



HAL
open science

Com que frequência devo acessar as minha redes sociais?

Eduardo Hargreaves, Daniel Menasché, Giovanni Neglia

► To cite this version:

Eduardo Hargreaves, Daniel Menasché, Giovanni Neglia. Com que frequência devo acessar as minha redes sociais?. BraSNAM 2019 - 8th Brazilian Workshop on Social Network Analysis and Mining, Jul 2019, Belem, Brazil. hal-02411237

HAL Id: hal-02411237

<https://hal.inria.fr/hal-02411237>

Submitted on 14 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Com que frequência devo acessar as minha redes sociais?

Eduardo Hargreaves¹, Daniel Menasché¹, Giovanni Neglia²

¹DCC, UFRJ, Rio de Janeiro, Brasil, ²INRIA, Sophia Antipolis, França

Abstract. *The time users spend on social networks is in vogue. The retention of social networks has important implications, encompassing economic, psychological and infrastructure aspects. In this article, we consider the problem of determining the optimal rate at which users should access a social network. To this aim, 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, referred to as the value of access (VoA). Next, we pose an optimization problem wherein the utility of users equals the VoA minus the costs to access the network. Using the proposed framework, we provide insights on the optimal access rate. Our results are parameterized using Facebook data, indicating the predictive power of the approach.*

Resumo. *O tempo que usuários passam nas redes sociais é um tema em voga. A retenção das redes sociais tem importantes implicações, que extrapolam o campo social atingindo aspectos econômicos, psicológicos e de infraestrutura da rede. Neste artigo, consideramos o problema da determinação da taxa ótima de acesso a uma rede social. Para tal, propomos um modelo analítico que permite determinar, em função da taxa com que as fontes geram conteúdos, a chance de um usuário acessar a rede e obter conteúdos novos. Em seguida, usamos os resultados do modelo para alimentar um problema de otimização, no qual assumimos que cada acesso à rede social envolve custos. Usando o modelo e o problema de otimização, provemos insights sobre a taxa ótima de acesso. Nossos resultados são parametrizados usando dados do Facebook, que indicam o poder preditivo da abordagem proposta.*

1. Introdução

O tempo que usuários passam nas redes sociais é um tema em voga. A última versão do iOS (12) e versões beta do Android não só emitem relatórios sobre o tempo gasto nos aplicativos instalados nos *smartphones* como também permitem ao usuário limitar o seu tempo de uso [Ceres 2018]. O tempo gasto nas redes sociais tem importantes implicações que extrapolam o campo social atingindo aspectos econômicos, psicológicos e de infraestrutura da rede. Sob o ponto de vista econômico, a atenção é um recurso escasso numa era na qual a informação é abundante [Simon 1971], i.e., a atenção é um dos recursos mais disputados na *economia da atenção* [Shapiro and Varian 1998].

Nas plataformas de redes sociais, a disputa pela atenção ocorre por meio da contagem de cliques e pelo tempo gasto visualizando publicações. Usuários utilizam as plataformas de redes sociais para criar e consumir o conteúdo criado pelas fontes que o usuário segue. No entanto, se um usuário acessa muito suas redes sociais, a chance de não haver conteúdo novo criado no intervalo entre os acessos aumenta. Em última análise, o acesso muito frequente dos usuários pode ser um problema para os usuários (e.g., refletindo

vícios ou consumindo recursos desnecessariamente) e para os gerenciadores das redes sociais (e.g., a falta de informação nova pode gerar uma queda da qualidade da experiência dos usuários prejudicando o engajamento com a plataforma). *Desta forma, trata-se de um importante desafio determinar a frequência ideal com que usuários devem acessar uma rede social, levando em conta os múltiplos e complexos fatores envolvidos.*

A literatura sobre retenção de usuários em redes sociais foca primordialmente no efeito da retenção dos usuários sobre o lucro advindo de propagandas e publicações promovidas ou patrocinadas [Jiang et al. 2013]. Em essência, assume-se que quanto maior for a frequência com que um usuário acessa a rede, maior será a chance de um usuário consumir um produto patrocinado, e as plataformas de redes sociais obtêm suas receitas exibindo anúncios aos seus usuários [Facebook 2019, Twitter 2019]. O usuário potencialmente arca com um custo monetário e de tempo para acessar as plataformas de redes sociais, e muitas vezes acaba recebendo conteúdo repetido, ou conteúdo que não necessariamente lhe convém. *Não é de nosso conhecimento nenhum trabalho anterior que tenha analisado a frequência ótima com que usuários devam acessar suas redes sociais, tendo em vista o valor intrínseco associado a cada acesso do usuário às redes.*

Neste artigo, consideramos o problema da determinação da taxa ótima de acesso a uma rede social. Para tal, propomos um modelo analítico que permite determinar, em função da taxa com que as fontes geram conteúdos, a chance de um usuário acessar a rede e obter conteúdos novos. Em seguida, usamos os resultados do modelo para alimentar um problema de otimização, no qual assumimos que cada acesso à rede social envolve custos. Usando o modelo e o problema de otimização, provemos *insights* sobre a taxa ótima de acesso. Nossos resultados são parametrizados usando dados do Facebook, que indicam o poder preditivo da abordagem proposta.

Em resumo, nossas principais contribuições são listadas a seguir.

