

Analisando Métricas Olímpicas – Exercício de Métodos Digitais a Partir do Desempenho de Três Medalhistas Brasileiras nas Redes Sociais¹

Márcio Carneiro dos Santos²
Universidade Federal do Maranhão, São Luís, MA.

RESUMO

Descreve-se um experimento de análise exploratória de dados a partir de métricas extraídas das redes sociais de três medalhistas olímpicas brasileiras, a saber, Rayssa Leal, Mayra Aguiar e Rebeca Andrade. Foram utilizadas as ferramentas Twarc e Instaloader, bem como códigos personalizados em linguagem Python para organizar e gerar visualizações, dentro da abordagem dos Métodos Digitais. Nesta outra competição, idealizada apenas para fins exploratórios e acadêmicos, a atleta de Imperatriz, segunda maior cidade do Maranhão, ganhou o ouro e em todas as provas. O exercício teve o principal objetivo de apresentar um encaminhamento metodológico viável em situações onde existe grande volume de dados e dificuldade para operar sem a ajuda de recursos computacionais intensivos.

PALAVRAS-CHAVE: Twitter, Instagram, Métodos Digitais.

Introdução

Para qualquer atleta, participar de uma olimpíada já é um momento mais que especial. Ganhar uma medalha olímpica então pode transformar contos de fada em realidade. Além do feito que fica para a história, há também, naturalmente, um direcionamento da atenção do público para aquela pessoa, potencializado pelo olhar da mídia em geral o que, inevitavelmente, também se reflete nas redes sociais.

Como pesquisador da Comunicação essa é uma oportunidade de exercitar a abordagem dos Métodos Digitais para fazer uma EDA, do inglês Exploratory Data Analysis, uma aproximação inicial que na Data Science fazemos toda vez que nos deparamos com um conjunto novo de dados.

¹ Trabalho apresentado no GP de Conteúdos Digitais e Convergências Tecnológicas, XXI Encontro dos Grupos de Pesquisas em Comunicação, evento componente do 44º Congresso Brasileiro de Ciências da Comunicação.

² Professor do PPGCOMPro/UFMA, doutor pelo programa TIDD da PUC-SP, e-mail: márcio.carneiro@ufma.br

Numa EDA, coletamos, organizamos e tentamos identificar padrões, através do auxílio de métricas e das visualizações com gráficos, que nos ajudam a entender melhor o que aquela massa de informações nos traz, servindo principalmente para geração de insights e indicação de questões para aprofundamento. O uso dos Métodos Digitais, diante de muitos objetos de estudo da Comunicação, torna-se importante, principalmente em situações de volume, variedade e velocidade na geração dos dados, nas quais abordagens manuais e feitas sem o auxílio de recursos computacionais mostram-se pouco efetivas.

Na metodologia foram aplicadas as ferramentas, Twarc e Instaloder, para coletar dados do Twitter e do Instagram, respectivamente, e daí criar um pódio comparativo do desempenho nestas redes de 3 medalhistas olímpicas brasileiras, todas com vitórias recentes, a saber, Rayssa Leal, prata no skate, Mayra Aguiar, bronze no judô e Rebeca Andrade, prata na ginástica. Nesta outra competição, idealizada apenas para fins exploratórios e acadêmicos, a atleta de Imperatriz, segunda maior cidade do Maranhão, ganhou o ouro e em todas as provas.

Pódio Twitter

No caso do Twitter, foi utilizada uma métrica bem simples: a contagem dos posts mencionando cada atleta, coletados por minuto, através da ferramenta Twarc, que na sua versão 2, traz a função *counts*.

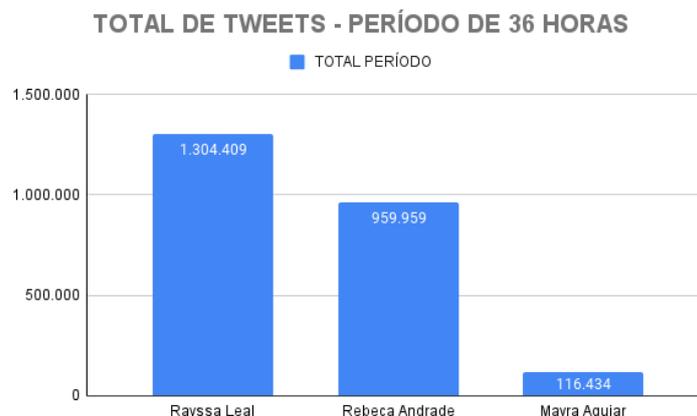


Figura 1 - Contagem de Tweets num período de 36 horas (12 antes e 24 depois da competição que gerou a medalha) – Fonte: Do autor.

