

# Caracterização e Comparação das Campanhas do Outubro Rosa e Novembro Azul no Twitter

Roberto Walter<sup>1</sup>, Karin Becker<sup>1</sup>

<sup>1</sup>Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)  
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brazil

roberto.wtr@gmail.com, karin.becker@inf.ufrgs.br

**Abstract.** *The Pink October and Blue November campaigns seek to spread education and awareness regarding breast and prostate cancers. Given the success of the Pink October campaigns and the low male engagement in the Blue November, in this article, we present a pioneer comparative evaluation of these campaigns in the online context. Considering Twitter as a platform that provides community engagement, we study the demographic characterization of gender and age, geographic location, activity periods, and user engagement using tweets from the 2017 Pink October and Blue November campaigns. We have identified that these campaigns have reached their target audience, i.e. men and women over 40 years of age. We also discovered patterns in the activities throughout the respective period within campaigns, as well as differences when different countries are considered. Pink October's outreach is definitely greater, measured in terms of the number of tweets, re-tweets, as well as the ability to spread through the social network*

## 1. Introdução

Estudos estimam que uma em cada oito mulheres irá desenvolver o câncer de mama durante sua vida [Altekruse et al. 2010]. A realização de exames de prevenção é fundamental para reverter este quadro. A campanha do Outubro Rosa<sup>1</sup> tem anualmente focado em aumentar a participação feminina nestes exames, bem como educar as pessoas e aumentar os cuidados em geral referentes ao câncer de mama [Jacobsen and Jacobsen 2011, National Breast Cancer Foundation 2017]. Esta campanha organiza caminhadas, eventos esportivos, e distribui materiais de comunicação e vestimentas rosas. O principal objetivo é motivar mulheres a realizar exames que detectam o câncer em estágios ainda iniciais, bem como angariar recursos financeiros. Estes esforços têm obtido muito êxito. Por exemplo, nos EUA o número de mulheres que realizaram exames preventivos subiu de 26% em 1987, para aproximadamente 72,4% em 2010 [for Disease Control and Prevention 2011]. Outro estudo [Glynn et al. 2011] reporta que no mês de outubro há mais interesse no tópico de câncer de mama em pesquisas *online*. No entanto, este mesmo nível de consciência não é observado para outros tipos de câncer, como o de próstata e o de pulmão.

Aproximadamente 63% dos pacientes de câncer buscam informações oncológicas na *internet* [Castleton et al. 2011], e alguns trabalhos têm se dedicado a examinar a importância do uso de mídias sociais, como o Twitter, para propagar o conhecimento sobre a saúde pública [Himmelboim and Han 2014,

<sup>1</sup>O Outubro Rosa teve início ainda nos primeiros anos de 1990, quando outubro foi reconhecido oficialmente pelo governo dos EUA como o mês de consciência sobre o câncer de mama.

Laranjo et al. 2014, Bravo and Hoffman-Goetz 2016]. O Twitter tem sido utilizado em campanhas *online* de diferentes propósitos, como no movimento Black Lives Matter por equidade racial [Olteanu et al. 2016], contra violência baseada no gênero [ElSherief et al. 2017], primavera árabe [Lotan et al. 2011], e também nas campanhas do Outubro Rosa [Thackeray et al. 2013, Nastasi et al. 2017] e Novembro Azul<sup>2</sup> [Bravo and Hoffman-Goetz 2016, Prasetyo et al. 2015]. As campanhas do Outubro Rosa e Novembro Azul (na sequência, referenciadas como OR e NA respectivamente) são estimuladas no Twitter através do uso de *hashtags* como *#breastcancerawareness* e *#cancerdemama* para o OR, e *#Movember* e *#NovembroAzul* para o NA. Estas campanhas aumentaram o conhecimento da população permitindo um trabalho mais efetivo de prevenção, exames, conhecimento sobre tratamentos, pesquisas e posições políticas [Edge 2006]. Porém, divergências nos padrões de desenvolvimento dessas campanhas podem ser fatores que contribuem para os baixos índices de engajamento dos homens em exames de detecção do câncer de próstata nos seus estágios iniciais.

Alguns trabalhos foram desenvolvidos com o intuito de apresentar indicadores sobre períodos de atividades, engajamento, conteúdo e caracterização dos usuários da campanha do OR [Thackeray et al. 2013, Nastasi et al. 2017], ou comparação de conteúdo do Twitter referentes a diferentes tipos de câncer [Borgmann et al. 2016]. No entanto, não há uma avaliação de atividade durante o mês de campanha, de caracterização e localização dos perfis, e do engajamento dos usuários para a campanha do NA. Também não foram identificadas abordagens comparativas entre estas campanhas, apesar de seu propósito similar. A comparação das propriedades de ambas campanhas pode permitir compreender fatores que explicam o sucesso do OR, que expandiu os seus números desde o início das campanhas, e a participação ainda tímida no NA. Desta forma, é importante averiguar se existem divergências nos padrões de desenvolvimento das campanhas, e a partir disso identificar se o público alvo do NA está sendo abordado da forma adequada como está sendo o público alvo do OR. Neste contexto de campanhas convencionais e *online*, será realizada uma avaliação e comparação das campanhas *online*.