Nova métrica: Criamos uma nova métrica, que chamamos de *valor de um acesso* (em inglês, *value of access*, ou VoA), para mensurar a quantidade de informação transmitida por uma plataforma de redes sociais. A métrica é avaliada usando dados colhidos do Facebook;

Modelagem analítica: Propomos um modelo analítico para estimar a quantidade esperada de novas impressões por cada visualização da sua *timeline*. O modelo produz fórmulas fechadas simples para a métrica de interesse;

Frequência ótima de acesso: Apresentamos um problema de otimização que produz como resultado a frequência ótima de acesso, balanceando entre o custo por acesso e a utilidade advinda do mesmo. Usando o modelo, indicamos analiticamente propriedades da solução ótima.

O restante do artigo está organizado da seguinte forma. Na Seção 2 introduzimos o modelo analítico que é validado na Seção 3. O problema da frequência ótima de acesso é apresentado na Seção 4, e exemplos numéricos são apresentados na Seção 5. Trabalhos relacionados estão na Seção 6. Uma discussão sobre as limitações deste trabalho é feita na Seção 7 e a Seção 8 conclui o trabalho.

2. Modelo

Seja i um usuário que segue um conjunto \mathcal{J} de fontes. A plataforma de redes sociais escolhe as publicações deste conjunto de fontes seguidas por i e as exibe na *timeline* do

usuário i . Esse conjunto de fontes cria publicações seguindo um processo de Poisson com uma taxa agregada de λ publicações por unidade de tempo. Uma *timeline* da tamanho K retorna K publicações por acesso do usuário i .

Seja T a variável aleatória que caracteriza o intervalo entre visualizações. O intervalo médio entre visualizações é dado por \bar{T} . Adicionalmente, denotamos por $f_T(t)$ a função de densidade de probabilidade de T . Eliminamos o subscrito quando não houver ambiguidades, denotando a densidade simplesmente como $f(t)$ para facilitar a notação.

Seja μ a frequência com que o usuário de interesse acessa a rede social. Tal frequência é também referida como a taxa de acesso (ou taxa de visualizações), e é nossa principal quantidade de interesse. Adicionalmente, definimos ρ como o número médio de postagens criadas entre duas visualizações do usuário à *timeline*. Temos então

$$\mu = \frac{1}{\bar{T}}, \quad \rho = \frac{\lambda}{\mu}. \quad (1)$$

Cada publicação visualizada em uma *timeline* é chamada de *impressão*. Seja N o número de novas impressões por visualização (i.e., por acesso do usuário). Uma impressão é considerada nova cada vez que é visualizada pela primeira vez. Seja A o número de publicações criadas pelas fontes no intervalo entre duas visualizações. A tabela 1 sumariza a notação usado ao longo deste trabalho.

Tabela 1. Tabela de notações (acessos e visualizações são sinônimos).

| variável | descrição |
|-----------|--|
| V | Valor do Acesso (VoA); neste trabalho, assumimos igual ao número de <i>novas</i> impressões por acesso |
| K | tamanho da <i>timeline</i> (número de impressões por acesso) |
| λ | taxa de criação de publicações (publicações por unidade de tempo) |
| A | quantidade de publicações criadas entre acessos |
| μ | taxa média de acessos |
| $1/\mu$ | intervalo médio entre acessos |
| ρ | quantidade média de publicações novas por acesso |

2.1. Quantificando o valor do acesso

A seguir, quantificamos a quantidade de informação nova recebida em cada visualização da *timeline*. Note que a quantidade de novas impressões serve como estimativa para a quantidade de informação recebida, que denotamos por *valor do acesso* (VoA). Daqui por diante, nos referimos indistintamente ao VoA e à quantidade esperada de novas impressões por visualização.

O modelo proposto a seguir produz a quantidade esperada de nova impressões por visualização (\bar{V}), em função do tamanho da *timeline* (K), da taxa de criação de publicações (λ) e da taxa de visualizações da *timeline* (μ).

O modelo analítico baseia-se nas seguintes premissas:

(P1) *Fluxo Poisson de geração de conteúdo*: o conteúdo é gerado pelas fontes seguindo um fluxo Poisson com taxa agregada λ .

(P2) *Novos conteúdos exibidos caso haja espaço* caso haja ao menos um conteúdo novo que ainda não tenha sido exibido, e o usuário tenha espaço na *timeline* para incluir

tal conteúdo, o conteúdo será exibido, i.e., o conteúdo irá gerar uma nova impressão. Note que o modelo é agnóstico com relação a ordem com que os conteúdos são exibidos na *timeline*.

(P3) *Conteúdos já exibidos só serão re-exibidos caso não existam novos* a única motivação para re-exibir um conteúdo é preencher o espaço da *timeline* caso não existam conteúdos novos.

Para calcular VoA, primeiramente condicionamos no tempo entre dois acessos. Seja τ uma constante que caracteriza o tempo entre acessos (por hora assumido como dado). Nesse caso, dadas as premissas do modelo, o número de novas impressões por acesso é dado por uma Poisson truncada,

$$E(V|\tau) = \left(\sum_{i=0}^{K-1} iP(V = i|\tau) \right) + KP(A \geq K|\tau) \quad (2)$$

Essa equação é decorrente do fato de que, dado que o intervalo entre visualizações teve duração igual a τ , um usuário pode receber no máximo K novas impressões na sua *timeline* de tamanho K . Ou seja:

$$P(V = i|\tau) = \begin{cases} P(A < i|\tau), & \text{se } i < K \\ P(A \geq i|\tau), & \text{se } i = K \\ 0, & \text{se } i > K \end{cases} \quad (3)$$

As três condições acima correspondem a situações nas quais (i) o conteúdo gerado é insuficiente para preencher a *timeline* e (ii) a *timeline* é preenchida só com novos conteúdos. A terceira e última condição indica que (iii) só há espaço para K conteúdos na *timeline*, i.e., a probabilidade de termos mais de K novidades por acesso é zero.

