

Análise de perfis de influência no Twitter: Brasil x Costa Rica (22/6/18)

No dia 22/06 durante a partida de futebol entre Brasil e Costa Rica (às 09:00), foram capturados 1000 *twitters* a cada 10 minutos desde as 08:00 até 00:00. A fim, de se analisar o comportamento sentimental dos usuários da rede acerca da seleção brasileira de futebol.

No Tabela 1 tem-se os dados que foram sumarizados por sentimento a cada 10 minutos entre às 08:00 até 12:00. Já na Figura 1 tem-se a análise dos dados e os principais eventos da partida sinalizados.

Tabela 1 - Dados apurados de 10 em 10 minutos

Horário	Positivo	Neutro	Negativo	Total	%Positivo	%Neutro	%Negativo	%Total
08:00	492	119	171	782	63%	15%	22%	100%
08:10	99	106	90	295	34%	36%	31%	100%
08:20	123	104	125	352	35%	30%	36%	100%
08:30	172	127	170	469	37%	27%	36%	100%
08:40	174	127	207	508	34%	25%	41%	100%
08:50	171	139	246	556	31%	25%	44%	100%
09:00	467	144	176	787	59%	18%	22%	100%
09:10	299	139	342	780	38%	18%	44%	100%
09:20	364	145	364	873	42%	17%	42%	100%
09:30	403	191	363	957	42%	20%	38%	100%
09:40	439	147	326	912	48%	16%	36%	100%
09:50	697	188	251	1136	61%	17%	22%	100%
10:00	436	135	371	942	46%	14%	39%	100%
10:10	538	159	209	906	59%	18%	23%	100%
10:20	654	148	169	971	67%	15%	17%	100%
10:30	542	133	149	824	66%	16%	18%	100%
10:40	513	186	225	924	56%	20%	24%	100%
10:50	787	480	770	2037	39%	24%	38%	100%
11:00	454	134	489	1077	42%	12%	45%	100%
11:10	443	130	286	859	52%	15%	33%	100%
11:20	403	144	243	790	51%	18%	31%	100%
11:30	395	94	266	755	52%	12%	35%	100%
11:40	369	107	405	881	42%	12%	46%	100%
11:50	332	112	350	794	42%	14%	44%	100%
12:00	181	74	570	825	22%	9%	69%	100%
Totais	9.947	3.712	7.333	20.992	47%	18%	35%	100%

Um detalhe importante dos dados utilizados é que foram levados em conta todos os twitters que envolviam a seleção brasileira dentro dessa janela temporal. Sabe-se que além da postagem inicial de um twitter, ainda pode-se ter repostas (*replies*), citações (*quote*) e compartilhamento (*retweet*). Todas essas operações caracterizam interações na rede e propagação de postagem pelos demais usuários.



Figura 1 - Análise dos dados apurados da partida Brasil x Costa Rica (22/06)

Como ficariam os mesmos dados não se levando em conta as propagações, ou seja, citações, respostas e compartilhamentos? Na Tabela 2, vê-se os dados sem a inclusão de citações, respostas e compartilhamentos – foram excluídos os usuários de contas verificadas (pois, deseja-se mensurar influência de perfis não verificados para a propagação na rede).

Pode-se constatar uma grande diminuição da quantidade de postagens, essas postagens correspondem a cerca de 30% das postagens sobre a seleção brasileira que foram capturadas. Ou seja, pode haver a indicação do grau de influência de alguns perfis sobre os assuntos da rede. E, a capacidade de “propagação” ou “reverberação” de temas pelos usuários da rede.

O que causa estranheza foi o mesmo gráfico de análise sentimental anterior, agora levando-se em conta apenas os posts iniciais (sem citação, respostas ou compartilhamentos). Na Figura 2, vemos os dados em um gráfico.

