

Filters for Social Media Timelines: Models, Biases, Fairness and Implications

Eduardo Martins Hargreaves^{1,2}, Daniel Sadoc Menashé²

¹Petrobras – Rio de Janeiro – RJ– Brazil.

²Departamento de Ciência da Computação – Universidade Federal do Rio de Janeiro (UFRJ)
Rio de Janeiro – RJ– Brazil.

Resumo. *As mídias sociais têm um impacto significativo, no estilo de vida, comportamento e opinião de bilhões de usuários. Para lidar com o fluxo de informações entre seus membros, elas desenvolveram algoritmos de personalização que filtram os conteúdos vistos pelos usuários. Apesar de sua grande abrangência, estes filtros não são de domínio público motivando pesquisas para entender e melhorar suas propriedades. Nesta tese, unindo teoria de filas, modelos de caches e funções de maximização de utilidade de redes, propomos uma metodologia reprodutível que engloba medições, modelos analíticos e funções de utilidade para projetar os filtros das mídias sociais. Usando o Facebook como estudo de caso, nossos resultados indicam que existe um viés significativo nas timelines de seus usuários, sendo mais forte no seu topo, motivando a proposta de um novo e transparente método de filtragem que pode ser controlado pelos usuários. Entre as implicações, indicamos a precisão do nosso modelo para a realização de análises contrafactuais, a sua capacidade de auditar as mídias sociais e a sua versatilidade na construção de diversos filtros respeitando de forma transparente a preferência dos usuários.*

Abstract. *Social media have a significant impact, on the lifestyle, behaviour and opinion of billions of users. To handle the flow of information between its members, social media developed personalization algorithms that filter the contents that flow into users' timelines. Despite the far-reach of social media filters, such algorithms lack transparency, motivating research to understand and improve its properties. In this thesis, bridging queuing theory, caching models and network utility maximization, we propose a reproducible methodology encompassing measurements, analytical models and a utility-based method to design timelines filters. Using Facebook as a case study, our empirical results indicate that a significant bias exists and it is stronger at the topmost position of News Feed motivating the proposal of a novel and transparent fairness-based timeline design which can be controlled by users. Among the implications, we indicate the accuracy of our model to make counterfactual analysis, the capability of auditing social media and its versatility in designing multiple filters accounting for users preferences in an open and transparent way.*

1. Introduction

Social media are online platforms in which users connect to each other to build social relationships with other users who share similar interests. Examples of such social me-

dia platforms include: Friendster, Orkut, MySpace, Facebook, Twitter, LinkedIn, Instagram, etc. Amongst many possible interests, members of social media make use of platforms for reading and sharing political opinion and news, particularly during election periods [Woolley and Howard 2017]. The 2017 Brazilian ICT report published by Cetic.br says that 50% of the Brazilian population make use of Internet to read news, while 68% of Internet users have shared some kind of content in social media in 2016 [Cetic.br 2017].

With almost 3.5 billions of users and an overall penetration of 45% of the world population, social media members spent on average 2 hours and 16 minutes per day [Kemp 2019]. To deal with user generated content, social media developed personalization algorithms to organize the available information to improve user experience. For instance, Facebook developed a filtering and personalization product called *News Feed*, Twitter developed the *top Tweets first* feature, LinkedIn developed the *LinkedIn Feed*, etc. In essence, these algorithms are recommender systems that filter posts to users in their timelines. Unlike classical recommendation systems, e.g., for music and movies, social media filtering algorithms pose their own set of challenges: the need to account for content continuously generated by multiple sources (publishers), users consume many publications quickly without providing explicit feedback about the recommendations.

The mechanisms used by social media platforms to filter messages presented to users are not in the public domain, motivating researches that range from basic statistics to reverse algorithm engineering. For instance, the Cambridge Analytica scandal involving Facebook and the possibility that Facebook's had influenced both the 2016 US presidential elections ¹ and in *Brexit* illustrate the fundamental need to guarantee a certain level of supervision of social media by its users and society as a whole [Garfinkel et al. 2017, Diakopoulos 2016]. This monitoring becomes even more essential because users prefer to receive news via *gateways* such as social media rather than directly from news portals. These new sort of gateways are responsible for over 30% of all web site visits [Newman et al. 2018].

In this thesis, we take important steps towards measuring, modeling and evaluation of the influence of the filters made by social media algorithms using Facebook News Feed as a case study. We believe that a better understanding of social media customization algorithms can be instrumental in the design of new mechanisms, which may eventually be actively controlled by users. Furthermore, obtaining this knowledge through public and reproducible measurements is an essential step for increasing the awareness of users about the filtering process they are submitted to.