Note que a análise realizada até então ainda não faz uso da premissa de Poisson. Fazendo uso da premissa de Poisson (P1), obtemos a seguinte expressão a partir de (2):

$$E(V|\tau) = \left(\sum_{i=0}^{K-1} ie^{-\lambda\tau} \frac{(\lambda\tau)^i}{i!} \right) + Ke^{-\lambda\tau} \frac{(\lambda\tau)^K}{K!} + KP(A > K|\tau) \quad (4)$$

$$= \left(\sum_{i=0}^K ie^{-\lambda\tau} \frac{(\lambda\tau)^i}{i!} \right) + K(1 - P(A \leq K|\tau)) \quad (5)$$

$$= \left(\sum_{i=0}^K ie^{-\lambda\tau} \frac{(\lambda\tau)^i}{i!} \right) + K \left(1 - \sum_{i=0}^K e^{-\lambda\tau} \frac{(\lambda\tau)^i}{i!} \right) \quad (6)$$

$$= K - \sum_{i=0}^K (K - i)e^{-\lambda\tau} \frac{(\lambda\tau)^i}{i!} \quad (7)$$

Finalmente, aplicando-se o teorema de probabilidades totais temos uma fórmula para o cálculo de \bar{V} :

$$\bar{V} = \int_{\tau=0}^{\infty} E(V|\tau)f(\tau)d\tau \quad (8)$$

A seguir, consideramos casos especiais da expressão acima, em função da distribuição do tempo entre acessos dos usuários.

2.2. Caso especial I: intervalo entre visualizações determinístico

No caso em que o tempo entre acessos é determinístico, o número de *posts* criados entre dois acessos é distribuído de acordo com uma distribuição Poisson com média $\tau\lambda$. Temos então que

$$\bar{V} = E(V|\tau) \quad (9)$$

A expressão de \bar{V} é dada por (7). Embora a expressão de \bar{V} seja simples, ela não admite fórmula fechada sem uso de somatórios.

2.3. Caso especial II: intervalo entre visualizações exponencialmente distribuído

A seguir, consideramos o tempo entre acessos exponencialmente distribuído. Seja $f(t)$ uma variável aleatória exponencialmente distribuída com média $1/\mu$, i.e., $f(t) = \mu e^{-\mu t}$. Substituindo-se $f(t)$ em (8):

$$\bar{V} = \int_{\tau=0}^{\infty} \left(K - \sum_{i=0}^K (K-i) e^{-\lambda\tau} \frac{(\lambda\tau)^i}{i!} \right) \mu e^{-\mu\tau} d\tau. \quad (10)$$

A expressão acima é passível de simplificações (vide Apêndice A), de tal forma que obtemos uma expressão final bem compacta para \bar{V} ,

$$\bar{V} = \frac{\lambda}{\mu} \left(1 - \left(\frac{\lambda}{\lambda + \mu} \right)^K \right). \quad (11)$$

Interpretação probabilística Como o intervalo entre publicações e o intervalo entre visualizações são exponencialmente distribuídos com taxas iguais a λ e μ , a razão $\lambda/(\lambda + \mu)$ é igual a probabilidade de haver uma nova publicação antes do usuário visualizar novamente sua timeline [Durrett 2012]. Então, pela falta de memória da distribuição exponencial, $(\lambda/(\lambda + \mu))^K$ é a probabilidade de haver K publicações entre duas visualizações. Neste último caso, uma publicação é perdida, ou seja, não vira uma impressão.

Seja $p = 1 - (\lambda/(\lambda + \mu))^K$ a probabilidade de o usuário visualizar a sua timeline antes da chegada de K publicações consecutivas. Nesse caso, todas as publicações viram impressões. Quando $q = 1$, todas as publicações viram impressões e teremos, na média, λ/μ novas impressões por visualização.

Pode-se verificar nesse caso que os seguintes limites se aplicam:

Tamanho da *timeline* $\lim_{K \rightarrow \infty} \bar{V} = \lambda/\mu$. Nesse regime, o usuário consegue ler todas as novidades criadas pelas fontes entre dois acessos, e o número médio de novidades por acesso é igual a λ/μ ;

Geração de conteúdo $\lim_{\lambda \rightarrow \infty} \bar{V} = K$ e $\lim_{\lambda \rightarrow 0} \bar{V} = 0$ Nesse regime, as fontes geram conteúdo a taxa alta (resp., baixa), e o número de novidades por visualização é máximo (resp., mínimo), e igual a K (resp., 0);

Frequência de acesso $\lim_{\mu \rightarrow 0} \bar{V} = K$ e $\lim_{\mu \rightarrow \infty} \bar{V} = 0$ Nesse regime, os usuários acessam o sistema com frequência baixa (resp., alta), e o número de novidades por visualização é máximo (resp., mínimo), e igual a K (resp., 0).

