  é substituída em (1) e em (2) para a obtenção da visibilidade, através da Proposição 1, e da ocupação, através da Proposição 4.

Cada ponto nas Figura 5 corresponde a um usuário e uma fonte. Um ponto  $(x, y)$  indica que, para o usuário e a fonte em questão, o modelo prediz uma ocupação  $x$  (respectivamente, visibilidade), e empiricamente observamos ocupação  $y$  (resp., visibilidade). Os erros advém do fato, por exemplo, de nosso modelo assumir que as publicações nunca são reordenadas. Ainda assim, o fato de a maioria dos pontos estar próxima à reta  $x = y$  ilustra o poder preditivo do modelo. O erro médio quadrático da ocupação foi igual a 0.0971 e o da visibilidade foi igual a 0.0527. É importante observar que não é possível comparar qual métrica o modelo represento melhor através do erro médio quadrático porque a visibilidade é uma probabilidade, logo varia entre zero e um, e a ocupação é um valor que varia entre 0 e o tamanho da *timeline*, que no exemplo é 10.

## 6. Quantificando o viés do Facebook

A seguir, combinamos o modelo proposto com os dados colhidos para realizar análise contrafactual de ocupações. Para tal, calculamos o viés entre as ocupações medidas empiricamente e as ocupações previstas pelo modelo, usando a quantidade de publicações criadas por cada fonte (API do Facebook). O objetivo é comparar a ocupação após a filtragem do Facebook (ocupação medida empiricamente) com aquela que nosso modelo prediz como sendo a ocupação condizente com um sistema sem filtros (ocupação analítica obtida com o modelo proposto).

O viés é definido como a diferença  $\hat{N}_{ij} - N_j$ , onde  $\hat{N}_{ij}$  é a ocupação média empírica da fonte  $j$  na *timeline*  $i$ ,  $N_j$  é dado por (4) e  $\hat{N}_{ij}$  é dado por  $I_{ij}/S_i$ . Observe



**Figura 5. Validação do modelo: (a) Ocupação, (b) Visibilidade**

que o índice  $i$  referente aos usuários é suprimido na expressão de  $N_j$  pois na análise de ocupação ignorando filtros assumimos que a ocupação é igual para todos os usuários.

As Figuras 6 e 7 mostram que, de um modo geral, o viés positivo reflete a orientação dos usuários. No entanto, a fonte que mais produziu publicações, Il Giornale, foi penalizada em todos os usuários, e a segunda fonte que mais produziu, o Il Fatto Quotidiano só não foi penalizado no usuário que curtiu a sua página. O usuário neutro sofreu viés tanto positivo quanto negativo. A fonte M5S teve um forte viés positivo no usuário neutro. No entanto, o forte viés negativo do Il Fatto Quotidiano anulou esse efeito de forma que o viés total do M5S no usuário neutro foi negativo.

## 7. Análise de séries temporais

A seguir, estudamos aspectos temporais associados às *timelines*. Até esse ponto, nosso enfoque foi na análise de métricas estacionárias. Nessa seção, focamos nos aspectos transientes das publicações exibidas nas 10 primeiras posições da News Feed dos 6 usuários participantes do experimento.

A Figura 8 mostra a evolução da ocupação das fontes ao longo do tempo em cada um dos seis usuários. Nessa figura, as fontes foram agrupadas de acordo com a sua respectiva orientação política para facilitar a visualização. Os valores da ocupação foram suavizados pela função *geom\_smooth* através do método *gam* (*Generalized Additive Models With Integrated Smoothness Estimation*) [Wickham, Hadley; Grolemund 2017]. Essa figura além de confirmar as observações feitas na Seção 4, mostra que as fontes de centro-esquerda tiveram um aumento da ocupação em todos os usuários a partir do início de março, e as fontes de esquerda apresentaram um pico de ocupação no usuário de esquerda no início de fevereiro.

Para ilustrar uma série temporal da ocupação de uma fonte em uma determinada News Feed, a Figura 9 apresenta a ocupação do Movimento 5 Estrelas (M5S), partido que recebeu mais votos na eleição italiana de 2018, no usuário indeciso. A partir do início de março, alguns *snapshots* chegaram a ter 50% de sua posições ocupadas com publicações do Movimento 5 Estrelas. No entanto, de uma forma geral, houve pouca oscilação no valor da ocupação média.