2. Motivation

Accordingly to [Eslami et al. 2015], 60% of Facebook users believe that they express their publication's preferences only through their connections and social ties, being unaware of the News Feed algorithm influence. The hidden characteristic of personalization algorithms, combined with proprietary codes, creates a *filter bubble* that potentially reinforces the perceived users' opinions [Pariser 2011] by creating *feedback loops* [Rossi et al. 2018] where users can not *opt out* while social media members.

Trapped in filter bubbles, users can not reach a common vision of reality, as it is difficult to understand what each other believe in and why each other have their beliefs.

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

Therefore, the ability to critically understand, and deal with contrary opinions, especially in the context of public sphere, is potentially damaged [Lichfield 2018].

Therefore, the News Feed algorithm and other social media filtering algorithms create a new form of public sphere affecting the way that public discourse is perceived, potentially creating biases towards sources and ultimately impacting how democracy works as well as the society as a whole [O’Neil 2016]. This ubiquity of filters in our everyday lives motivates measurements, modeling and analysis in order to understand and improve social media properties.

3. Objectives

Our goal is to provide insights on the filtering that occurs in social media through a reproducible methodology with datasets made publicly available.² 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?
3. determine how often users should check their timelines. If users do not access social media frequently, important posts might be missed. On the other hand, if users access too often, the chance that there will be no new relevant content created in the interval between accesses is high. Given a set of publishers followed by users, which is the optimal frequency for the users to check their timelines ?

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 counterfactual 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 [Dehghan et al. 2019], and accounts for fairness among publishers through utility functions.

To address the third objective, we propose an analytical model that allows us to determine, depending on the rate at which sources generate content, the chance of a user accessing the network and obtaining new content. The proposed model yields a closed form expression for the value of an access (VoA), which we use as basis for an optimization problem that provides *insights* on the optimal access rate.

4. Contributions

In this research, we propose a methodology that encompasses measurements and models for the analysis of filters in social media. Our key contributions are summarized as follows.

²<https://github.com/EduardoHargreaves/Effect-of-the-OSN-on-the-elections>

An analytical model for filtered timelines derived from queuing theory is proposed to quantify the visibility and occupancy of publishers in the users News Feeds in function of publishers filtered arrival rates (λ_{ij}) and News Feed size (K). In the model, the occupancy of publisher j at user i (N_{ij}) is given by: $N_{ij} = \lambda_{ij}K / \sum_j \lambda_{ij}$. The model allows us to conduct a counterfactual analysis to assess the metrics of interest under different filtering mechanisms and is validated using data from the Italian election experiment.

Empirical findings are reported using data collected from a measurement campaign conducted during the 2018 Italian elections. We observed that *a)* the filtering algorithm tends to select information that is aligned with user’s perceived political orientation, *b)* this effect is more prominent at the topmost News Feed position and *c)* neutral users are also exposed to non-uniform filtering.

A fairness-driven mechanism design is proposed, bridging the proposed models, measurements, caching models and network utility maximization theory. 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 (w_{ij}) and timeline size under a utility maximization framework. For instance, under proportional fairness ($\alpha = 1$), the optimal occupancy (N_{ij}^*) of publisher j at user i is given by: $N_{ij}^* = w_{ij}K / \sum_k w_{ik}$.

A model-based bias assessment is conducted using the Italian dataset. The dataset is used to parameterize the proposed model, and yields a baseline publisher visibility (i.e., without the influence of the News Feed algorithm). The measured visibility is then contrasted against the baseline to quantify the *bias*, i.e., how publishers’ occupancies are affected by user’s orientations as they are perceived by the News Feed algorithm.

Value of access: Is proposed a new metric, referred to as *value of access*, or VoA, to measure the expected amount of new information received by a social media user (\bar{V}) per access in terms of posts creation rate (λ), users access rate μ and timeline size (K). We present the following analytical model with closed forms expressions: $\bar{V} = (\lambda/\mu)[1 - (\lambda/(\lambda + \mu))^K]$. Both metric and model are evaluated using data collected from Facebook during the 2018 Brazilian presidential elections.

Optimal access frequency is determined: We pose an optimization problem to find the optimal access rate (μ^*), balancing between the cost per access (c) and the utility derived from it (\bar{V}). The model shows that, if a user scrolls the screen to see K posts per access, the optimal access rate is: $\mu^* = \lambda[(\sqrt[K+1]{K} / \sqrt[K+1]{c}) - 1]$.

5. Related work

In what follows, we present a broad overview of previous work related to the topics discussed in the thesis. More detailed bibliographic review is presented at each chapter of [Hargreaves 2019].

Social media timeline models: Performance analysis of network of timelines was proposed by Giovanidis et al. [Giovanidis et al. 2019] and the competition for visibility at timelines were previously studied in [Altman et al. 2013, Dhouchak et al. 2017]. All these related works evaluates the *organic* performance of timelines. This thesis propose models for the metrics of interest for *filtered* network of timelines.