Tabela 2. Tabela das páginas seguidas

| Página | Endereço |
|-----------------------|---|
| Jair Bolsonaro | https://www.facebook.com/jairmessias.bolsonaro/ |
| PSL | https://www.facebook.com/PartidoSocialLiberalBR/ |
| Lula | https://www.facebook.com/Lula/ |
| Fernando Haddad | https://www.facebook.com/fernandohaddad/ |
| PT | https://www.facebook.com/pt.brasil/ |
| Geraldo Alckmin | https://www.facebook.com/geraldoalckmin/ |
| PSDB | https://www.facebook.com/Rede45/ |
| Marina Silva | https://www.facebook.com/marinasilva.oficial/ |
| Rede Sustentabilidade | https://www.facebook.com/RedeSustentabilidade18/ |
| Ciro Gomes | https://www.facebook.com/cirogomesoficial/?ref=br_rs |
| PDT | https://www.facebook.com/pdt.org.br/ |
| Henrique Meirelles | https://www.facebook.com/hmeirellesoficial/ |
| MDB | https://www.facebook.com/MDBNacional15/ |
| Álvaro Dias | https://www.facebook.com/ad.alvarodias/ |
| Podemos | https://www.facebook.com/podemos19/ |
| Guilherme Boulos | https://www.facebook.com/guilhermeboulos.oficial/ |
| PSOL | https://www.facebook.com/psol50/ |
| Cabo Daciolo | https://www.facebook.com/depudadocabodaciolo/ |
| Patriotas | https://www.facebook.com/Patriota51Oficial/ |
| João Amoës | https://www.facebook.com/JoaoAmoedoNOVO/ |
| Partido Novo | https://www.facebook.com/NOVO30/ |

3. Parametrização do Modelo, Simulação e Validação

O modelo analítico proposto foi validado utilizando-se as publicações das páginas dos candidatos à Presidência da República do Brasil no Facebook, bem como as de seus respectivos partidos políticos, durante o primeiro turno da campanha eleitoral das eleições presidenciais de 2018. Os dados foram obtidos através da API do Facebook analisando-se apenas as publicações realizadas entre os dias 27 de setembro de 2018 e 10 de outubro de 2018. As páginas, e os seus respectivos endereços encontram-se sumarizadas na Tabela 2.

A Figura 1(a) mostra o total de publicações por dia durante o período dos experimentos de todas as páginas selecionadas. O total de publicações variou entre 30 e 193 publicações por dia. A página *CiroGomesOficial* foi a que mais publicou e a página *Patriota51Oficial* foi a que produziu menos conteúdo. A quantidade de publicações oscilou entre o mínimo de 4 e o máximo de 222 publicações por página.

Após a seleção das fontes, o comportamento de uma timeline FIFO [Hargreaves et al. 2018] de tamanho $K = 10$ foi simulada utilizando-se o universo das publicações coletadas. Cada *fotografia* é uma instanciãção de uma *timeline* FIFO de tamanho K . Uma fotografia consiste na seleção de uma publicação *referência* (escolhida de forma uniforme e aleatória dentre as publicações disponíveis) e das $K - 1 = 9$ publicações imediatamente anteriores, a partir do universo de publicações coletadas.

Dado o intervalo entre amostras desejado, $1/\mu$, medido em horas, determinamos o número de fotografias que iremos tirar do sistema. Num período de 14 dias, tiramos $24 \times 14/\mu$ fotografias. As fotografias são então ordenadas no tempo, tomando como base as fotografias de referência, e a partir desta lista de fotografias calculamos o VoA.

Nos experimentos considerados, $1/\mu$ variou entre 1 hora e 24 horas. Para cada

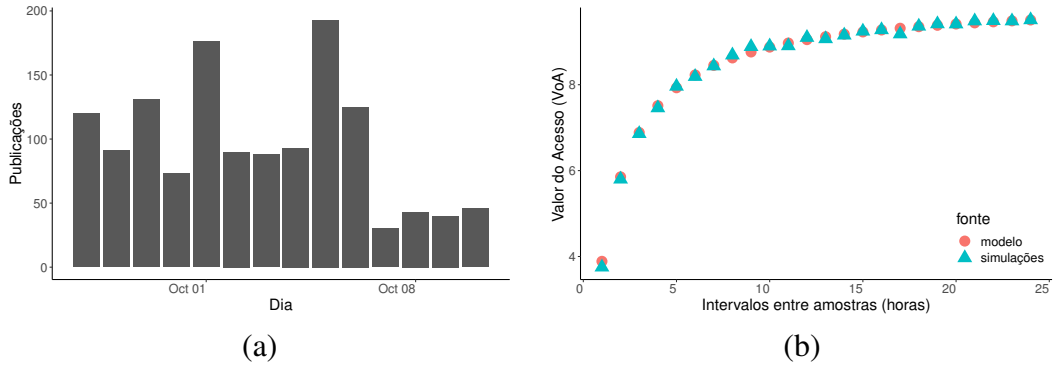


Figura 1. (a) Quantidade total de publicações por dia das páginas selecionadas e (b) parametrização e validação do modelo com dados do Facebook

intervalo foram realizadas 30 rodadas de simulação. Os resultados encontram-se na Figura 1(b). Os círculos em vermelho são os dados previstos pelo modelo (eq. (11), com $\lambda = 4,487$ e $K = 10$) e os triângulos azuis são as médias dos valores obtidos nas simulações. A pequena diferença entre os dados previstos pelo modelo e os dados obtidos via simulação indicam a precisão do modelo proposto. Note que na medida em que $1/\mu$ aumenta, o VoA tende à assíntota $K = 10$, ou seja, a *timeline* será preenchida completamente com informações novas, com alta probabilidade, e os ganhos marginais são cada vez menores.

