

Método W-Entropia para medir a influencia dos usuários brasileiros nas redes sociais

Li Weigang¹, Deborah Mendes¹, Zheng Jianya¹

¹Departamento de Ciência da Computação - Universidade de Brasília
Brasília – DF

weigang@unb.br, deborahmdf@gmail.com, darcy.zhengjianya@gmail.com

Abstract. *As the tendency worldwide, a large proportion of brazilian users in the Web access the social networks. Each social network has a parameter for classifying an individual, but the method of the measurement is different and presents only one perspective. This paper uses W-entropy method to calculate the influence among various social networks. This method is defined based on the information theory and using the data from Twitter, Facebook and Google search. Comparing to the StarCount method, W-entropy ranking shows the suitable and adequate in the case study.*

Resumo. *A cada ano aumenta o número de brasileiros que possui acesso à web e uma grande parte deles utilizam as redes sociais. Cada rede social possui um parâmetro para classificar um indivíduo, porém cada medida é diferente e apresenta somente uma perspectiva, é necessário uma medida que mostre a influência absoluta entre todas as redes sociais. Este artigo utiliza o método W-entropia para calcular a influência entre diversas redes sociais. Este método é definido baseado na teoria da informação e utiliza dados do Twitter, Facebook e busca do Google. A comparação com o método do Starcount mostra que o W-Entropia se apresenta mais apropriado em um estudo de caso.*

1. Introdução

Mais de 46,3 milhões brasileiros acessam regularmente a internet em casa ou no trabalho, este número aumentou 16% em relação a 2011. Uma das causas deste aumento, é a utilização das redes sociais, por exemplo, a rede social Facebook possui mais de 30,9 milhões de usuários no Brasil e o Twitter possui mais de 14,2 milhões [IBOPE 2011]. As empresas precisam de uma maneira de medir sua presença nas redes sociais e a eficácia de suas ações de marketing.

Com todo este crescimento das redes sociais, é necessário analisar a influência das empresas nas redes sociais para se ter o conhecimento de quais as marcas que possuem uma maior presença na web, quais estratégias de marketing são mais eficientes e para que as empresas possam analisar a concorrência e analisar sua situação nas redes sociais. Também é necessário analisar a influência das pessoas, para que, por exemplo, empresas possam encontrar pessoas adequadas para divulgar sua marca.

Neste artigo foi utilizado o Índice W-Entropia [Weigang et al. 2011], uma medida que se baseia na Entropia de Shannon [Shannon 1948] para calcular a influência nas diversas redes sociais considerando o desequilíbrio das informações.

2. StarCount Squared

O StarCount Squared [StarCount 2012] é um website que tem como principal propósito o entretenimento, ele mostra quem são as celebridades mais populares. Seu site é atualizado diariamente e utiliza dados do Facebook, do Twitter e o número de visualizações do Youtube. Ele coloca porcentagens para cada usuário, onde o usuário mais popular possui o índice de 100%. Seu índice é calculado basicamente por uma média ponderada dos parâmetros. O site leva em consideração o número de visualizações do Youtube, porém este parâmetro favorece somente as pessoas que possuem vídeos no YouTube.

A Tabela 1 mostra a classificação deste website. Ele permite classificar os usuários por país, porém possui poucos usuários brasileiros cadastrados.

Tabela 1. Classificação do StarCount Squared

Posição	Usuário	Seguidores	Fãs	YouTube	Índice
1	Ricardo Kaká	9.444.215	15.461.975	0	56.37%
2	Luciano Huck	4.988.156	6.123.888	0	39.34%
3	Paulo Coelho	3.724.061	8.005.791	0	38.40%
4	Ivete Sangalo	5.884.881	1.734.970	7.189.732	36.09%
5	Neymar Júnior	3.845.001	3.284.139	4.818.215	32.22%

3. Índice W-Entropia

As redes sociais que são estudadas nesse artigo são o Twitter, Facebook e a busca do Google. Se o número de seguidores do mais popular no Twitter for P_j e o número de seguidores de um indivíduo X for X_j , então a razão do Twitter do indivíduo X é $p_j = X_j/P_j$. Considerando o Facebook, se o número de fãs do membro mais popular for P_{j+1} , o número de fãs de um indivíduo X for X_{j+1} , então a razão do Facebook é $p_{j+1} = X_{j+1}/P_{j+1}$ e assim em diante. O peso de cada área das redes sociais é a_1, a_2, \dots, a_n , $\sum a_j = 1$. A razão média do indivíduo X é:

$$m = \sum a_i \times p_j, j = 1, 2, \dots, n. \quad (1)$$

A entropia da informação pode ser empregada para quantificar a distribuição desequilibrada da transmissão da informação entre as diferentes redes sociais. A entropia é definida como um coeficiente de correção para essa transmissão.