Facebook experiments: This thesis enables the design of experiments that seek

to quantify the filtering that occurs in Facebook News Feed through a reproducible methodology that does not require the participation of the platform that will be audited.

This fact is particularly important because most of the prior work that quantifies the effect of Facebook on information diffusion with massive datasets [Bakshy et al. 2012, Bakshy et al. 2015, Adam D. I. Kramer and Hancock 2014, Sun et al. 2009, Bond et al. 2012] 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.

Bias on social media: Multiple sources of algorithmic bias were described in [Baeza-Yates 2018] and forms to audit it were investigated in [Diakopoulos 2013, Sandvig et al. 2014]. The models proposed in this thesis enables a bias analysis and further foster accountability, as they allow what-if analysis with *unfiltered* timelines occupancies as *baselines*.

Fairness on computer networks and algorithms: The literature on fairness accounts for its implications on social networks [Valenzuela et al. 2009], risk score estimation [Kleinberg et al. 2017, Barabas et al. 2017], recommender system [Krishnasamy et al. 2016, Singh and Joachims 2018], individuals classification in order to prevent discrimination [Dwork et al. 2012] and computational policy [Gilbert et al. 2018].

In this thesis we apply the Network Utility Maximization framework proposed by Kelly [Kelly 1997] to social media timelines. The framework was already applied to caches by [Dehghan et al. 2019]. Both approaches are a form of *occupancy-based* fairness instead the traditional *rate-based* fairness developed by Kelly.

Social media and democracy: In [Epstein and Robertson 2015, Kulshrestha et al. 2017] it was shown that search engine rank manipulation can influence the vote of undecided citizens and decide election won by small margins. Fuchs made and critical analysis of the impact of social media in Arab Spring and in the Occupy Movement [Fuchs 2017]. The Facebook Ads interface was exploited by the Russian Intelligence Research Agency (IRA) to make micro-targeted ads to influence users during the US elections in 2016 [Ribeiro et al. 2017]. A detailed analysis of network propaganda during the 2016 presidential election can be found in [Benkler et al. 2018].

This thesis relies on the 2018 Italian and Brazilian general elections as case studies for the models developed and incorporates new results for the influence of the Facebook News Feed algorithm on the users' information diets during elections campaigns.

6. Publications and Awards

The results of the thesis were published in some of the most prestigious conferences and journals of the network and performance evaluation communities; . Next, follows the exhaustive list of papers published during my research.

Journal Publications

- Eduardo Hargreaves, Claudio Agosti, Daniel Menasché, Giovanni Neglia, Alexandre Reiffers-Masson and Eitan Altman. Fairness in online social

network timelines: Measurements, models and mechanism design. *Performance Evaluation* (2019). Volume 129, February 2019, Pages 15-39. DOI:10.1016/j.peva.2018.09.009. Qualis A2

- Eduardo Hargreaves, Claudio Agosti, Daniel Menasché e Giovanni Neglia. Modelando, Auditando e Prevendo a Visibilidade de Fontes no Facebook: Um Estudo de Caso nas Eleições Italianas. *iSys - Revista Brasileira de Sistemas de Informação*. Qualis B3.

International Conferences and Workshops Papers

- Eduardo Hargreaves, Daniel Sadoc Menasché Giovanni Neglia. How often should I check my social networks?. *27th IEEE International Symposium on the Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS 2019)*. DOI: 10.1109/MASCOTS.2019.00028. Qualis A2.
- Eduardo Hargreaves and Claudio Agosti and Daniel Menasche and Giovanni Neglia and Alexandre Reiffers-Masson and Eitan Altman; Fairness in Online Social Network Timelines : Measurements , Models and Mechanism Design. *36th International Symposium on Computer Performance, Modeling, Measurements and Evaluation 2018, IFIP Performance 2018*
- Eduardo Hargreaves, Claudio Agosti, Daniel Menasche, Giovanni Neglia, Alexandre Reiffers-Masson and Eitan Altman. Biases in the Facebook News Feed: a Case Study on the Italian Elections, Fosint-SI 2018, in conjunction with ASONAM 2018, Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) DOI:10.1109/ASONAM.2018.8508659
- Alexandre Reiffers-Masson, Eduardo Hargreaves, Eitan Altman, Wouter Carls, Daniel Sadoc Menasché; Timelines are Publisher-Driven Caches: Analyzing and Shaping Timeline Networks, NetEcon 2016, in conjunction with ACM SIGMETRICS 2016, ACM SIGMETRICS Performance Evaluation Review, DOI:10.1145/3040230.3040237