A Figura 10 mostra a razão entre o número de publicações das orientações

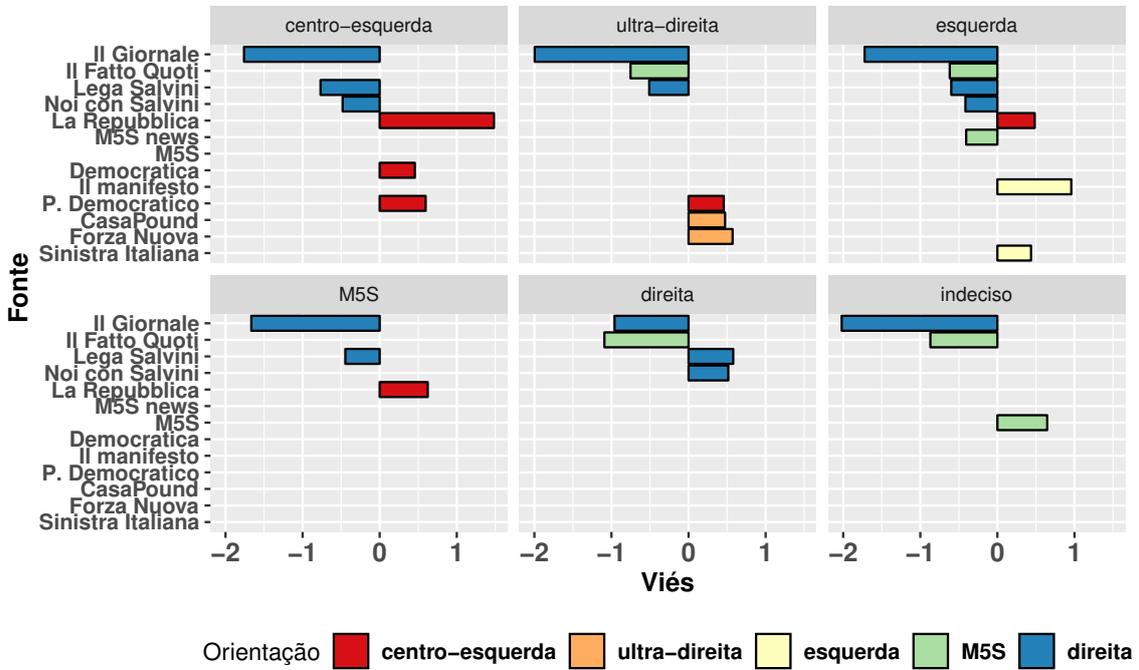


Figura 6. Viés por usuário e por fonte (estão representados apenas os casos em que o valor absoluto do viés foi maior do que 0.4)