Antes de calcular a entropia, é necessário ajustar o primeiro conjunto de valores p_1, p_2, \dots, p_n pois a soma de todos estes termos deve ser igual a 1.

$$q_j = \frac{p_j}{(n+1)}, j = 1, \dots, n \quad (2)$$

$$q_{n+1} = 1 - \sum p_j, j = 1, \dots, n \quad (3)$$

Onde q_1, q_2, \dots, q_n representam os valores numéricos da transmissão da informação entre diferentes redes sociais. O q_{n+1} é uma porcentagem que representa a falta de informação sendo transferida entre diferentes redes.

$$h(q_1, q_2, \dots, q_n, q_{(n+1)}) = -\sum q_j \times \text{Log}_{(n+1)} q_j, j = 1, \dots, n+1 \quad (4)$$

A variável h apresentada na fórmula 4 pode assumir um valor entre 0 e 1. Quando a informação de um indivíduo está sendo transmitida igualmente entre as redes sociais, $h = 1$. Quando a informação do indivíduo está sendo transmitida de forma desigual, onde a maior parte dos p_1, p_2, \dots, p_n são iguais a 0, então $h = 0$. Baseado na fórmula 1 e 4, o Índice W-Entropia pode ser definido como:

$$W - Entropia = h \times m \quad (5)$$

Para simplificar essa fórmula para a aplicação, o valor da fórmula 5 foi modificado em relação ao $W - Entropia_{max}$, que é o valor máximo do índice, e multiplicado por 100, que é o resultado da seguinte equação:

$$IndiceW - Entropia = 100 \times \frac{W - entropia}{W - Entropia_{max}} \quad (6)$$

O Índice W-Entropia é a medida de influência de um membro das redes sociais.

4. Ranking do Índice W-entropia

Para o cálculo neste estudo, foram coletados o número de fãs do Facebook, o número de seguidores do Twitter e o resultado da busca do Google. O Índice W-Entropia foi calculado com os seguintes pesos: o Facebook possui um peso de 45%, o Twitter possui 30% e o Google Search 25%. O *ranking* é exibido na Tabela 2.

- Somente uma empresa está no *ranking* do W-Entropia, isto mostra que as empresas brasileiras precisam melhorar sua presença na web.
- A tabela do StarCount Squared (Tabela 1) utiliza o parâmetro de visualizações do YouTube, porém dos 5 brasileiros selecionados, três deles possuem este parâmetro igual a zero. O ideal é escolher um parâmetro que seja aplicável a todos os usuários, por exemplo, a busca do Google, que é aplicável a qualquer pessoa/empresa.
- O índice do StarCount deixa de fora diversos brasileiros, por exemplo, o cantor Michel Teló que possui uma presença significativa nas redes sociais, com quase 3 milhões de fãs no Facebook e mais de 1 milhão de seguidores.
- O escritor Paulo Coelho é conhecido mundialmente, ele obteve terceiro lugar no índice do StarCount devido a uma diferença com o número de seguidores do Luciano Huck. No rank W-Entropia ele está em segundo lugar, na frente do apresentador brasileiro, Luciano Huck.

Tabela 2. Classificação gerada com o Índice W-entropia

Posição	Usuário	Seguidores	Fãs	Google	Índice W-Entropia
1	Ricardo Kaká	9.444.215	15.461.975	13.700.000	100
2	Paulo Coelho	3.724.061	8.005.791	24.300.000	85
3	Michel Teló	1.312.221	2.862.730	37.500.000	60
4	Ivete Sangalo	5.884.881	1.734.970	15.500.000	56
5	Luciano Huck	4.988.156	6.123.888	863.000	53

5. Comparação dos resultados do Índice W-entropia com o StarCount

A classificação do StarCount é baseada na média dos parâmetros do Facebook, Twitter e YouTube, a simples média destas medidas não é suficiente para fornecer uma medida justa. A média não reflete a distribuição de informação das pessoas/marcas através das multi-plataformas, o número de fãs de uma pessoa é uma medida diferente do número de seguidores. O cenário do W-Entropia se mostra ser mais realista e abrangente.

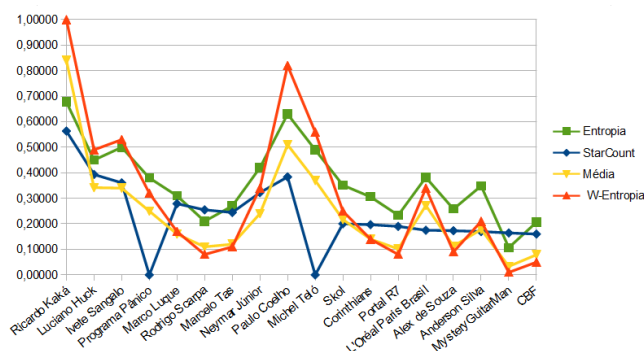


Figura 1. Comparação com a média, Starcount, W-entropia e entropia.

6. Conclusões

Este artigo apresentou uma pesquisa sobre o Índice W-Entropia, que mede a influência dos membros das redes sociais. O conceito e aplicação do W-entropia foram descritos comparando com o ranking StarCount.

As redes sociais estão sempre se modificando, é importante atualizar o Índice W-Entropia à medida que isto acontece. Um dos trabalhos futuros nesta área será estudar novas medidas que podem ser incluídas no cálculo, como por exemplo, número de retweets e número de pessoas que estão falando de um determinado usuário.

Referências

- IBOPE (2011). Pesquisa sobre a utilização das redes sociais. <http://www.ibope.com.br/calandraWeb/servlet/CalandraRedirect?temp=0&proj=PortalIBOPE&pub=T&db=caldb>. [Online; Acessado: 01/04/2012].
- Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 623656.
- StarCount (2012). Starcount squared. <http://www.starcountssquared.com/>. [Online; Acessado: 01/04/2012].
- Weigang, L., Zheng, J., and Li, D. (2011). Analysis of w-entropy index: the impact of members on social networks. *In Proceedings of the IADIS International Conference WWW/INTERNET, Rio de Janeiro, Brazil, 171-178. Best Paper Award.*