Brazilian Conferences and workshops

- Eduardo Hargreaves, Daniel Sadoc Menasché Giovanni Neglia. Com que frequência devo acessar minhas redes sociais?, in: Proceedings of the Brazilian Workshop on Social Network Analysis and Mining (BraSNAM). Belém, Pará. Julho, 2019. Qualis B5
- Eduardo Hargreaves, Daniel Sadoc Menasché, Giovanni Neglia, Claudio Agosti, Visibilidade no Facebook: Modelos, Medições e Implicações, in: Proceedings of the Brazilian Workshop on Social Network Analysis and Mining (BraSNAM). Natal, Rio Grande do Norte. Julho, 2018. **Best Paper award**. Qualis B5

Referências

- Adam D. I. Kramer, J. E. G. and Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *PNAS*.
- Altman, E., Kumar, P., Venkatramanan, S., and Kumar, A. (2013). Competition over timeline in social networks. *ASONAM '13*.
- Baeza-Yates, R. (2018). Bias on the web. *Communications of the ACM*, 61(6):54–61.

- 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 *WWW '12*, page 519, New York, NY, USA. ACM Press.
- Barabas, C., Dinakar, K., Virza, J. I., Zittrain, J., et al. (2017). Interventions over predictions: Reframing the ethical debate for actuarial risk assessment. *arXiv:1712.08238*.
- Benkler, Y., Faris, R., and Roberts, H. (2018). *Network Propaganda*, volume 6. Oxford University Press, Oxford, England, UK.
- 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.
- Cetic.br (2017). ICT households - Survey on the Use of Information and Communication Technologies in Brazilian Households.
- Dehghan, M., Massoulie, L., Towsley, D., Menasche, D. S., and Tay, Y. C. (2019). A utility optimization approach to network cache design. *IEEE/ACM Trans. Netw.*, 27(3).
- Dhouchak, R., Kavitha, V., and Altman, E. (2017). A Viral Timeline Branching Process to study a Social Network. *{ITC} PhD Workshop 2017*.
- 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).
- Dwork, C., Hardt, M., Pitassi, T., Reingold, O., and Zemel, R. (2012). Fairness through awareness. In *Proceedings of ITCS*, pages 214–226. ACM.
- Epstein, R. and Robertson, R. E. (2015). The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *PNAS*, 112(33):E4512–21.
- Eslami, M., Aleyasen, A., Karahalios, K., Hamilton, K., and Sandvig, C. (2015). Feedvis: A path for exploring news feed curation algorithms. In *Proceedings of CSCW*. ACM.
- Fuchs, C. (2017). *Social Media: A Critical Introduction*. Sage, 2 nd edition.
- 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.
- Gilbert, G., Ahrweiler, P., Barbrook-Johnson, P., Narasimhan, K., and Wilkinson, H. (2018). Computational modelling of public policy: Reflections on practice. *Journal of Artificial Societies and Social Simulation*, 21(1):1–14.
- Giovanidis, A., Baynat, B., and Vendeville, A. (2019). Performance analysis of online social platforms. In *2019 Proceedings of IEEE INFOCOM*.
- Hargreaves, E. (2019). *Filters for Social Media Timelines: Models, Biases, Fairness and Implications*. PhD thesis, PPGI/UFRJ.
- Kelly, F. (1997). Charging and rate control for elastic traffic. *European Transactions on Telecommunications*, 8(1):33–37.

- Kemp, S. (2019). Digital 2019 reports. Hootsuite, we are social. Retrieved August 19, 2019 from: <https://datareportal.com/reports/digital-2019-global-digital-overview>.
- Kleinberg, J., Mullainathan, S., and Raghavan, M. (2017). Inherent Trade-Offs in the Fair Determination of Risk Scores. *ITCS 2017*.
- Krishnasamy, S., Sen, R., Shakkottai, S., and Oh, S. (2016). Detecting Sponsored Recommendations. *ACM Transactions on Modeling and Performance Evaluation of Computing Systems*, 2(1):1–29.
- 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. *CSCW 2017*, pages 417–432.
- Lichfield, G. (2018). Technology is threatening our democracy. How do we save it? *MIT Technology Review*.
- 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: What the Internet Is Hiding from You*. The Penguin Group, East Rutherford, NJ, USA.
- 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.
- Rossi, W. S., Polderman, J. W., and Frasca, P. (2018). The closed loop between opinion formation and personalised recommendations. pages 1–24.
- Sandvig, C., Hamilton, K., Karahalios, K., and Langbort, C. (2014). Auditing Algorithms : Research Methods for Detecting Discrimination on Internet Platforms. *International Communication Association*, pages 1–20.
- Singh, A. and Joachims, T. (2018). Fairness of Exposure in Rankings. In *KDD ’18*, pages 2219–2228, New York, New York, USA. ACM Press.
- 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.
- Valenzuela, S., Park, N., and Kee, K. F. (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):875–901.
- Woolley, S. C. and Howard, P. N. (2017). Computational propaganda worldwide: executive summary. (11).