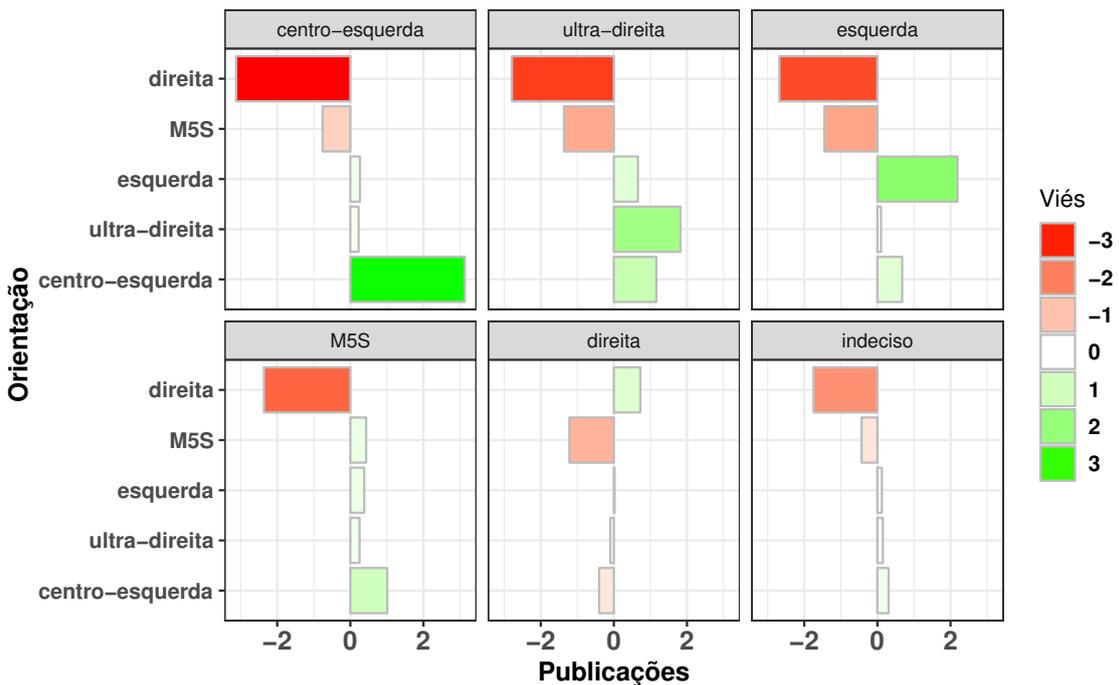
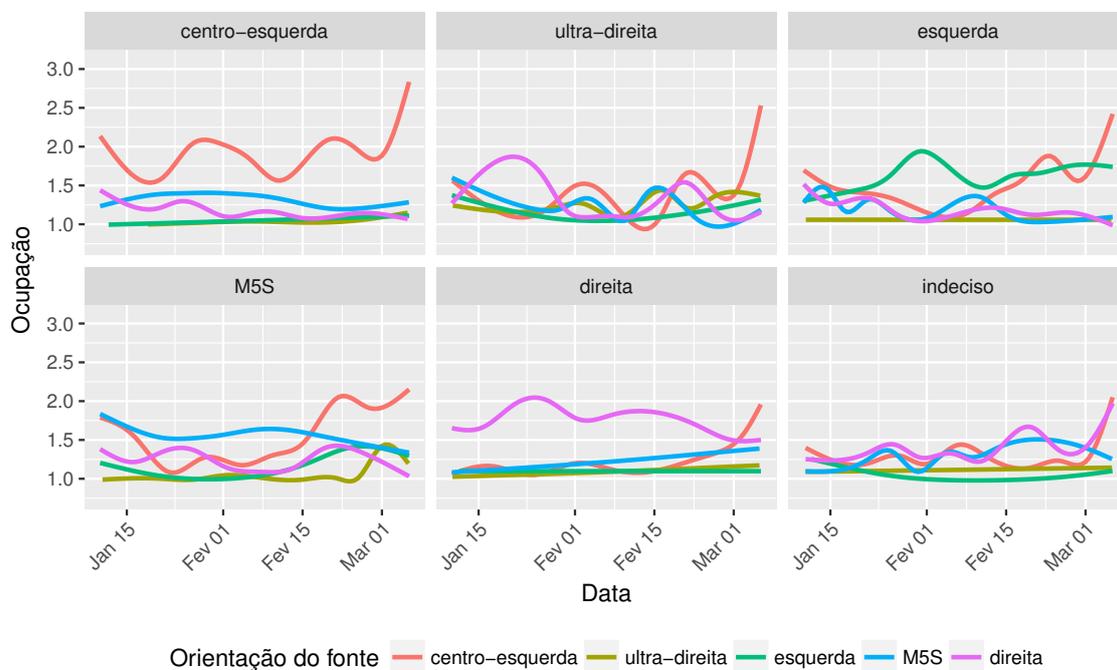
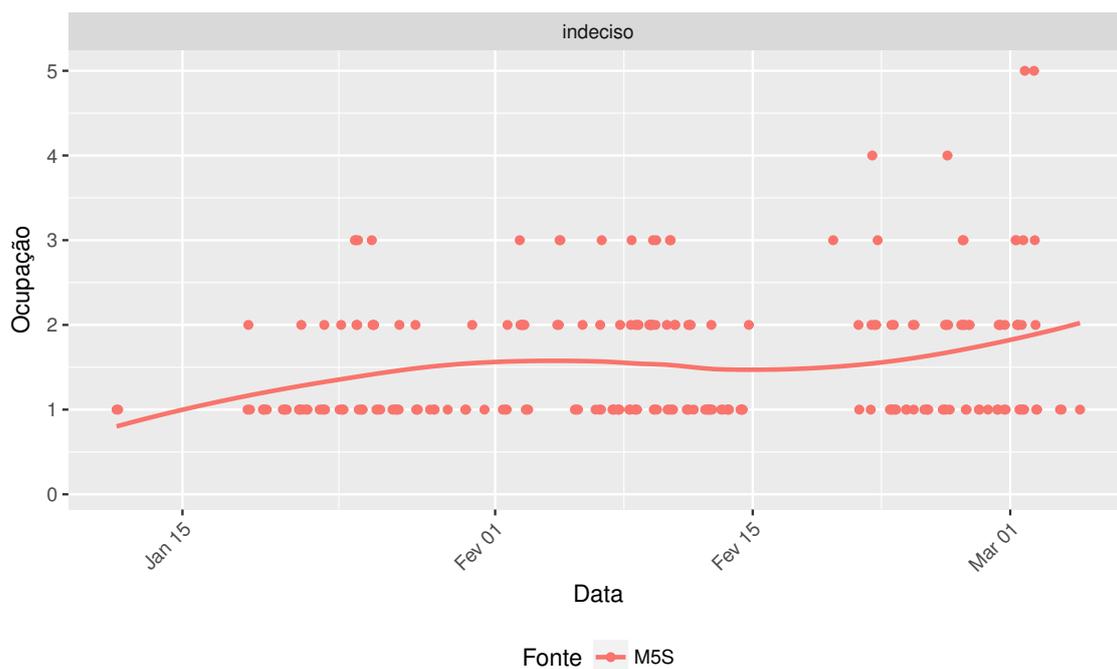