4. Determinando a Taxa de Acesso Ótima dos Usuários

Nesta seção introduzimos o problema da taxa ótima de acesso às redes sociais. Consideramos usuários que possuem um custo para acessar suas redes sociais. Tal custo advém do fato de que usuário deixam de fazer outras atividades para se dedicar a estes acessos, gastam energia dos seus equipamentos, e incorrem custos monetários para cobrir seus planos de dados. Seja c o custo incorrido pelos usuários por acesso.

A *utilidade* por unidade de tempo experimentada por um usuário é dada pelo número esperado de impressões novas por visualização (\bar{V}), vezes a taxa de visualizações por unidade de tempo (μ), menos o custo $c\mu$ necessário para a recuperação destas impressões. Temos então que,

$$U(\bar{V}, \mu) = \mu\bar{V} - c\mu \quad (12)$$

Substituindo-se a equação (11) na equação (12), podemos reescrever a função de utilidade em termos λ e K :

$$U(\mu) = \lambda \left(1 - \left(\frac{\lambda}{\lambda + \mu} \right)^K \right) - c\mu. \quad (13)$$

Portanto, dados λ e K , o problema de determinação da taxa ótima de acesso às redes sociais consiste em maximizar a utilidade ajustando a taxa de visualização μ ,

$$\max U(\mu), \quad \mu \geq 0, \quad (14)$$

onde $U(\mu)$ é dada pela equação (13). A utilidade é maximizada quando sua derivada parcial com relação à variável de controle é igual a zero, i.e., quando $\partial U/\partial \mu = 0$. A derivada parcial $\partial U/\partial \mu$ é dada por

$$\frac{\partial U}{\partial \mu} = K \left(\frac{\lambda}{\lambda + \mu} \right)^{K+1} - c \quad (15)$$

Fazendo $\partial U / \partial \mu = 0$ obtemos a taxa de acesso ótima, μ^* ,

$$\mu^* = \lambda \left(\frac{\sqrt[K+1]{K}}{\sqrt[K+1]{c}} - 1 \right). \quad (16)$$

A seguir, avaliamos numericamente o comportamento da taxa de acesso ótima, em função dos diferentes parâmetros do sistema.

5. Exemplos Numéricos

Nesta seção apresentamos exemplos numéricos obtidos a partir do modelo proposto. Nossos objetivos são (i) realizar análise de sensibilidade das métricas de interesse dados os parâmetros do modelo (proposto na Seção 2) e (ii) ilustrar a sensibilidade das soluções do problema da taxa ótima de acesso (introduzido Seção 4) em função dos parâmetros.

A seguir, avaliamos o incremento marginal de VoA resultante do usuário fazer o *scroll down* de sua tela para capturar mais impressões por acesso, i.e., avaliamos o impacto do aumento de K no VoA. A Figura 2(a) mostra este impacto para um valor fixo de ρ . No limite quando $K \rightarrow \infty$, $\bar{V}(\rho, K) = \rho$. Note também que em todos os cenários considerados a convergência para valores próximos do assintótico se dá para $K \leq 100$.

Por exemplo, quando $\rho = 10$, isto é, a taxa de acesso é 10 vezes menor que a taxa de publicação, $\bar{V} \approx \rho$ para $K \geq 40$. Desta forma, mesmo que um usuário acesse a rede de forma infrequente, à taxa $\mu = \lambda/10$, ele ainda assim pode absorver quase todas as novidades, desde que possa recuperar $K = 40$ impressões por acesso.

Quando os usuários capturam K impressões por acesso, o máximo VoA alcançável é K . Como o VoA experimentado pelos usuários se compara esse valor?

Na curva pontilhada inclinada na Figura 2(a) temos o VoA correspondente à situação na qual todas as publicações recebidas são novas. A Figura 2(a) mostra que mesmo para pequenos valores de K , ex., $K = 10$, o VoA experimentado já se distancia de K . O distanciamento é causado tanto por flutuações na quantidade de publicações entre acessos quanto pelas variações do intervalo entre acessos. *Note que se os acessos e as publicações forem determinísticos e sincronizados, as curvas pontilhadas e cheias irão sobrepor-se. Entretanto, publicações surgem de forma orgânica, em função de fatores exógenos, e lidar com a imprevisibilidade é um aspecto intrínseco do problema.*

A Figura 2(b) mostra o aumento de VoA em função de ρ . Quando $\rho \rightarrow \infty$, $\bar{V}(K, \rho) = K$. Note que quanto maior o valor de K , maior é o valor de ρ necessário para se alcançar a assíntota $\bar{V} = K$. De fato, quanto maior for K , mais ativas precisam ser as fontes para se preencher com novidades as *timelines* dos usuários.

A Figura 3(a) mostra a utilidade experimentada pelo usuário por unidade de tempo, em função da taxa de acesso μ , para cinco valores de K . Nesta figura, λ foi parametrizado com os dados retornados pela API do Facebook conforme descrito na Seção 3. O custo c é fixo igual a 1. Esta figura mostra que, dados os demais parâmetros constantes, quanto maior o valor de K , menor é a taxa ótima de acesso. Por exemplo, quando $K = 20$, $\mu^* = 0.688$ acessos/hora e quando $K = 2$, $\mu^* = 1.1663$ acessos/hora. Para um dado par (λ, μ) , quanto maior o valor de K , maior é a utilidade. Isto ocorre porque assumimos em nosso modelo que o custo de acesso é independente de K . Funções de custo mais complexas são objeto de trabalhos futuros.