O objetivo deste trabalho é permitir um melhor conhecimento sobre as campanhas *online* do NA, além de caracterizar e compará-las às campanhas do OR. Nossa pesquisa busca entender e comparar o engajamento dos usuários proporcionado através de *tweets*, suas características demográficas, e os períodos de mais atividades em cada campanha. Para isso, é avaliada uma base de dados com aproximadamente 680,000 *tweets*, coletados entre setembro e dezembro de 2017. Em nossa análise buscamos responder as seguintes perguntas sobre o NA, comparando-o com o OR:

- **QP 1:** Os usuários envolvidos nas campanhas apresentam características de perfil demográfico e geográfico similares?
- **QP 2:** As campanhas apresentam características temporais similares?
- **QP 3:** As campanhas apresentam abrangência similar na rede social?

Este estudo constitui uma experiência pioneira comparando campanhas de câncer no Twitter, com as seguintes contribuições:

- Complementamos estudos anteriores (e.g. [Nastasi et al. 2017, Jacobson and Mascaro 2016][Thackeray et al. 2013][Prasetyo et al. 2015] com

<sup>2</sup>O Novembro Azul surgiu na Austrália, em 2003, chamado Movember, a aproveitando as comemorações do Dia Mundial de Combate ao Câncer de Próstata, realizado a 17 de novembro.

uma avaliação comparativa entre OR e NA, e análises adicionais: *i*) o público envolvido em cada campanha, verificando se corresponde ao respectivo público-alvo; *ii*) os padrões temporais de atividades em cada campanha, detectando diferenças/semelhanças por campanha e país; *iii*) diferenças na cobertura de tweets em cada campanha, mostrando que o NA tem alcance limitado e que organizações e celebridades não desempenham um papel proeminente. Essas análises podem ser estendidas a outras campanhas sobre câncer.

- Geramos métricas para uma campanha com resultados positivos (OR) e a contrastamos com uma campanha similar (NA) com menos engajamento. Essa comparação ajuda a entender os fatores que influenciam o alcance da campanha do NA no Twitter e, conseqüentemente, aumenta o engajamento da população-alvo em exames de detecção precoce do câncer.
- Estabelecemos um *baseline* para essas campanhas, o que nos permite monitorar sua evolução no ano futuro.

Este artigo está organizado da seguinte forma: na próxima seção são apresentados trabalhos relacionados e as técnicas utilizadas neste trabalho. A Seção 3 define os dados, métodos e proposta deste artigo. A Seção 4 expõem os experimentos e a Seção 5 apresenta as conclusões e trabalhos futuros.

## 2. Trabalhos Relacionados

Vários são os trabalhos que estudam o perfil de engajamento de usuários do Twitter em diferentes causas (e.g. equidade racial, violência baseada no gênero, câncer). A Tabela 1 lista alguns trabalhos com seu contexto de aplicação, destacando desafios comuns a todos: coleta e integração de dados, avaliação de dados temporais, caracterização de usuários por categorias, engajamento de usuários nas campanhas, determinação de informações demográficas, incluindo geolocalização.

Estes trabalhos buscam primariamente criar um *dataset* válido e representativo do contexto estudado. A maioria usa o Twitter, restringindo a tweets que contenham palavras chaves ou *hashtags* em um período determinado, usando a API de integração do Twitter<sup>3</sup> ou outras ferramentas integradas na API. As análises temporais em geral são avaliações de volume de *tweets* e ativismo dos usuários durante períodos de campanhas *online*. O engajamento dos usuários geralmente é medido pela atividade de postagem de *tweets* e *retweets*, detalhado no contexto de informações demográficas como sexo e idade, geolocalização ou categorização de usuários. Assim, busca-se definir as características de participantes e de perfis que possam ter mais ou menos influência/participação nas mobilizações estudadas. Contudo, essas definições demográficas e geográficas são um problema. Por um lado, ao capturar o perfil de um usuário, o Twitter não solicita informações de sexo e idade, e permite que a localização geográfica seja informada em um campo aberto, que muitas vezes é inválida. Portanto, alguns recursos como o Face++<sup>4</sup>, Google Maps, ou plataformas coletivas de definição de dados têm sido utilizadas para buscar uma definição destes dados do usuário com informações que o mesmo possa ter informado em seu perfil, como por exemplo a imagem de perfil ou localização. O Twitter também permite que o usuário geo-referencie os seus *tweets*, mas o número de tweets geo-referenciados é menor que 2% [Schulz et al. 2015]. Assim, utiliza-se frequentemente

---

<sup>3</sup><https://developer.twitter.com>

<sup>4</sup><https://www.faceplusplus.com/>

**Tabela 1. Visão geral de dados e métodos de trabalhos relacionados.**

Obra	Integração Dados	Contexto	Análise Temporal	Engajamento	Tipo de Usuário	Demografia			Estado Político	Geolocalização
						Idade	Sexo	Técnica Definição		
ElSherief et al. 2017	API Twitter	Violência baseada no gênero	-	Sim	-	Sim	Sim	Face++	-	-
De Choudhury et al. 2016	API Twitter	Equidade Racial	Sim	-	-	-	-	-	Estado	Auto-relatado pelo usuário e filtro em Nominatim library
Olteanu et al. 2016	API Twitter	Equidade Racial	Sim	-	Sim	Sim	Sim	Crowdworkers (Crowdfollower)	-	-
Lotan et al. 2011	API Twitter	Primavera Árabe	Sim	Sim	Sim	-	-	-	País	Auto-relatado pelo usuário
Glynn et al. 2011	Google Insights for Search	Outubro Rosa	Sim	-	-	-	-	-	-	-
Thackeray et al. 2013	API Twitter	Outubro Rosa	Sim	Sim	Sim	-	-	-	-	-
Borgmann et al. 2016	Symplur, Tweet Archivist, Twitonomy	Oncologia Urológica	Sim	Sim	Sim	-	-	-	País	Twitonomy
Nastasi et al. 2017	Symplur	Outubro Rosa	-	Sim	Sim	-	-	-	País/Continente	-
Prasetyo et al. 2015	API Twitter	Novembro Azul	-	Sim	Sim	-	-	-	País/Continente	Auto-relatado pelo usuário
Jacobson and Mascaro 2016	API Twitter	Novembro Azul	Sim	Sim	-	-	Sim	-	-	-