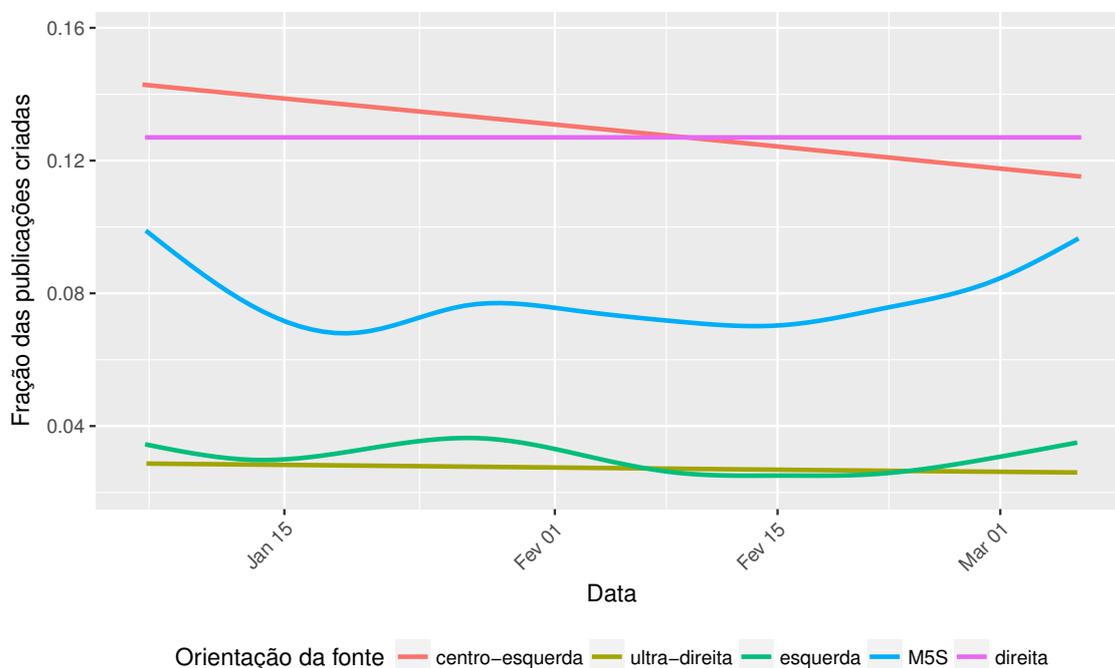
Figura 7. Viés por usuário e por orientação política



**Figura 8. Série Temporal da ocupação das orientações políticas nos 6 usuários**



**Figura 9. Série temporal da ocupação do movimento 5 estrelas (M5S) no usuário indeciso**



**Figura 10. Série temporal das publicações das fontes selecionadas**

políticas publicadas em um intervalo de tempo e o total de publicações realizadas neste mesmo período. Os valores destas frações também foram suavizados pela função *geom\_smooth* através do método *gam*. As fontes de centro-esquerda e de direita foram as que mais publicaram conteúdo no período analisado, reforçando o indicado na Figura 2. É importante ressaltar que a quantidade relativa de publicações criadas pelas fontes de centro-esquerda diminuiu no final do período analisado enquanto a ocupação apresenta uma súbita alta em todos os usuários neste período, conforme mostra a Figura 8.

Em particular, a análise dessa seção é útil tanto para usuários quanto para fontes. Do ponto de vista dos usuários, é útil poder prever, por exemplo, a distribuição de conteúdo em sua *timeline* dentro de uma determinada janela de tempo futura. Tal previsão serve, por exemplo, para planejar acessos ao Facebook. Tal planejamento, por sua vez, é útil para evitar acessar o sistema com muita frequência (o que é custoso, em termos de atenção) ou muito infrequentemente (perdendo-se possivelmente informações valiosas que deixem de estar no topo da *timeline*).

Do ponto de vista das fontes, a possibilidade de prever a distribuição de conteúdo é útil para fins estratégicos. Considere, por exemplo, um fonte que decida escolher, estrategicamente, quando postar conteúdo de tal forma que suas publicações admitam o mínimo de competição com publicações de um fonte que produz determinado tipo de conteúdo.

A Figura 11 apresenta a previsão da ocupação da fonte *La Repubblica* no usuário de centro-esquerda. A previsão é feita no *snapshot*  $t + 1$  usando todos os valores de ocupações desde o primeiro *snapshot* até o *snapshot*  $t$ . Usando-se esta previsão, um usuário pode, por exemplo, desistir de visualizar a sua News Feed porque a previsão indica que terão poucas publicações de uma determinada fonte na sua próxima visualização. As previsões foram feitas utilizando-se o modelo ARIMA cujos parâmetros

ótimos foram calculados com a função *auto.arima()* do pacote de previsões *Forecast* [Hyndman and Khandakar 2008] do R. A primeira previsão foi realizada após colhidos os primeiros 60 *snapshots*. Na Figura 11, os pontos vermelhos representam as medições, os azuis as previsões e os verdes, os erros. As linhas contínuas representam as suavizações destes pontos através do método *gam* da *geom\_smooth*. Seja  $\hat{N}_{ij}(t + 1)$  a ocupação prevista no *snapshot*  $t + 1$  e  $N_{ij}(t + 1)$  a ocupação medida neste mesmo *snapshot*. O erro  $\epsilon(t + 1)$  no *snapshot*  $t + 1$  é calculado pelo valor absoluto da diferença entre a medição  $N_{ij}(t + 1)$  e a predição  $\hat{N}_{ij}(t + 1)$  conforme a equação (5).