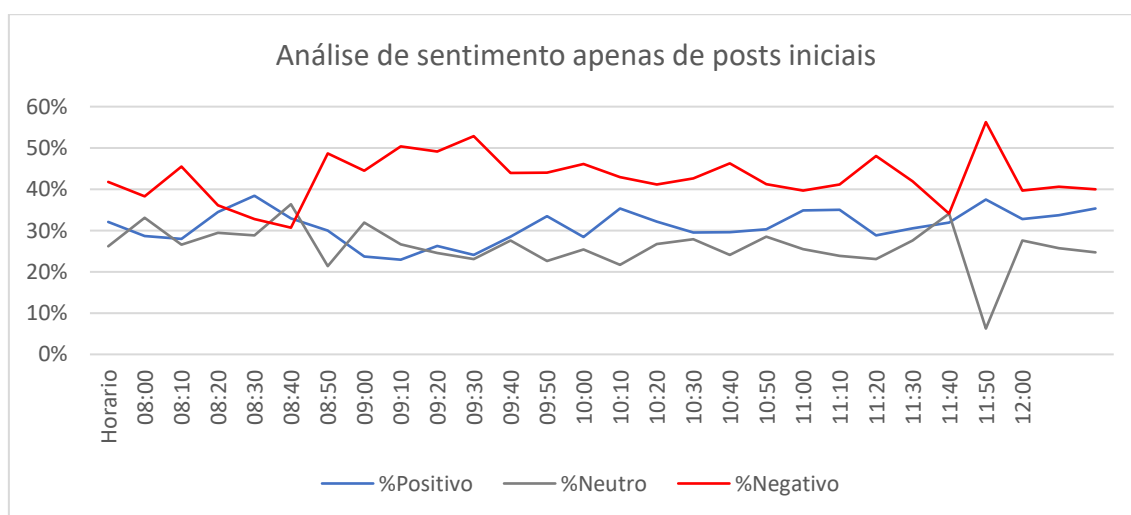


Figura 2 - Análise sentimento apenas de posts iniciais de perfis não verificados

Ao realizar a comparação com o gráfico da Figura 1, nota-se uma grande diferença entre a positividade e negatividade. Certamente, devido as interações nas postagens de usuários concordando ou discordando daquilo que foi postado, gerando as interações na rede.

Tabela 2 - Dados de posts retirando citações, respostas e compartilhamentos

Horário	Positivo	Neutro	Negativo	Total	%Positivo	%Neutro	%Negativo	%Total
08:00	66	54	86	206	32%	26%	42%	100%
08:10	33	38	44	115	29%	33%	38%	100%
08:20	40	38	65	143	28%	27%	45%	100%
08:30	62	53	65	180	34%	29%	36%	100%
08:40	68	51	58	177	38%	29%	33%	100%
08:50	58	64	54	176	33%	36%	31%	100%
09:00	66	47	107	220	30%	21%	49%	100%
09:10	49	66	92	207	24%	32%	44%	100%
09:20	86	100	189	375	23%	27%	50%	100%
09:30	123	115	230	468	26%	25%	49%	100%
09:40	93	89	204	386	24%	23%	53%	100%
09:50	87	84	134	305	29%	28%	44%	100%
10:00	111	75	146	332	33%	23%	44%	100%
10:10	66	59	107	232	28%	25%	46%	100%
10:20	70	43	85	198	35%	22%	43%	100%
10:30	71	59	91	221	32%	27%	41%	100%
10:40	88	83	127	298	30%	28%	43%	100%
10:50	243	198	380	821	30%	24%	46%	100%
11:00	69	65	94	228	30%	29%	41%	100%
11:10	93	68	106	267	35%	25%	40%	100%
11:20	69	47	81	197	35%	24%	41%	100%
11:30	45	36	75	156	29%	23%	48%	100%
11:40	51	46	70	167	31%	28%	42%	100%
11:50	44	47	47	138	32%	34%	34%	100%
12:00	6	1	9	16	38%	6%	56%	100%
12:10	38	32	46	116	33%	28%	40%	100%
12:20	34	26	41	101	34%	26%	41%	100%
12:30	30	21	34	85	35%	25%	40%	100%
Totais	1.857	1.626	2.746	6.229	30%	26%	44%	100%

Na Tabela 3, vê-se os cinco posts (daqueles aqui analisados) que mais foram compartilhados durante o evento. E, continuaram a ser compartilhados após o evento. Analisando a postagem positiva com maior número de compartilhamentos tem-se um total de 886 compartilhamento, o que equivale a aproximadamente 9% do total de postagem positivas no intervalo de tempo de 08:00 às 12:30.