as informações disponíveis no perfil, ainda que a informação possa ser imprecisa (e.g. o usuário está em local diferente do relatado no perfil, quando da postagem).

Campanhas relacionadas ao câncer no Twitter aumentaram a conscientização da população sobre o câncer, prevenção e tratamentos [Edge 2006]. Diferenças em tópicos de tweets de acordo com o tipo de câncer foram examinadas em [Borgmann et al. 2016]. As campanhas OR/NA foram caracterizadas em termos de atividades, nível de engajamento, e categorias de usuários [Thackeray et al. 2013, Nastasi et al. 2017]. Uma análise das atividades de angariação de fundos relacionadas com NA em diferentes países foi desenvolvida em [Prasetyo et al. 2015]. O presente trabalho complementa estes trabalhos relacionados, desenvolvendo uma análise comparativa entre as campanhas OR e NA, a fim de auxiliar na compreensão de fatores que contribuem a um maior engajamento da população-alvo nas campanhas on-line.

### 3. Dados e Métodos

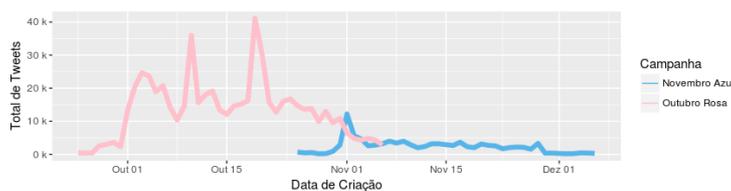
Neste trabalho, analisamos as características temporais das atividades das duas campanhas de câncer, associadas as respectivas informações demográficas e de localização dos usuários e seu engajamento, bem como o alcance das campanhas.

Os dados de *tweets* e perfil dos usuários foram coletados do Twitter a partir da API de consulta aos dados públicos. A consulta foi efetuada com um conjunto inicial de *hashtags* que foram levantadas pelos autores como relacionadas às campanhas do OR e NA nos anos anteriores, e ao examinar as relações no Twitter, agregou-se um novo conjunto de *hashtags* resultando no conjunto da Tabela 2.

As *hashtags* contemplam as principais campanhas nos idiomas inglês, português, espanhol e alemão. A definição do período foi pelo mês da campanha, precedido e sucedido de uma semana, o que resultou no período de 24 de setembro de 2017 a 07 de

Idioma	OR	NA
Português	#OutubroRosa, #Outubrorosa2017, #PrevençãoContraoCâncerDeMama, #outubrorosabr, #cancerdemama	#CâncerDePróstata, #NovembroAzul, #NovembroAzul2017
Inglês	#breastcancerawareness, #thinkpink, #pinktober, #walkagainstcancer, #breastcancer, #IDriveFor, @AmericanCancer, #breastcancerawarenessmonth, #projectpinkblue, #raceforthecure, #BCSM, #BRCA, #pinktober, #chokecancer	#Movember, #ProstateCancer, #BlueNovember, #beatcancer
Espanhol	#OctubreRosa, #miluchaesrosa, #luchacontraelcancerdemama	#noviembreazul
Alemão	#RosaOktober, #bröstkancer	#Prostatakrebs

**Tabela 2. Conjunto de *hashtags***



**Figura 1. Distribuição de Tweets por Data de Criação**

dezembro de 2017. No total foram coletados 678.968 *tweets* referentes a ambas campanhas, postados por 213.905 usuários diferentes. A distribuição do número de postagens por data é mostrada na Figura 1.

A fim de avaliar a abrangência das campanhas, foram mantidos tanto os *tweets* originais, quanto os *retweets*. Após a obtenção dos dados de *tweets* e usuários do Twitter, aplicamos o *Face++* para definir a idade e gênero dos usuários a partir de suas fotos do perfil. Essa escolha foi motivada pela acurácia mínima do *Face++* conforme reportado em [Fan et al. 2014]. Aproximadamente 44% do nosso *dataset* não teve uma definição demográfica pelo *Face++*. A API do Google Maps<sup>5</sup> foi utilizada para obtenção do país do usuário a partir da informação de localidade que o usuário informa em um campo aberto no seu perfil no Twitter (apenas 0.33% dos *tweets* coletados são georreferenciados).

Além disto, foi criado um atributo nas informações do usuário que indica a sua categoria, onde ele pode ser classificado como Celebridade, Organização, ou Indivíduo. Para o perfil se enquadrar como Celebridade, da mesma forma como definido em [Thackeray et al. 2013], o perfil deve ser verificado como verdadeiro pelo Twitter, possuir mais de 100 mil seguidores, e um gênero (masculino ou feminino) identificado pelo *Face++*. O usuário da categoria Organização é um perfil verificado pelo Twitter e que não possui o gênero identificado pelo *Face++*. Os demais perfis são classificados como Indivíduo ou Desconhecido.