$$\epsilon(t + 1) = |\hat{N}_{ij}(t + 1) - N_{ij}(t + 1)|. \quad (5)$$

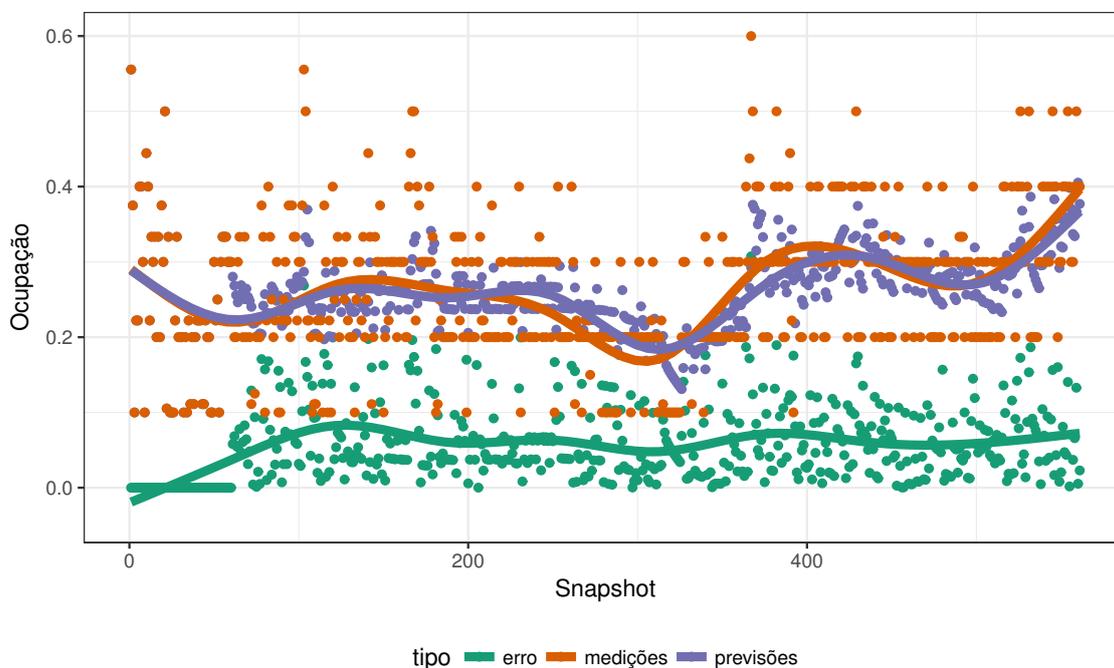
A Figura 11 mostra que as previsões acompanharam o aumento da ocupação da fonte de centro-esquerda na parte final do período do experimento. Consequentemente, o erro é aproximadamente constante durante todo o período e independentemente do tamanho do conjunto de treinamento.

Cabe também recordar que nosso modelo foi parametrizado usando informações sobre impressões. Entretanto, coletar impressões é custoso. É preciso desenvolver e manter atualizadas ferramentas de *web scrapping* além do esforço para gerenciar os usuários envolvidos nos experimentos. Dessa forma, idealmente é interessante ter um mecanismo para prever impressões a partir do padrão de publicações dos fontes. Nesta seção apresentamos primeiros passos nesse sentido. A partir das séries temporais de publicações dos fontes, buscamos correlações entre elas e as séries de impressões. Assim, em trabalhos futuros vislumbramos que seja possível caracterizar (parcialmente) o comportamento da *timeline* diretamente a partir de informações publicamente disponíveis sobre os fontes (e.g., padrão de criação de publicações) e usuários (e.g., padrão de *likes*).

## 8. Trabalhos relacionados

O livro de [O’Neil 2016] mostrou diversas situações nas quais algoritmos podem reforçar preconceitos e tomar decisões que podem influenciar a sociedade como um todo. No contexto de mídias sociais e política, em [Epstein and Robertson 2015] foi demonstrado que manipulações em mecanismos de buscas são capazes de influenciar eleitores indecisos. Robôs fazendo propaganda política tentando influenciar eleições foram estudados no mundo todo em [Woolley and Howard 2017] e particularmente no Brasil em [Arnaudo 2017]. Um estudo realizado por [Eslami et al. 2015] mostrou que mais de 60% das pessoas não tinham conhecimento da existência dos filtros de personalização do Facebook. Em [Bakshy et al. 2015] foi identificado que as escolhas individuais são o maior fator de influência sobre o nível de exposições a posições políticas opostas no Facebook. O nosso trabalho, por outro lado, mostrou que o algoritmo reforça as preferências dos usuários. Esse reforço de preferências pode ajudar a criar as *filter bubbles* identificadas em [Pariser 2011].

Formas de auditar o viés de algoritmos foram investigados em [Diakopoulos 2013, Sandvig et al. 2014]. Em [Epstein and Robertson 2015] foi mostrado que a manipulação do ranking dos mecanismos de busca pode influenciar o voto de cidadãos indecisos. Esse viés foi ainda quantificado em [Kulshrestha et al. 2017]. Com a ajuda do nosso modelo, pudemos quantificar o viés no *News Feed* do Facebook.