Tabela 3 - Cinco posts mais compartilhados

Post ID	Sentimento	Sentimento	User Screen Name	Retweeted
1010139579867951104	NEUTRO	Atualmente, Philippe Coutinho é o melhor jogador da seleção brasileira.	InfosFutebol	1011
1010128354954088449	POSITIVO	EU ACORDANDO CEDO PRA VER O JOGO DA SELEÇÃO BRASILEIRA #soujoga10nacopa https://t.co/U4jD7zfQa6	ESCANTEIOCURTO	886
1010141617087483914	POSITIVO	SELEÇÃO BRASILEIRA A GENTE COMPROU BRIGA COM TODOS OS PAÍSES, JÁ ESTÁVAMOS ATÉ PENSANDO ONDE FICARIA MELHOR A SEXTA... https://t.co/rSUAtbyU9z	Aguinaldinho	819
1010161535212314626	NEGATIVO	AGORA TA OFICIALMENTE COMPROVADO QUE O ÓDIO É O QUE MOVE A SELEÇÃO BRASILEIRA o juiz não marcou aquele pênalti, ne... https://t.co/Zl55yeKwmF	blckbwr	805
1010157565613953024	POSITIVO	O MELHOR JOGADOR DA SELEÇÃO BRASILEIRA SE CHAMA PHILIPPE COUTINHO!!!! https://t.co/DAQShcWhpo	InfosFutebol	659

Analisando novamente a Figura 2 pode-se notar que os perfis influenciadores tiveram uma tendência a negatividade. No entanto, os usuários da rede reverberaram as postagens positivas. Mostrando assim, uma tendência positivista dos usuários em relação a seleção brasileira (pelo menos durante a partida) – mas, isto é altamente efêmero, reveja a análise da Figura 1.

Na Figura 3, é possível visualizar (de 10 em 10 minutos, dentro da janela de tempo de interesse), a distribuição sentimental das postagens de perfis que tiveram posts compartilhados. Logo, após o encerramento da partida é possível notar o crescimento das postagens negativas. Ou seja, mesmo com a vitória a reputação da seleção ainda é bem negativa na rede.

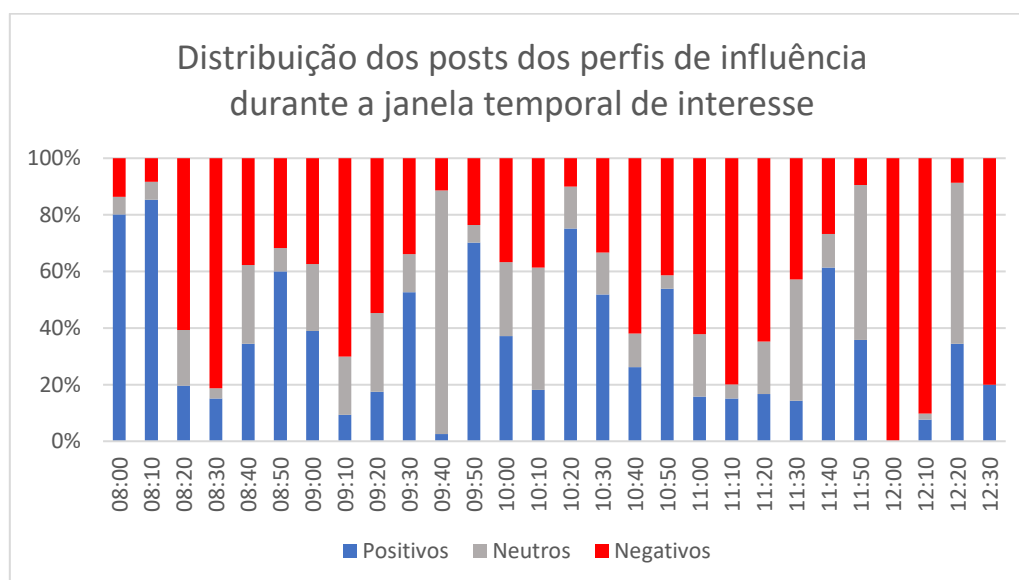


Figura 3 - Distribuição dos posts dos perfis de influência durante a janela temporal de interesse

Uma análise interessante é mensurar quantidade de compartilhamentos que ocorrem dentro de um período de tempo (aqui foi escolhido a primeira hora) após a postagem inicial que gerou os compartilhamentos. A Figura 4, exibe o percentual de compartilhamento (nesse caso de

postagens positivas) distribuídos pelo tempo decorrido, em minutos, desde o momento da postagem inicial.