## 4. Experimentos

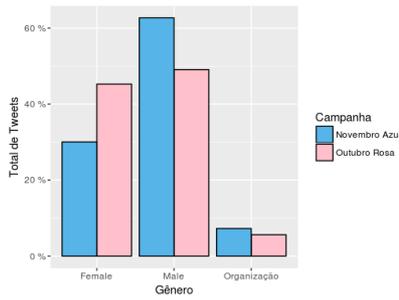
Nesta seção são apresentados os experimentos para as questões de pesquisa (QP) definidas anteriormente, que buscam atingir os objetivos definidos na seção anterior.

### 4.1. QP 1: Os usuários envolvidos nas campanhas apresentam características de perfil demográfico e geográfico similares?

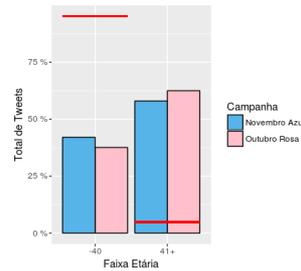
Na Figura 2 pode-se visualizar a distribuição por gênero e campanha, em termos de porcentagem do total de tweets. Identifica-se que cada gênero engaja mais na campanha da qual é alvo, isto é, mulheres na OR e homens no NA. Contudo, observa-se também que nas duas campanhas o gênero que mais engaja é o masculino, representando 62.7% dos usuários engajados no NA. Mesmo na campanha OR, o gênero masculino representa o grupo com maior envolvimento (49.09%). Há também um grupo menor, representado pelas organizações, que possuem participação equilibrada em ambas campanhas, com 7.26% no NA e 5.63% no OR.

No que diz respeito à participação das faixas etárias, os usuários foram separados em um grupo com até 40 anos (-40), e acima de 40 anos (41+). Esses grupos foram

<sup>5</sup><https://developers.google.com/maps/>



**Figura 2. Distribuição por Gênero X Campanha**



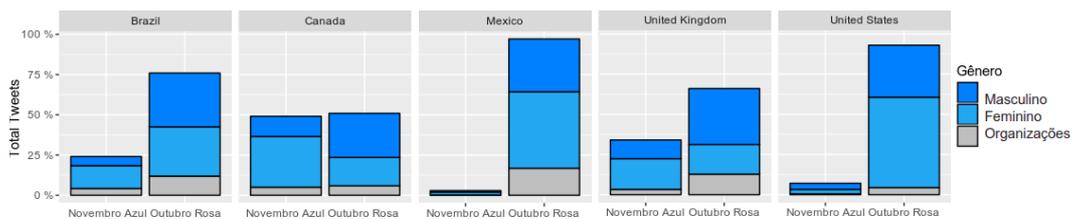
**Figura 3. Distribuição por Faixa Etária x Campanha**

criados para verificar se as pessoas com mais de 40 anos, que são o público alvo da realização dos exames preventivos, possuem participação similar entre as campanhas do OR e NA, e os *tweets* de propósito geral. A Figura 3 apresenta a participação dos dois grupos etários nas campanhas em termos de percentagens sobre o total de *tweets*. Pode-se visualizar uma maior participação do grupo etário 41+, com 62.46% no OR e 57.95% no NA. A distribuição de cada grupo nas duas campanhas é muito próxima, sendo que no OR o grupo 41+ tem uma participação levemente maior, quando comparado ao NA. Para estabelecer um *baseline*, indicamos a participação destas mesmas faixas etárias na postagem de *tweets* em geral de acordo com um estudo [Sloan et al. 2015] (95.2% e 4.8% para os grupos -40 e 41+, respectivamente). Esse alto índice de participação, comparado a *tweets* em geral, indica que ambas campanhas estão atingindo os seus públicos alvos.

As Figuras 4 e 5 apresentam a distribuição de gênero e faixa etária para os cinco países com o maior número de *tweets* sobre as campanhas. Pode-se verificar que com exceção do Canada, em todos os países a campanha do OR atrai maior participação quando comparada a do NA. Os Estados Unidos e o México são os países que apresentam as maiores diferenças, onde a participação no NA está abaixo dos 10%.

Quando detalhada por gênero (Figura 4), observamos uma maior participação de organizações no OR, com destaque para o Brasil, México e UK. No NA, o envolvimento é mínimo. O NA engaja mais sua população alvo, i.e. masculina, exceto nos Estados Unidos, onde é igualitária. Já a campanha OR tem comportamentos distintos conforme o país: maior participação feminina em UK e Canada, igualitária no Brasil, e minoritária nos Estados Unidos. Quanto à idade (Figura 5), observa-se uma predominância de participação do público alvo (41+), exceto pelo Brasil e México, onde há um equilíbrio entre as duas populações.