Para busca configurei um período médio de 36 horas para cada atleta, sendo 12 horas antes do horário marcado para a competição e 24 horas após este horário, de forma a tentar entender a repercussão da vitória na plataforma, basicamente extraíndo o total de publicações, a média por minuto e o pico de posts.

Como se pode ver na figura 1, Rayssa Leal ficou com o ouro, com mais de um milhão e trezentos mil posts no período. Dividindo esse total por 2160, que é o número de minutos em 36 horas (36 x 60) achamos a média geral da performance, figura 2, que refletiu os resultados brutos.

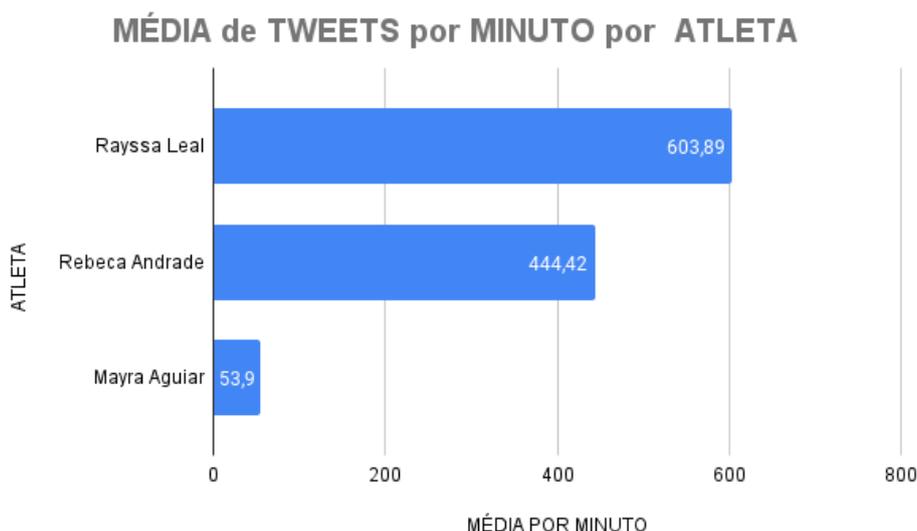


Figura 2 – Média de Tweets por Minuto / Atleta – Fonte: Do autor

Os picos máximos, ou seja, os maiores números de posts em um único minuto podem ser vistos abaixo, figura 3.

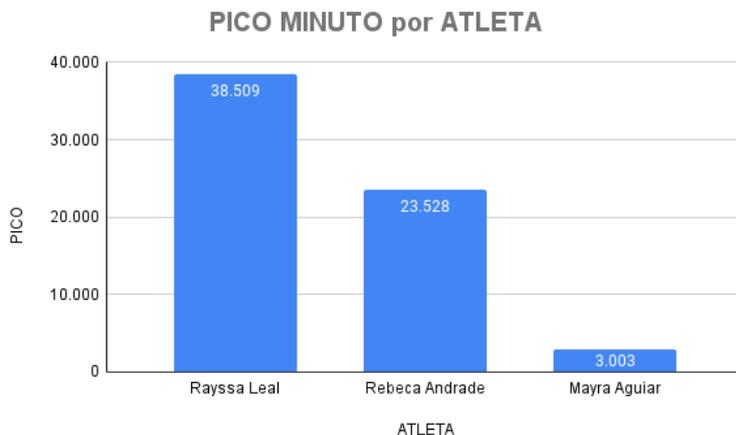


Figura 3 – Picos Máximos de Posts - Fonte: Do autor

No caso do ouro no Twitter , ganho pela Rayssa Leal, é possível ver o horário exato em que ocorreu, segundo os registros da plataforma. É importante observar que a API do Twitter trabalha com o horário em UCT e que para convertermos isso para o horário de Brasília (UCT-3) precisamos diminuir 3 horas do tempo informado. Assim, podemos ver que, à 01:32 da manhã do dia 26 , a fadinha teve 38.509 posts, isso num único minuto, figura 4.