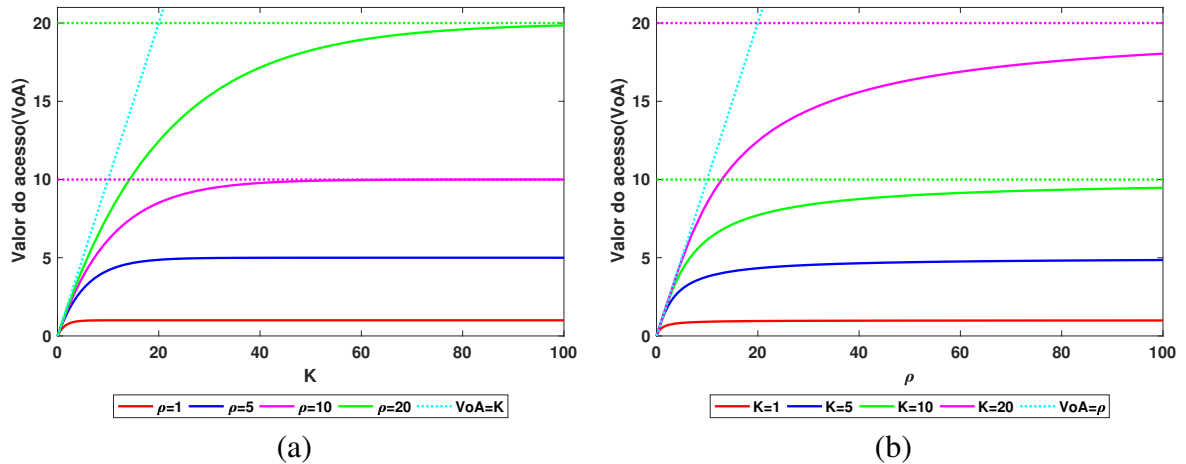


Figura 2. Análise de sensibilidade: impacto de diferentes parâmetros (K, ρ) no valor de um acesso (VoA).

A Figura 3(b) mostra a utilidade experimentada pelo usuário por unidade de tempo em função da taxa de acesso μ para quatro valores de λ (com $K = 10$). Neste caso, quanto maior λ , maior é a taxa de acesso ótima. Fixando-se um valor para o par (K, μ) , a utilidade aumenta com o aumento de λ . De fato, fontes mais ativas irão favorecer um aumento de VoA, lembrando que tal valor é limitado por K .

A Figura 3(c) mostra a taxa de acesso ótima (μ^*) em função de K . Note que, para valores de K próximos a um, na medida em que K aumenta, μ^* aumenta, pois pode-se preencher a *timeline* com mais novidades. Entretanto, um aumento adicional de K gera uma diminuição de μ^* pois a probabilidade de haver conteúdos repetidos aumenta.

A Figura 3(d) mostra a taxa de acesso ótima (μ^*) em função de λ . Essa figura ilustra o comportamento linear de μ^* em função de λ (vide (16)). Fixando-se λ , μ^* diminui com o aumento de K , i.e., *se o usuário absorver mais publicações por acesso, ele pode acessar a rede com menos frequência*.

Impactos práticos no design de redes sociais: Os exemplos numéricos contidos nesta seção ilustram que o VoA poder ser limitado ora por ρ ora por K . Temos então dois regimes de operação. Consideremos o regime em que $\bar{V} \approx \rho$. Neste regime, assumimos que K é dado e fixo, e que o aumento de K não resulta em aumento de \bar{V} . Se o usuário for capaz de processar K novas impressões por acesso, ele não precisa fazer uso dos filtros das redes sociais. Já no regime em que $\bar{V} \approx K$ o usuário potencialmente está perdendo muitas publicações novas. Ele precisa da ajuda do sistema de recomendação da rede social para filtrar as publicações mais relevantes. A compreensão desses regimes auxilia no projeto das *timelines*, caracterizando as faixas nas quais algoritmos de recomendação como o News Feed do Facebook passam a ser necessários para agregar valor aos usuários.

6. Trabalhos Relacionados

A seguir, dividimos os trabalhos relacionados em três categorias, referentes aos impactos da frequência com que usuários acessam redes sociais. Em seguida, apresentamos também um breve apanhado sobre modelagem de linhas do tempo (*timelines*). *Não é de nosso conhecimento nenhum trabalho que tenha proposto modelos para estimar a frequência ótima de acesso de usuários a redes sociais, e que tenha parametrizado tais modelos com dados reais.*