Conclui-se assim que ambas as campanhas atingem seu público alvo, no tocante



**Figura 4. Distribuição de gênero por país**

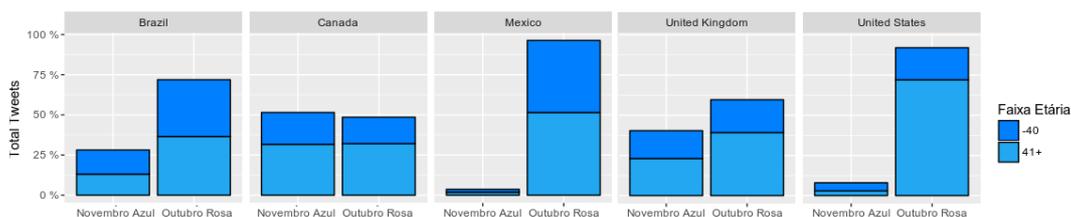


Figura 5. Distribuição de faixa etária por país

à gênero e faixa etária, mas que os níveis de consciência e engajamento são afetados pela cultura ou políticas próprias a cada país. A participação de organizações na causa do NA ainda é tímida.

#### 4.2. QP 2: As campanhas apresentam características temporais similares?

Os números da distribuição de *tweets* por data de postagem conforme foram apresentados na Figura 1, onde se pode verificar que a quantidade de *tweets* nos dias que precedem as campanhas são relativamente baixos, comparados ao respectivo período oficial de campanha. As duas campanhas apresentaram um pico no início de seu respectivo mês. No caso do NA, após os primeiros dias a participação teve um declínio e estabilizou. A campanha do OR, por outro lado, teve três (3) picos (um deles no início da campanha), sendo o declínio da participação observado somente ao final do mês.

O padrão de semelhança entre os países quando considerado o engajamento total nas duas campanhas (Figuras 4 e 5) não é observado quando a participação é detalhada ao longo do período da campanha, como mostra a Figura 6. Canadá, Reino Unido e Estados Unidos, que apresentam proporções de participação total bastante diversos, revelam similaridade nas suas atividades detalhadas por campanha. Para o NA, observa-se um pico de *tweets* nos primeiros dias da campanha, seguida de uma posterior estabilização, enquanto que no OR, existe uma estabilidade ao longo da campanha. O Brasil apresenta um grande pico no início do NA, bem maior que o pico inicial da campanha OR, seguidos de bastante instabilidade de participação ao longo de ambas campanhas. O México também apresenta bastante instabilidade de atividade durante os períodos, mas em um padrão distinto.

Para confirmar estas semelhanças e diferenças de atividades nos países, calculamos as correlações entre as séries do OR e NA de todos pares de países pela medida de Normalized Cross-Correlation. Os resultados são apresentados na Tabela 3 para o OR e NA nas áreas rosa e azul, respectivamente. É possível averiguar que de fato o Canadá, Reino Unido e Estados Unidos apresentam similaridade entre as atividades temporais das campanhas, com correlações acima de 0.90 no NA e 0.70 no OR. Em termos das diferenças, pode-se confirmar também que o Brasil não possui correlações fortes com os demais países, e que o México apresenta valores ainda menores.

A Figura 7 apresenta o percentual acumulado dos *tweets* ao longo dos dias da campanha e confirma a maior atividade do NA no início da campanha. Pode-se verificar que no 10º dia o NA já atinge 28.45% dos *tweets* postados, enquanto que a campanha do OR atinge aproximadamente esse percentual somente no 16º dia da sua campanha. No entanto, as duas curvas apresentam regularidade no seu crescimento durante o mês. As exceções desse crescimento regular são esse crescimento acelerado já mencionado do NA no início do seu mês, um declínio do NA mais acentuado no final, e um crescimento do

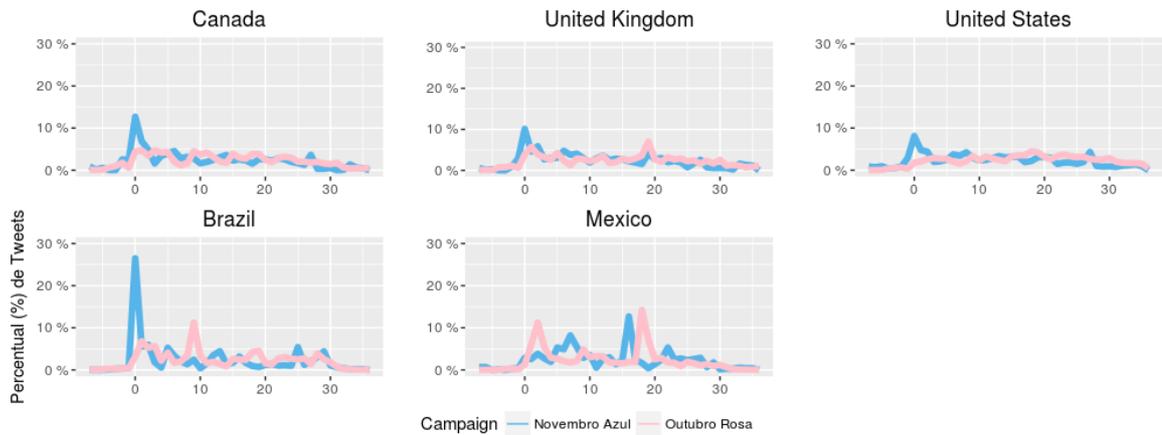


Figura 6. Distribuição Países

Tabela 3. Correlação do nível de atividade por campanha. As correlações OR e NA estão representadas nas cores rosa e azul.

	Canadá	Reino Unido	Estados Unidos	Brasil	México
Canadá		0.8131533	0.7100765	0.7614922	0.6226711
Reino Unido	0.9250876		0.7123116	0.6804556	0.6579495
Estados Unidos	0.9094715	0.9402248		0.5670101	0.551835
Brasil	0.8316607	0.7470957	0.718368		0.6280569
México	0.3545886	0.3862714	0.4534587	0.216265	

OR por volta do 25<sup>o</sup> dia.