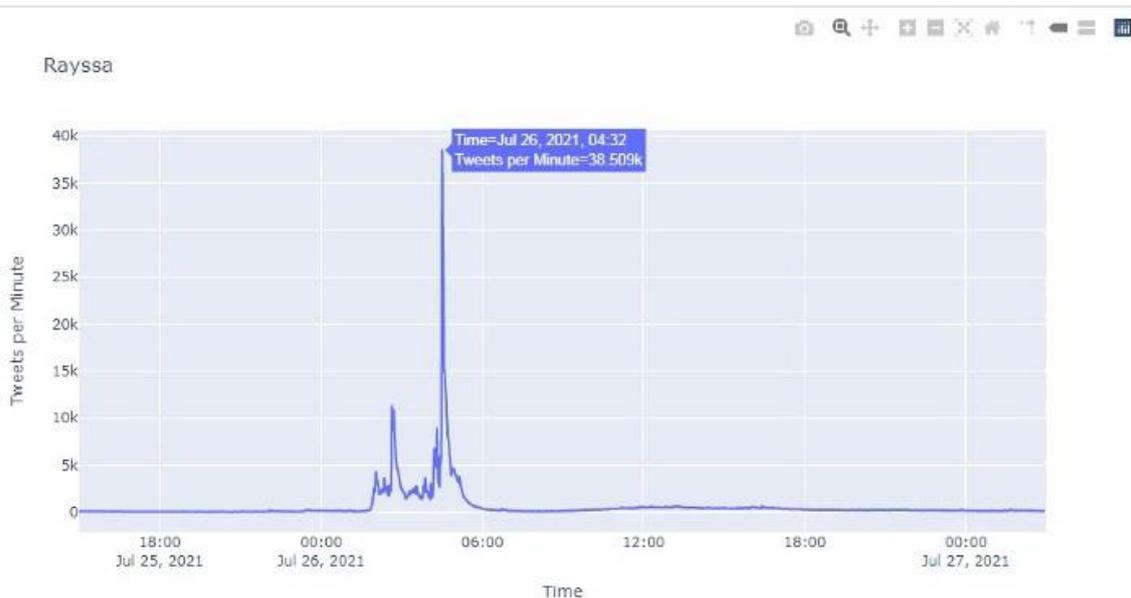


Figura 4 – Horário dos Picos Máximos de Posts - Fonte: Do autor

Deve-se considerar também que o horário das transmissões das provas pode ter afetado a performance neste quesito. No caso da Mayra, que ganhou mais de madrugada, observou-se um padrão com dois picos, um durante a competição e outro de dia, depois da ampla disseminação da informação. A competição bem sucedida da Rebeca, que foi transmitida na manhã seguinte, deve ter concorrido também para uma redução dos resultados da atleta do Judô, figura 5.