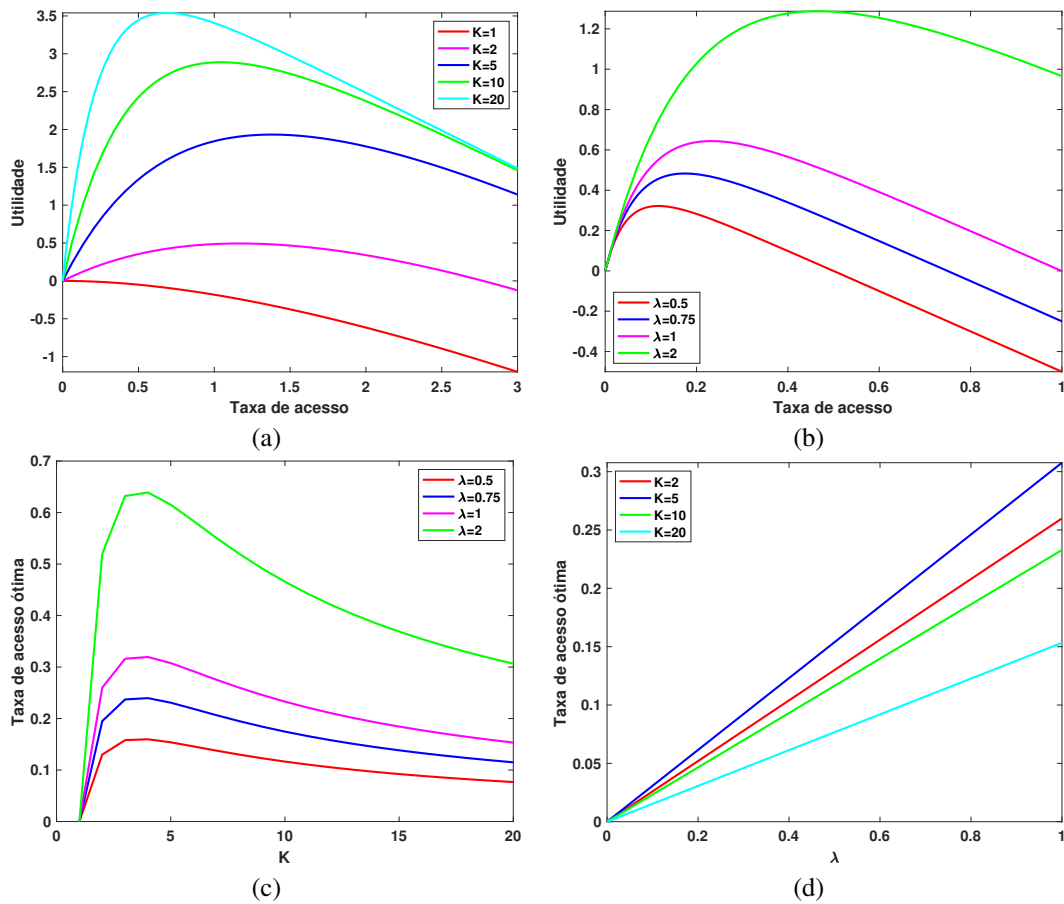


Figura 3. Análise de sensibilidade: impacto de diferentes parâmetros (K, λ, μ) na utilidade dos usuários ((a) e (b)) e na taxa ótima de acesso ((c) e (d)).

Psicologia e vício A taxa adequada para usuários acessarem uma rede social é um tema que vem sendo amplamente discutido na esfera da psicologia [Salehan and Negahban 2013]. Segundo Sean Parker, fundador do Napster e ex-presidente do Facebook, *o Facebook é um sistema projetado para explorar as vulnerabilidades dos seres humanos* [Solon 2017]. Segundo Tristan Harris, ex-funcionário do Google, *artefatos tecnológicos podem sequestrar a mente de seus usuários* [Harris 2016]. Embora nesse trabalho nosso enfoque seja exclusivamente relacionado às questões técnicas relativas à quantidade de novidade que um usuário obtém cada vez que acessa uma rede social, as implicações psicológicas deste trabalho são evidentes. Em particular, se o usuário souber de antemão que irá receber pouca novidade ao acessar a rede em determinado momento, ele terá mais incentivos para postergar seus acessos. Alternativamente, se um usuário aprender, em retrospectiva, que seus acessos geraram poucas novidades, ele poderá melhor planejar seus acessos futuros.

Economia da atenção e anúncios A taxa com que usuários acessam as redes sociais tem implicações imediatas em aspectos econômicos, tendo em vista que as propagandas geram lucros em função do *click rate*. O *click rate*, por sua vez, é fortemente influenciado pela taxa com que usuários acessam a rede social. Se os anunciantes tiverem acesso a informações sobre a taxa recomendada aos usuários, para acessar a rede social, ao longo do dia, estes podem melhor decidir quando pagar para que determinadas propagandas sejam exibidas, e.g., dependendo do público alvo [Rendle 2012, Jiang et al. 2013].

Infraestrutura e idade da informação (AoI) A taxa com que usuários acessam as redes sociais tem impactos diretos no tráfego que passa pela rede. Tendo em vista que as redes sociais são atualmente uma das maiores fontes de tráfego, recomendações sobre a taxa com que usuários devem acessar suas páginas pode ter efeitos em cascata na rede como um todo. A área de idade da informação (age of information, ou AoI) tem ganhando bastante visibilidade na comunidade científica, e vislumbramos que as contribuições deste trabalho são relevantes neste novo domínio [Kosta et al. 2017].

Modelagem e avaliação de desempenho de timelines A *News Feed* do Facebook foi modelada como uma fila FIFO em [Hargreaves et al. 2018]. Em [Hargreaves et al. 2019] foi mostrado que timelines FIFO são um caso particular de timelines do TTL (*time-to-live*) sendo que estas são mais versáteis e flexíveis por permite o reordenamento das publicações. O presente trabalho adiciona a taxa de visualizações dos usuários ao modelo, criando novas perspectivas de análises e possibilitando a formulação do problema da taxa de acesso ótima dos usuários .