Assim, identifica-se que alguns países possuem atividades de *tweets* similares nas campanhas, e que em outros há diferenças nestes períodos de aplicação. Em geral, a campanha do OR apresentou mais regularidade de atividades, e o NA teve oscilações de forma mais ou menos acentuada em cada um dos países.

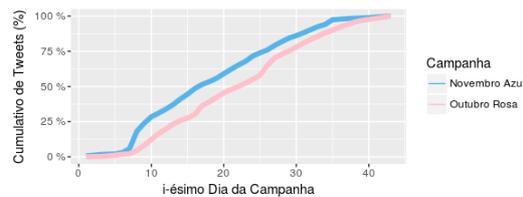


Figura 7. Percentual Acumulado (%) dos Tweets por Dia de Campanha

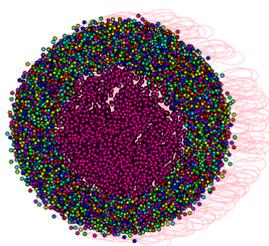
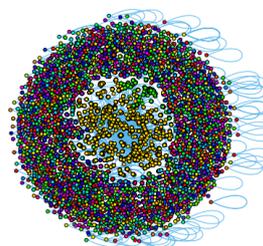
#### 4.3. QP 3: As campanhas apresentam abrangência similar na rede social?

Foi realizada uma avaliação de abrangência para saber em quais campanhas os *tweets* atingem mais pessoas. Quando avaliado o número de *tweets* por categoria de usuário, em todas as categorias o número de usuários e *tweets* é maior na campanha do OR, comparado ao NA, como mostra a Tabela 4. Outro indicador é a média de *tweets* por usuário, a qual também é maior em todas as categorias para o OR. Isto indica que, além da maior participação de usuários nas campanhas, em média os usuários envolvidos na campanha do OR se envolvem com mais postagens, se comparados aos participantes no NA.

**Tabela 4. Características dos *tweets* por tipo de usuário**

	Organizações		Indivíduos		Celebridades		Total	
	OR	NA	OR	NA	OR	NA	OR	NA
Nº Usuários	2364	1007	172826	46025	257	81	175447	47113
Nº de Tweets	17016	3505	564646	96329	643	142	582305	99976
Média	7.19	3.48	3.26	2.09	2.50	1.75	3.31	2.12

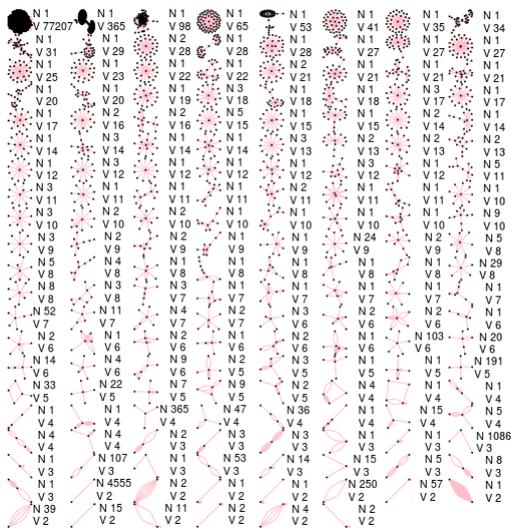
Sabendo que os *tweets* podem ser retuitados e atingir uma rede maior do que apenas os seguidores imediatos do perfil que postou a mensagem, buscou-se compreender as relações por meio dos *retweets* entre os usuários. Para cada campanha, foi construído um grafo que contém a estrutura das relações de *retweets* entre os usuários, apresentados nas Figuras 8 e 9. Cada vértice do grafo representa um usuário, onde as arestas que os conectam representam os *retweets* entre os usuários. As cores dos vértices representam um tipo de estrutura, definida pelo número de usuários e de *retweets*.

**Figura 8. Conexões via *retweets* entre usuários do OR****Figura 9. Conexões via *retweets* entre usuários do NA**

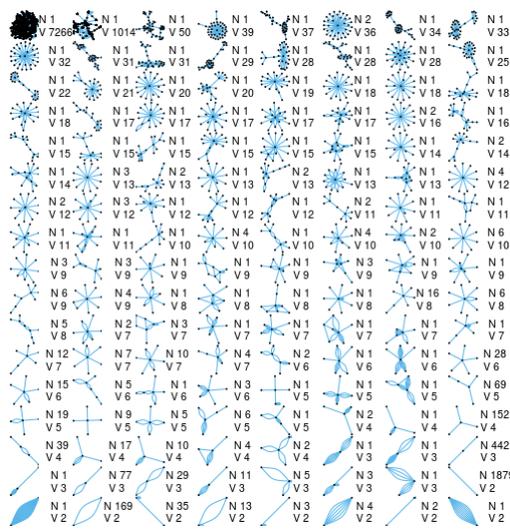
Todos os subgrafos da rede de conexão entre os usuários, juntamente com o número de usuários ( $V$ ) e de repetições no grafo ( $N$ ), são apresentados nas Figuras 10 e 11, ordenados nas linhas por número de usuários em cada estrutura. Se tomarmos as primeiras linhas de cada figura, pode-se verificar que as estruturas do OR possuem um maior número de vértices, e que estas são mais utilizadas do que as maiores estruturas em vértices do NA. Isto indica que o nível de conexão entre os usuários do OR é maior do que a do NA, com os usuários compartilhando bem mais informação de um modo geral. Outro padrão interessante é o alto número de usuários centralizados que são retuitados, e de *retweets* cruzados, onde usuários retuíam-se entre si.