**Figura 11. Previsão da ocupação no *snapshot*  $t+1$ . Os pontos roxos representam as previsões, os vermelhos os valores medidos e os verdes representam os erros.**

O comportamento dos usuários em busca de visibilidade foi estudado em [Eslami et al. 2016, Bucher 2012, Sleeper et al. 2013]. Tais estudos são baseados principalmente em pequenos experimentos. As exceções são os artigos [Bakshy et al. 2012, Bakshy et al. 2015, Kramer et al. 2014, Sun et al. 2009, Bond et al. 2012] que trabalharam com extensos *datasets* fornecido pelo Facebook por meio de acordos restritivos de não divulgação. Portanto, grandes *datasets* para avaliar as visibilidades das fontes do Facebook não estão disponíveis publicamente. Nosso trabalho visa contribuir preenchendo essa lacuna através da utilização da extensão do navegador para a coleta das publicações sem a participação do sistema que está sendo auditado.

A plataforma de anúncios do Facebook vem recebendo bastante atenção da literatura recentemente. Formas de discriminação de usuários através de anúncios foram estudadas em [Ali et al. 2019], uma análise da transparência da plataforma foi apresentada por [Andreou et al. 2018] e o impacto dos anúncios na privacidade dos usuários foi quantificado em [Venkatadri et al. 2018]. A interface de anúncios foi explorada pela *Russian Intelligence Research Agency* (IRA) para fazer anúncios microdirecionadas aos usuários durante as eleições dos Estados Unidos em 2016 [Ribeiro et al. 2017] e foi estudada por [Zagheni et al. 2017] para identificar o número de imigrantes em uma dada região. Neste trabalho, usamos a metodologia baseada em bots, ou personas [Carrascosa et al. 2015], para auditar um outro produto do Facebook, a *News Feed*.

Neste trabalho, mostramos como modelos de *caches* e filas podem ser usados no estudo de *timelines*. Existe uma vasta literatura sobre ambos [Martina et al. 2014, Dehghan et al. 2016, Harchol-Balter 2013]. Acreditamos que esta conexão aqui estabelecida permita estender-se resultados destes domínios para as *timelines* uma vez que a lite-

ratura de modelos analíticos nessa área é bem mais escassa. O primeiro a modelar uma timeline com um fila FIFO foi apresentado em [Altman et al. 2013]. Timelines FIFO com publicações de tamanho variável foram introduzidas em [Reiffers-Masson et al. 2017]. O modelo proposto neste trabalho, que é uma versão estendida do trabalho inicialmente apresentado em [Hargreaves et al. 2018], apresenta uma solução do sistema capaz de obter além da visibilidade, a ocupação das fontes. Em particular, não é de nosso conhecimento nenhum trabalho anterior que tenha feito uso de medições reais de *timelines*, conectando tais medições com modelos analíticos.

Um modelo baseado em temporizadores foi proposto em [Hargreaves et al. 2019] para controlar *timelines*. Ajustando esses temporizadores, é possível controlar o fluxo de publicações nas *timelines* de forma a introduzir mecanismos de justiça aos seus algoritmos de personalização equilibrando as preferências dos usuários, os objetivos das fontes e trazendo diversidade aos conteúdos exibidos. Todas as análises contidas em [Hargreaves et al. 2019, Reiffers-Masson et al. 2017] e em [Hargreaves et al. 2018] são baseadas no estado estacionário. No presente trabalho, em contrapartida, apresentamos uma análise das séries temporais das ocupações das fontes, bem como um mecanismo de predição das ocupações numa janela de tempo futura dado o histórico de ocupações.

## 9. Discussão

**Modelo FIFO** O modelo FIFO proposto é um passo preliminar para capturar a dinâmica do News Feed. Esse modelo simples já nos permite avaliar que tipo de análise pode ser conduzida combinando medições e ferramentas analíticas. Um modelo mais complexo, por exemplo, para contabilizar os diferentes tempos de residência para diferentes publicações usando temporizadores [Dehghan et al. 2016, Kelly 1997] foi apresentado em [Hargreaves et al. 2019]. No entanto, se todas as publicações de *timelines* temporizadas possuem o mesmo valor para o temporizador, em [Hargreaves et al. 2019] foi mostrado que *timelines* FIFO são um caso particular de *timelines* temporizadas.

**Pequeno número de usuários** Nossa metodologia experimental engloba um pequeno número de bots (usuários), já que cada bot precisa passar por um registro completo da conta do Facebook. Em particular, em nosso estudo de caso, consideramos seis bots. Acreditamos que esses bots já fornecem uma perspectiva representativa dos vieses introduzidos pelo Facebook. Em trabalhos futuros pretendemos aumentar o número de bots para dezenas de bots, incluindo bots com perfis semelhantes para comparar suas impressões. Também pretendemos comparar as *timelines* dois tipos de bots indecisos: um bot que não curte nenhuma publicação e um curte publicações de todas as orientações.

**Polarização dos usuários e outros cenários de casos de uso** Apesar da polarização dos usuários ter sido feita de modo arbitrário, não fazemos nenhum juízo de valor sobre o viés identificado pelo modelo. Da forma que o experimento foi planejado, qualquer critério para agrupar as fontes seria adequado, desde que os bots curtissem apenas as publicações das fontes selecionadas. Sendo assim, esse experimento pode ser aplicado em outros contextos. Inclusive, repetimos este experimento, com mais usuários por perfil e dois tipos de usuários indecisos, durante as eleições presidenciais brasileiras com o objetivo de averiguar a generalidade do modelo. Então utilizamos um método de aprendizado por máquina não supervisionado que analisa o conteúdo das publicações das fontes e identifica as semelhanças entre as fontes e, conseqüentemente, suas polarizações.