7. Discussão

As ideias apresentadas neste trabalho podem ser implementadas numa camada intermediária entre o usuário e a rede social, e.g., por meio de um aplicativo que consuma a *News Feed* do Facebook e aprenda os parâmetros da eq. (16) via aprendizado por máquina. Plataformas como Gobo [MIT 2018] já atuam como um *middleware* entre o Facebook e os usuários, e podem servir de base para implementar as ideias aqui expostas.

Os resultados contidos neste trabalho não consideram o impacto dos algoritmos de personalização das redes sociais online, como o *News Feed* do Facebook, no VoA. Em trabalhos futuros, pretendemos quantificar o VoA por perfil de usuário e por rede social, e.g., usando as plataformas mencionadas acima. A partir de tal caracterização do VoA por perfil e por rede, vislumbramos ser possível usar o VoA como critério de comparação entre perfis e redes sociais.

8. Conclusões

Redes sociais estão cada vez mais presentes no cotidiano de milhões de usuários da Internet, e uma parcela significativa da população mundial acessa redes sociais inúmeras vezes ao dia [Kemp 2018]. A cada acesso à rede social, seus usuários experimentam um determinado *valor*, sujeito a aspectos subjetivos e difíceis de mensurar. Dessa forma, mensurar o *valor do acesso* (VoA) é desafiador. Neste trabalho, contornamos este desafio propondo uma abordagem simples porém quantificável para estimar o VoA experimentado por um usuário de uma rede social. Em particular, assumimos que tal valor é igual a quantidade de impressões novas apresentadas pela rede social ao usuário. Usando tal abordagem, propusemos um modelo analítico para quantificar o VoA. Em seguida, parametrizamos o modelo usando dados reais, e indicamos como o modelo pode ser usado para obter a taxa de acesso ótima quando há custos associados à obtenção das impressões. Acreditamos que este trabalho seja mais um passo numa agenda de pesquisa que visa desenvolver plataformas de redes sociais respeitando as preferências do usuários. Esta agenda necessariamente envolve contemplar aspectos psicológicos, econômicos e de infraestrutura inerentes ao problema, que pretendemos atacar em trabalhos futuros.

Agradecimentos Este projeto contou com recursos do projeto THANES (INRIA/França, FAPERJ/UFRJ/Brasil), do CNPq e da Petrobras.

Referências

- Ceres, P. (2018). Screen time controls. *Wired*. <http://tinyurl.com/ioswired>.
- Durrett, R. (2012). *Essentials of Stochastic Processes*, volume 7 of *Springer Texts in Statistics*. Springer New York, New York, NY.
- Facebook (2019). About ad auctions. <http://bit.ly/2OulPUE>.
- 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*, 129:15–39.
- Hargreaves, E., Menasché, D., Neglia, G., and Agosti, C. (2018). Visibilidade no facebook: Modelos, medições e implicações. In *BraSNAM*, volume 7. SBC.
- Harris, T. (2016). How technology hijacks people’s minds: from a magician and Google’s design ethicist. <http://www.tristanharris.com/essays/>.
- Jiang, B., Hegde, N., Massoulié, L., and Towsley, D. (2013). How to optimally allocate your budget of attention in social networks. In *2013 Proceedings of IEEE INFOCOM*.
- Kemp, S. (2018). Digital in 2018. *Wired*. <http://bit.ly/2uChFUm>.
- Kosta, A., Pappas, N., Angelakis, V., et al. (2017). Age of information: A new concept, metric, and tool. *Foundations and Trends® in Networking*, 12(3):162–259.
- MIT (2018). Gobo – your social media, your rules. <https://gobo.social/>.
- Rendle, S. (2012). Social network and click-through prediction with factorization machines. In *KDD-Cup Workshop*, page 113.
- Salehan, M. and Negahban, A. (2013). Social networking on smartphones: When mobile phones become addictive. *Computers in Human Behavior*, 29(6):2632–2639.
- Shapiro, C. and Varian, H. (1998). *Information Rules: A Strategic Guide to the Network Economy*. Harvard Business School Press.
- Simon, H. A. (1971). Designing organizations for an information rich world. In Greenberger, M., editor, *Computers, communications, and the public interest*. Baltimore.
- Solon, O. (2017). Ex-Facebook president Sean Parker: site made to exploit human ‘vulnerability’. *The Guardian*. <http://bit.ly/2HJTufh>.
- Twitter (2019). Bidding and auctions. <http://bit.ly/2JXj6H8>.

A. Expressão de \bar{V} no modelo exponencial

A partir de (7) e (8) temos que

$$\bar{V} = K - \sum_{i=0}^K \frac{(K-i)}{i!} \lambda^i \mu \int_{\tau=0}^{\infty} e^{-(\lambda+\mu)\tau} \tau^i d\tau \quad (17)$$

A solução da integral pode ser encontrada em tabelas padrões de integração. Temos,

$$\bar{V} = K - \frac{\mu}{\lambda + \mu} \left(K \sum_{i=0}^K \frac{\lambda^i}{(\lambda + \mu)^i} - \sum_{i=0}^K \frac{i \lambda^i}{(\lambda + \mu)^i} \right). \quad (18)$$

Após manipulações algébricas, obtemos uma expressão simples para \bar{V} (vide (11)).