Pode-se reforçar estas suposições com uma apresentação da frequência de número de *retweets* por data de criação, mostrada na Figura 12 para ambas as campanhas. É verificado um maior número de *retweets* de *tweets* relacionados ao OR, onde a quantidade de *retweets* fica em torno de 300 em diversas datas, além de um número bem maior de *tweets* retuitados por dia. Por outro lado, no NA essa máxima de *retweets* fica um pouco acima de 50, ou seja, 16,6% do número do OR. Pode-se verificar também que a mediana (linha horizontal nos gráficos) é bem mais alta no OR (8 *retweets*), comparada ao NA (apenas 4). Esta superioridade também é observada em termos do 3º quartil, onde são observados 34 *retweets* para o OR, e apenas 8 para o NA.

Finalmente, a distribuição do grau de conexões a partir de *retweets* é exibida na Figura 13 para as duas campanhas, para cada uma das diferentes categorias. A maioria dos usuários são retuitados poucas vezes, enquanto apenas pouquíssimos usuários possuem várias conexões por *retweets*. Isso indica que a maioria dos usuários não têm in-



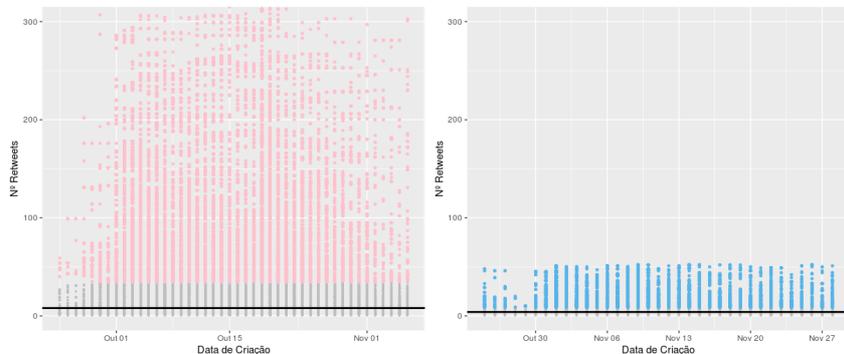
**Figura 10. Estruturas de conexão entre usuários do OR que se retuítam**



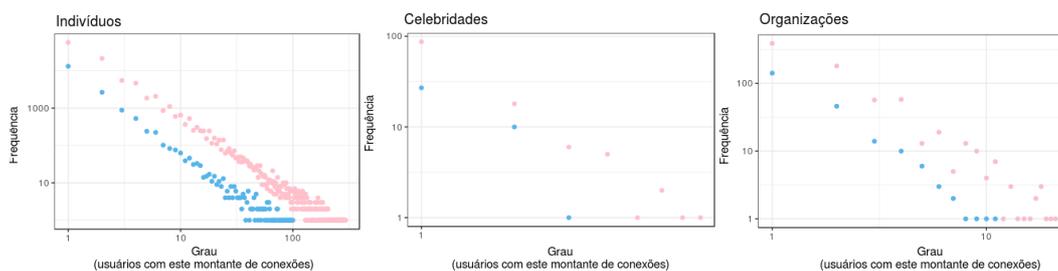
**Figura 11. Estruturas de conexão entre usuários do NA que se retuítam**

fluência para propagar os *tweets* na rede do Twitter. Ainda, em todos os pontos desta série, observa-se que o OR apresentou uma maior capacidade de propagação de informação comparado ao NA. Podemos identificar também que nas três (3) categorias de usuários, o OR prevalece sobre o NA em números de *retweets* para cada *tweet*. Para todas as categorias as diferenças ocorrem nos dois eixos, onde a frequência dos graus é maior para o OR, e também há casos no OR com *tweets* com grau não alcançado pelo NA. Assim, pode-se concluir que a abrangência de propagação dos *tweets* originais através de *retweets* é maior para o OR, pelo fato dos graus de *retweets* do OR serem mais frequentes do que o NA, e no OR os *tweets* atingirem graus de *retweets* que o NA não atinge.

Conclui-se assim que as campanhas não apresentam abrangência similar na rede. Além de mais numerosos, os usuários do OR possuem maior média de *tweets*. A propagação destes *tweets* também é maior na campanha do OR, pelo fato de o engajamento através de *retweets* ser maior no OR do que no NA tanto em quantidade de *retweets*, frequência, e também no tamanho das conexões de usuários através de *retweets* entre si.



**Figura 12. Frequência de Nº de Retweets por Data de Criação**



**Figura 13. Distribuição de grau das conexões entre os usuários (log-log)**

Estas conclusões se aplicam também em cada uma das categorias dos usuários, indicando que a influência de todas as categorias dos usuários é mais forte no OR.