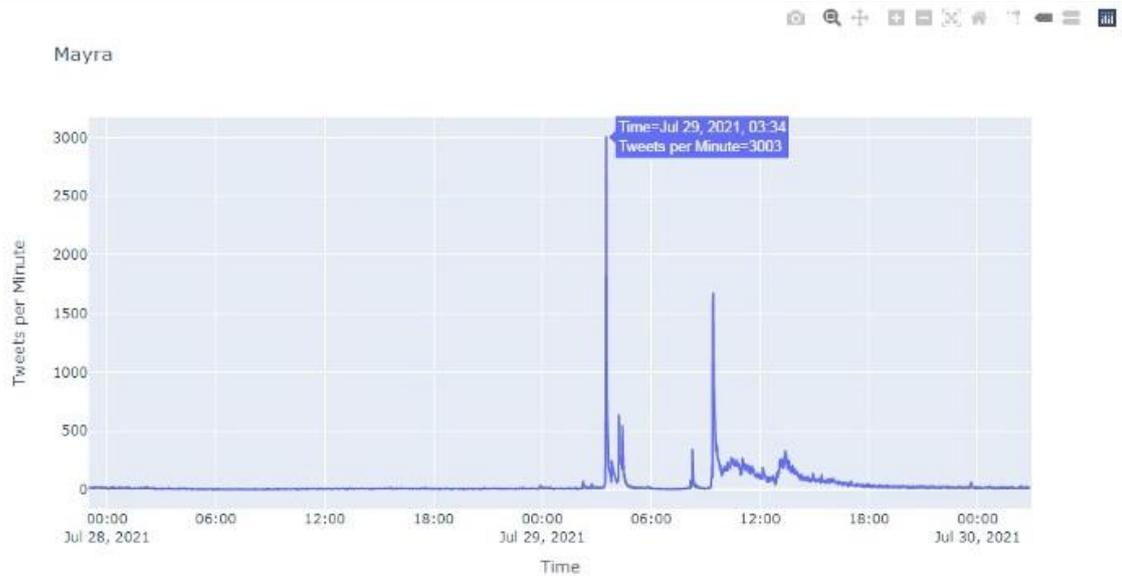


Figura 5 – Horário dos Picos Máximos de Posts - Fonte: Do autor

Pódio Instagram

Já no Instagram há mais dificuldades. A plataforma não gosta muito de fornecer seus dados e restringe bastante principalmente tentativas repetitivas de coleta. Por isso, a partir do Instaloader, foi construído um código em Python, o LABCOM_Insta que coleta informações dos perfis, como bio e número de seguidores. É pouco para associar com a vitória na Olimpíada mas , de qualquer forma, apresenta-se aqui apenas a título de demonstração da possibilidade.

Pelo noticiário fomos informados do extraordinário crescimento dos números nas redes sociais da Rayssa, fato que, de forma bem provável, também deve ter acontecido com as outras atletas, ainda que em escala diferente. Os números totais de seguidores, hoje (dia 31/07/2021) dos perfis estão abaixo. Rayssa já tem mais de 6 milhões e trezentos mil, contra cerca de um milhão e quatrocentos da Rebeca , segunda colocada, num aumento exponencial após a medalha, figura 6.

```
In [18]: segui
Out[18]: Perfis
rayssalealsk8      6334071
rebecarandrade    1408621
mayraaguiarjudo   179867
Name: Seguidores, dtype: int64
```

```
In [16]: df1['Seguidores'].plot.barh()
plt.title('Perfis por Número de Seguidores')
plt.xlabel('Seguidores')

Out[16]: Text(0.5, 0, 'Seguidores')
```

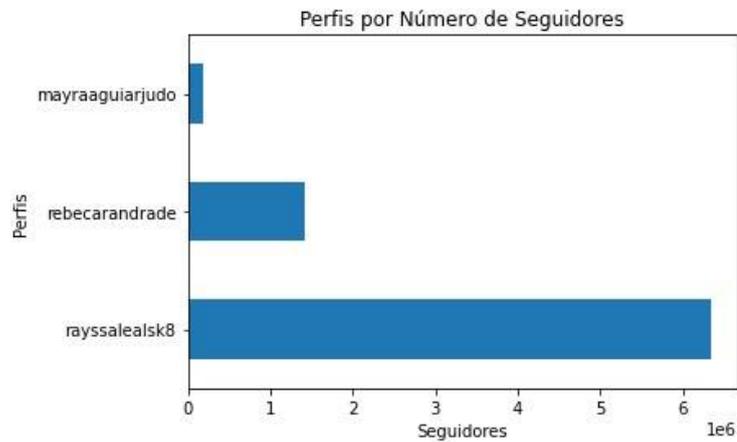


Figura 6 – Perfis por número de seguidores - Fonte: Do autor

Do pouco que é possível extrair, foi calculada uma métrica que, em tese, nos ajudaria a imaginar a efetividade de cada perfil para conseguir seguidores a partir do total de publicações que tem, ou seja, dividindo o total de seguidores pelo número de posts daquele perfil, figura 7.

```
In [13]: df1=df.sort_values(by='SegPorPosts',ascending=False)
In [14]: df1
```

Out[14]:

Perfis	Posts	IGTV	Seguidores	Seguindo	Bio	SegPorPosts
rayssalealsk8	378	6	6334071	788	• 13 years exin • Atletain@nikesb @monsterenergy @mrv @bancodobrasil in• rayssa.info@gmail.comin• ▲Acc. managed by mom	16756.80
rebecarandrade	873	6	1409621	1127	Olympic Silver Medalist AAinTokyo2020inm; Rebe.andrade@yahoo.cominPatrocínio/ Publi rp@felipepaulino.com inNão coloque limites em seus sonhos, coloque fe★	1613.54
mayraaguiajrudo	673	7	179967	1981	Judoca inee Tri-Medalhistas Olímpica in B-Campeã Mundialin; Sogipa - Bolsa Atleta - Adidas - Avon - XP Investimentos - #TeamCocaCola Athlete	267.28

```
In [15]: df1['SegPorPosts'].plot.barh()
plt.title('Perfis com Melhor Relação Seguidores por Posts')
plt.xlabel('Seguidores por Posts')
```

```
Out[15]: Text(0.5, 0, 'Seguidores por Posts')
```