Os resultados deste experimento serão publicados em trabalhos futuros.

**Conformidade com as políticas do Facebook** Nossa metodologia abrange a criação de usuários falsos, pois usamos um número limitado de contas experimentais do Facebook para seguir e gostar de páginas públicas. Essas contas têm um impacto muito baixo no sistema, pois não possuem amigos reais ou virtuais. Para minimizar possíveis danos, removemos nossas contas experimentais do Facebook depois que terminamos nossos experimentos

## 10. Conclusões e trabalhos futuros

O Facebook afeta milhões de usuários da Internet todos os dias, e qualquer decisão algorítmica sobre as *timelines* pode ter importantes impactos sociais e políticos. Neste trabalho, nos inspiramos na teoria de filas e nas análises de caches para propor um modelo baseado numa fila FIFO para modelar o comportamento das *timelines* dos usuários e quantificar suas métricas de ocupação e visibilidade. Através de um estudo de caso feito durante as eleições presidenciais Italianas de 2018, validamos o modelo utilizando dados reais coletados da *News Feed* do Facebook. Por meio da criação de perfis de usuários, pudemos utilizar o modelo para fazer análise contrafactual e quantificar o viés da *News Feed* do Facebook utilizando uma *timeline* não filtrada como referência. Finalmente, utilizamos um método de análise de séries temporais para fazer uma análise transiente das *timelines*. Dado o histórico de ocupações, este método se mostrou capaz de prever as ocupações das fontes na próxima visualização da *timeline*.

Devido a todas as contribuições apresentadas neste trabalho, acreditamos que este seja um importante passo no sentido de habilitar os usuários a fiscalizar, auditar e interagir com suas plataformas de redes sociais, tornando-os mais informados sobre os processos de filtragem aos quais as publicações por eles visualizadas estão submetidas. Como trabalhos futuros, pretendemos repetir a metodologia apresentada neste trabalho utilizando as eleições brasileiras como estudo de caso. As lições apreendidas no experimento nos motivou a superar algumas limitações enfrentadas durante a realização do estudo de caso apresentado neste artigo. Também pretendemos utilizar outros métodos para prever a ocupação das *timelines* durante horizonte de tempo maiores.

## Referências

- Ali, M., Sapiezynski, P., Bogen, M., Korolova, A., Mislove, A., and Rieke, A. (2019). Discrimination through optimization: How Facebook’s ad delivery can lead to skewed outcomes.
- Altman, E., Kumar, P., Venkatramanan, S., and Kumar, A. (2013). Competition over timeline in social networks. *ASONAM*, pages 1352–1357.
- Andreou, A., Venkatadri, G., Goga, O., Gummadi, K. P., Loiseau, P., and Mislove, A. (2018). Investigating Ad Transparency Mechanisms in Social Media : A Case Study of Facebook ’ s Explanations. (February).
- Arnaudo, D. (2017). Computational Propaganda in Brazil : Social Bots during Elections.
- Bakshy, E., Messing, S., and Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239):1130–1132.

- Bakshy, E., Rosenn, I., Marlow, C., and Adamic, L. (2012). The role of social networks in information diffusion. In *Proceedings of the 21st international conference on World Wide Web - WWW '12*, page 519, New York, NY, USA. ACM Press.
- Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E., and Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415):295.
- Bucher, T. (2012). Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. *New Media & Society*.
- Cadwalladr, C. (2019). Facebook's role in brexit - and the threat to democracy. Retrieved April 21, 2019 from: [https://www.ted.com/talks/carole\\_cadwalladr\\_facebook\\_s\\_role\\_in\\_brexit\\_and\\_the\\_threat\\_to\\_democracy?language=en](https://www.ted.com/talks/carole_cadwalladr_facebook_s_role_in_brexit_and_the_threat_to_democracy?language=en).
- Carrascosa, J. M., Cuevas, R., and Laoutaris, N. (2015). I Always Feel Like Somebody 's Watching Me Measuring Online Behavioural Advertising.
- Dehghan, M., Massoulie, L., Towsley, D., Menasché, D., and Tay, Y. C. (2016). A utility optimization approach to network cache design. In *Proceedings - IEEE INFOCOM*, volume 2016-July, pages 1–10.
- Diakopoulos, N. (2013). Algorithmic accountability reporting: On the investigation of black boxes. *Tow Center for Digital Journalism A Tow/Knight Brief*, pages 1–33.
- Diakopoulos, N. (2016). A view from computational journalism. *Communications of the Acm*, 59(2).
- Epstein, R. and Robertson, R. E. (2015). The search engine manipulation effect and its possible impact on the outcomes of elections. *Nat. Academy of Sciences of the United States of America*, 112(33):E4512–21.
- Eslami, M., Karahalios, K., Sandvig, C., Vaccaro, K., Rickman, A., Hamilton, K., and Kirlik, A. (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*, pages 2371–2382. ACM.
- Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., Hamilton, K., and Sandvig, C. (2015). I always assumed that I wasn't really that close to [her]. *Human Factors in Computing (HCI)*, pages 153–162.
- Facebook (2017). Controlling what you see in news feed. Retrieved December 31, 2017 from: <https://www.facebook.com/help/335291769884272>.
- Facebook (2018a). Graph api reference /user-id/home. Retrieved April 7, 2018 from: <https://developers.facebook.com/docs/graph-api/reference/v2.12/user/home>.
- Facebook (2018b). Log de alterações da graph api. Retrieved April 6, 2018 from: <https://developers.facebook.com/docs/graph-api/changelog>.
- Garfinkel, S., Matthews, J., Shapiro, S. S., and Smith, J. M. (2017). Toward algorithmic transparency and accountability. *Communications of the ACM*, 60(9):5–5.