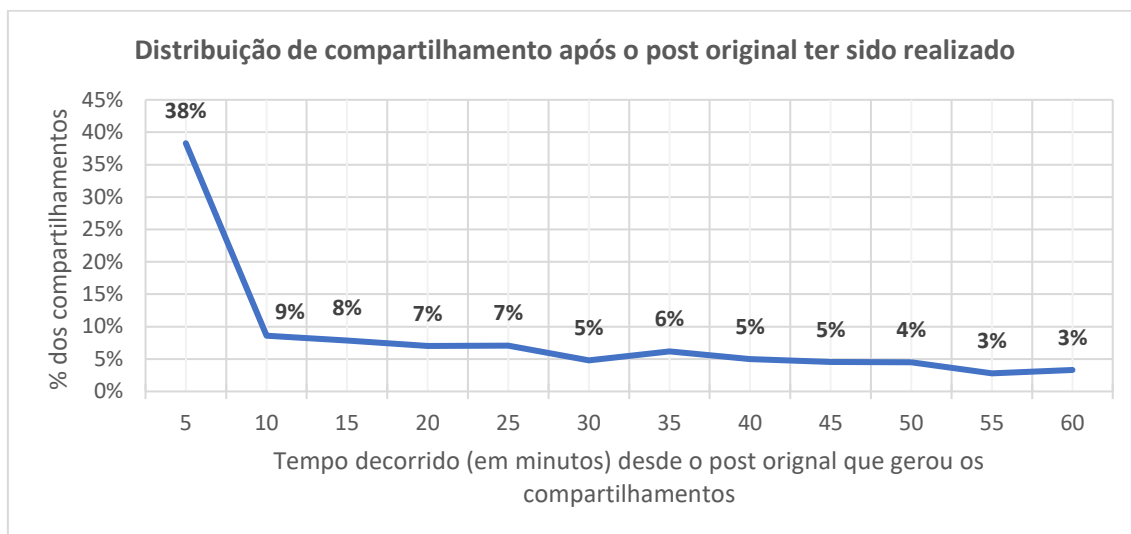


Figura 4 - Distribuição média de compartilhamentos positivos após a postagem inicial

É nítido a concentração de postagens nos primeiros 5 minutos. Este poderia ser um bom *delay* para estudo de efetividade de postagens em eventos de curta duração. No entanto, este mesmo comportamento pode ser preocupante para propagação das ditas *fake News*. Olhando apenas a postagens com até cinco minutos (38% do total de compartilhamentos) após a postem inicial pode-se perceber um acúmulo entre 0 e 3 minutos - Figura 5.

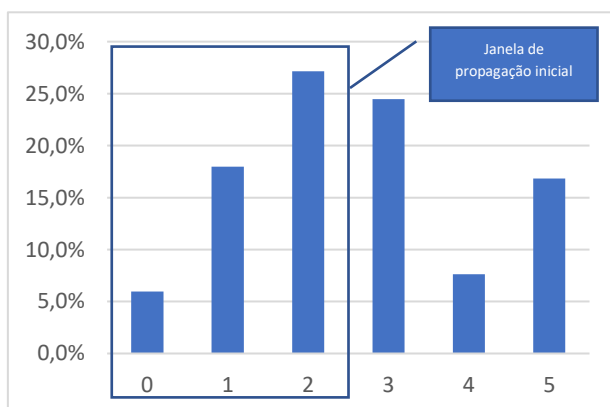


Figura 5 - Compartilhamentos nos 5 primeiros minutos

Este intervalo de tempo pode ser muito curto para que um usuário pouco informado ou mesmo imaturo consiga examinar o conteúdo da postagem e checar a validade das informações.

Isso também vale para as ferramentas que supostamente irão combater as chamadas *fake News*, a janela para a atuação das mesmas é este período. Aqui está-se obviamente extrapolando a

conclusão – ainda é necessário a realização de mais estudos para comprovar esta afirmação.

Se a restrição a uma postagem ocorrer até os primeiros minutos ou até mesmo segundos após a postagem a efetividade de se evitar a propagação na rede é alta. Mas, a qual custo? Qual seria a quantidade falsos positivos? Seria isto uma nova forma de censura?

Realizando a mesma análise para o compartilhamento de postagem negativas de perfis de influência pode-se observar um comportamento “mais agressivo” dos usuários nos compartilhamentos e a reverberação na rede. E, a indicação de um comportamento ondulatório da propagação da negatividade pela rede, por isso, aqui, foi estendido o período de tempo de 60 minutos para 120 minutos - Figura 6.