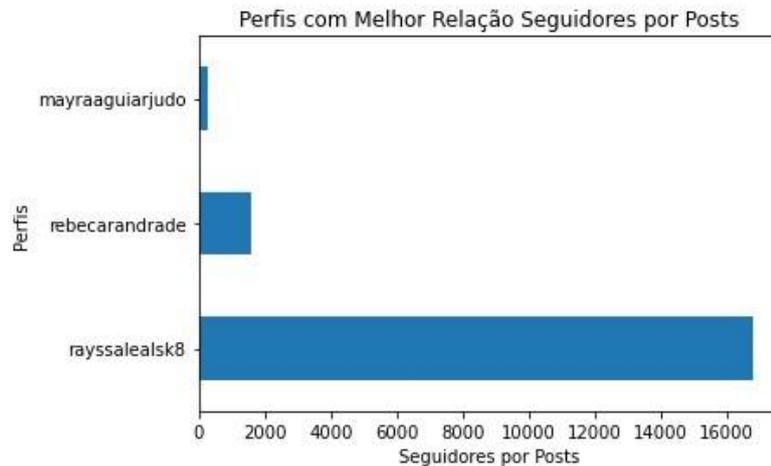


Figura 7 – Relação Número de Post por Número de Seguidores - Fonte: Do autor

Em mais essa métrica o Skate é vitorioso já que, mesmo tendo o menor número de publicações (378) a relação seguidores por post da Rayssa é de mais de 16.750 seguidores por post, contra 1.613 da Rebeca (com 873 posts) e 267 da Mayra (com 673 posts).

Apenas para registro metodológico, foram usados no Twitter os termos “rayssa” ou “fadinha” para a primeira concorrente, “rebeca” ou “rebeca andrade” para a segunda e no caso da Mayra, utilizamos combinações com o nome dela escrito com I e com Y já que as pessoas, sem saber direito, usaram as duas versões, o que diminuiria os números da medalhista de bronze se eu tivesse usado apenas a variante correta. Esse tipo de consideração é importante quando vamos realizar coletas de dados baseados em termos que podem ser digitados com divergência de grafia.

Considerações Finais

Para os pesquisadores da Comunicação, os desafios do ambiente digital, com volumosa e rápida geração de dados, não devem ser considerados como uma olimpíada ou qualquer tipo de competição mas, com certeza, uma excelente oportunidade de

expandir o ferramental metodológico e descobrir novos ângulos para olhar o que nos rodeia e intriga.

Para os que desejam seguir esse caminho, fica aqui o exercício metodológico de um jornalista, com formação em Comunicação, que um dia desejou fazer essa viagem e nunca mais voltou.

REFERÊNCIAS

Amadeu, S. (2019). Democracia e os códigos invisíveis: Como os algoritmos estão modulando comportamentos e escolhas políticas. Coleção Democracia Digital. São Paulo, SP: Edições Sesc.

Amaral, F. (2016). Aprenda mineração de dados – Teoria e prática. Rio de Janeiro, RJ: Alta Books.

Maier, S. R. (2002). Numbers in the News: A Mathematics Audit of a Daily Newspaper. *Journalism Studies*, 3(4), 507-519. <https://doi.org/10.1080/1461670022000019191>

Santos, M. (2014). Textos Gerados por Software – Surge Um Novo Gênero Jornalístico? In *Anais do XXXVII Congresso Brasileiro de Ciências da Comunicação*, Foz do Iguaçu. Recuperado de <http://www.intercom.org.br/papers/nacionais/2014/resumos/R9-2066-1.pdf>

Santos, M. (2016). Narrativas Automatizadas e a Geração de Textos Jornalísticos: A estrutura de organização do lead traduzida em código. *Brazilian Journalism Research*, 12(1), 160-185. <http://dx.doi.org/10.25200/BJR.v12n1.2016.757>

Santos, M. (2016). Internet das Coisas e Sistemas Inteligentes no Jornalismo. *Comunicação e Inovação*, 17, 21-39. Recuperado de http://seer.uscs.edu.br/index.php/revista_comunicacao_inovacao/article/view/3769

Santos, M. (2016). Jornalismo e Internet Das Coisas - Notas sobre tipologia e modelos de utilização. *Revista Latinoamericana de Ciencias de La Comunicación*, 12(22), 134-145. Recuperado de <https://www.alaic.org/revista/index.php/alaic/article/view/621/380>

Santos, M. (2019, 2 sem.). A datificação de um campo de conhecimento. *ORGANICOM USP*, 31, 145-157. Recuperado de <https://www.revistas.usp.br/organicom/article/view/161444/159283>