- Harchol-Balter, M. (2013). *Performance Modeling and Design of Computer Systems*. Cambridge University Press, Cambridge.
- Hargreaves, E., Agosti, C., Menasché, D., Neglia, G., Reiffers-Masson, A., and Altman, E. (2019). Fairness in online social network timelines: Measurements, models and mechanism design. *Performance Evaluation*, 129(xxxx):15–39.
- Hargreaves, E., Menasché, D., Neglia, G., and Agosti, C. (2018). Visibilidade no facebook: Modelos, medições e implicações. In *BraSNAM*, volume 7. SBC.
- Hyndman, R. J. and Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal Of Statistical Software*, 27(3):C3–C3.
- Kelly, F. (1997). Charging and rate control for elastic traffic. *European Transactions on Telecommunications*, 8(1):33–37.
- Kramer, A. D. I., Guillory, J. E., and Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*.
- Kulshrestha, J., Eslami, M., Messias, J., Zafar, M. B., Ghosh, S., Gummadi, K. P., and Karahalios, K. (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), Portland, USA. February 2017.*, pages 417–432.
- Martina, V., Garetto, M., and Leonardi, E. (2014). A unified approach to the performance analysis of caching systems. *IEEE INFOCOM*, pages 2040–2048.
- Newman, N., Richard Fletcher, A. K., Levy, D. A. L., and Nielsen, R. K. (2018). Digital news report 2018. Retrieved April 21, 2019 from: <http://www.digitalnewsreport.org>.
- O’Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Publishing Group, New York, NY, USA.
- Pariser, E. (2011). *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. Penguin Publishing Group.
- Reiffers-Masson, A., Hargreaves, E., Altman, E., Caarls, W., and Menasché, D. S. (2017). Timelines are Publisher-Driven Caches. *ACM SIGMETRICS Performance Evaluation Review*, 44(3):26–29.
- Ribeiro, F. N., Saha, K., Babaei, M., Henrique, L., Messias, J., Goga, O., Benevenuto, F., Gummadi, K. P., and Redmiles, E. M. (2017). On Microtargeting Socially Divisive Ads: A Case Study of Russia-Linked Ad Campaigns on Facebook.
- Sandvig, C., Hamilton, K., Karahalios, K., and Langbort, C. (2014). Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms. *Int. Comm. Association*, pages 1–20.
- Sleeper, M., Balebako, R., Das, S., McConahy, A. L., Wiese, J., and Cranor, L. F. (2013). The post that wasn’t: exploring self-censorship on facebook. *Proceedings of the 2013 conference on Computer supported cooperative work*, page 793.

- Sun, E., Rosenn, I., Marlow, C. A., and Lento, T. M. (2009). Gesundheit ! Modeling Contagion through Facebook News Feed Mechanics of Facebook Page Diffusion. *Proceedings of the 3rd International ICWSM Conference*, (2009):146–153.
- TechCrunch (2016). How facebook news feed works. Retrieved December 31, 2017 from: <https://techcrunch.com/2016/09/06/ultimate-guide-to-the-news-feed/>.
- TheGuardian (2018). The cambridge analytica files. Retrieved April 21, 2019 from: <https://www.theguardian.com/news/series/cambridge-analytica-files>.
- Tsaparas, P. (2017). Online social networks and media. <http://www.cs.uoi.gr/~tsap/teaching/cs-114/references.html>.
- Venkatadri, G., Andreou, A., Liu, Y., Mislove, A., Gummadi, K. P., Loiseau, P., and Goga, O. (2018). Privacy risks with facebook’s pii-based targeting: Auditing a data broker’s advertising interface. In *2018 IEEE Symposium on Security and Privacy (SP)*, pages 89–107. IEEE.
- Wickham, Hadley; Golemund, G. (2017). *R for Data Science*.
- Woolley, S. C. and Howard, P. N. (2017). Computational propaganda worldwide: executive summary.
- Zagheni, E., Weber, I., and Gummadi, K. (2017). Leveraging Facebook’s Advertising Platform to Monitor Stocks of Migrants. *Population and Development Review*, 43(4):721–734.