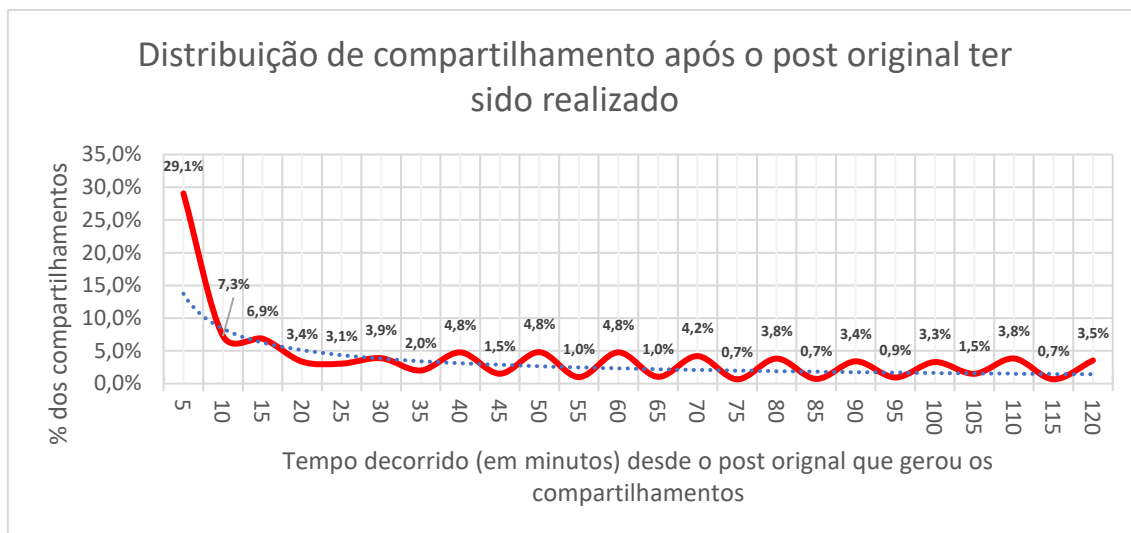


Figura 6 - Compartilhamento de postagens negativas

Se, o comportamento ondulatório for realmente comprovado teremos certeza de que postagens negativas e ou pejorativas, bem como as *fake News*, tende a reverberar pela rede. Mas, para isso deverá ser feito um estudo de fluxo específico para este caso (sendo abordado em outro momento).

Fazendo a mesma análise do compartilhamento de positividade dentro da janela de 5 minutos após a postagem original. Verifica-se que boa parte dos compartilhamentos ocorrem em até 2 minutos tendo uma percentagem significativa ocorrendo antes de 1 minuto.

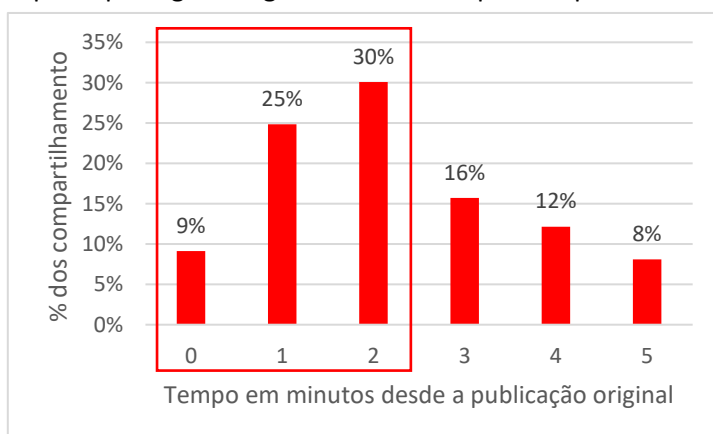


Figura 7 - Análise de compartilhamento após a primeira postagem

Mais uma vez, pode-se concluir que não existe tempo hábil para que um usuário possa analisar o conteúdo da postagem e verificar sua veracidade ou até mesmo julgar se deve ou não propagar o conteúdo da mesma.

O que, de maneira preliminar, pode-se perceber que os usuários tendem a repercutir postagem com conteúdo que para eles são verdadeiros ou que desejam que seja. A engenharia social é algo muito pertinente na propagação de postagens pela rede.

Será realizado a mesma análise nos próximos eventos e também na base já existente capturada para outros termos de interesse. Visando assim, comprovar as afirmações feitas aqui e até mesmo possibilitar a criação de modelos preditivos.