## 5. Conclusão e Trabalhos Futuros

Providenciamos neste artigo algumas das primeiras intuições sobre uma comparação das campanhas do Outubro Rosa e Novembro Azul no contexto *online* no Twitter. Nas nossas análises, o Twitter providenciou boas reflexões de múltiplos aspectos das campanhas. Enquanto nossas análises mostram mais engajamento do público masculino, e que a faixa etária alvo das campanhas (homens e mulheres acima dos 40 anos) é a que mais participa no engajamento, identificamos que há diferenças de aplicação das campanhas nos diferentes países, sendo que os EUA e México não possuem um grande apelo para a campanha do Novembro Azul. Nós também percebemos que as estruturas de rede e engajamento da campanha do Outubro Rosa são maiores que a do Novembro Azul, possibilitando maior propagação da campanha na rede. Esperamos que este trabalho irá direcionar para um progresso nas campanhas do Novembro Azul, e que as mesmas possam alcançar números expressivos como as campanhas do Outubro Rosa. No futuro, pretendemos expandir nossas análises para avaliar os tópicos dos *tweets*, e realizar um estudo comparativo longitudinal.

## References

- Altekruse, S., Kosary, C., Krapcho, M., Neyman, N., Aminou, R., Waldron, W., Ruhl, J., Howlader, N., Tatalovich, Z., Cho, H., et al. (2010). *Seer cancer statistics review, 1975–2007*. Bethesda, MD: National Cancer Institute, 7.
- Borgmann, H., Loeb, S., Salem, J., Thomas, C., Haferkamp, A., Murphy, D. G., and Tsaor, I. (2016). Activity, content, contributors, and influencers of the twitter discussion on urologic oncology. In *Urologic Oncology: Seminars and Original Investigations*, volume 34, pages 377–383. Elsevier.
- Bravo, C. A. and Hoffman-Goetz, L. (2016). Tweeting about prostate and testicular cancers: Do twitter conversations and the 2013 november canada campaign objectives align? *Journal of Cancer Education*, 31(2):236–243.
- Castleton, K., Fong, T., Wang-Gillam, A., Waqar, M. A., Jeffe, D. B., Kehlenbrink, L., Gao, F., and Govindan, R. (2011). A survey of internet utilization among patients with cancer. *Supportive Care in Cancer*, 19(8):1183–1190.
- Edge, L. (2006). Breast-cancer awareness: too much of a good thing? *Lancet Oncol*, 7:611.

- ElSherief, M., Belding, E. M., and Nguyen, D. (2017). # notokay: Understanding gender-based violence in social media. In *ICWSM*, pages 52–61.
- Fan, H., Cao, Z., Jiang, Y., Yin, Q., and Doudou, C. (2014). Learning deep face representation. *arXiv preprint arXiv:1403.2802*.
- for Disease Control, C. and Prevention (2011). Cancer screening — united states, 2010. *Morbidity and Mortality Weekly Report (MMWR)*, pages 61(03); 41–45.
- Glynn, R. W., Kelly, J. C., Coffey, N., Sweeney, K. J., and Kerin, M. J. (2011). The effect of breast cancer awareness month on internet search activity—a comparison with awareness campaigns for lung and prostate cancer. *BMC cancer*, 11(1):442.
- Himmelboim, I. and Han, J. Y. (2014). Cancer talk on twitter: community structure and information sources in breast and prostate cancer social networks. *Journal of health communication*, 19(2):210–225.
- Jacobsen, G. D. and Jacobsen, K. H. (2011). Health awareness campaigns and diagnosis rates: evidence from national breast cancer awareness month. *Journal of health economics*, 30(1):55–61.
- Jacobson, J. and Mascaro, C. (2016). Movember: Twitter conversations of a hairy social movement. *Social Media+ Society*, 2(2):2056305116637103.
- Laranjo, L., Arguel, A., Neves, A. L., Gallagher, A. M., Kaplan, R., Mortimer, N., Mendes, G. A., and Lau, A. Y. (2014). The influence of social networking sites on health behavior change: a systematic review and meta-analysis. *Journal of the American Medical Informatics Association*, 22(1):243–256.
- Lotan, G., Graeff, E., Ananny, M., Gaffney, D., Pearce, I., et al. (2011). The arab spring—the revolutions were tweeted: Information flows during the 2011 tunisian and egyptian revolutions. *International journal of communication*, 5:31.
- Nastasi, A., Bryant, T., Canner, J. K., Dredze, M., Camp, M. S., and Nagarajan, N. (2017). Breast cancer screening and social media: a content analysis of evidence use and guideline opinions on twitter. *Journal of Cancer Education*, pages 1–8.
- National Breast Cancer Foundation, I. (2017). Breast cancer awareness month.
- Olteanu, A., Weber, I., and Gatica-Perez, D. (2016). Characterizing the demographics behind the# blacklivesmatter movement. *OSSM*. <http://arxiv.org/abs/1512.05671>.
- Prasetyo, N. D., Hauff, C., Nguyen, D., van den Broek, T., and Hiemstra, D. (2015). On the impact of twitter-based health campaigns: A cross-country analysis of movember. In *Proceedings of the Sixth International Workshop on Health Text Mining and Information Analysis*, pages 55–63.
- Schulz, A., Schmidt, B., and Strufe, T. (2015). Small-scale incident detection based on microposts. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, pages 3–12. ACM.
- Sloan, L., Morgan, J., Burnap, P., and Williams, M. (2015). Who tweets? deriving the demographic characteristics of age, occupation and social class from twitter user metadata. *PloS one*, 10(3):e0115545.
- Thackeray, R., Burton, S. H., Giraud-Carrier, C., Rollins, S., and Draper, C. R. (2013). Using twitter for breast cancer prevention: an analysis of breast cancer awareness month. *BMC cancer*, 13(1):508.