Inteligência Analítica Aplicada ao
Processo de Participação Social

Detecção de Usuários Influentes e Conteúdos Relevantes

InWeb/UFMG CEWEB/NIC.BR MJ

Agosto de 2015

Conteúdo

1 Contexto

1.1 Descrição

A dinâmica dos processos de participação social muito se assemelha a redes sociais mais formais, sejam elas online ou presenciais. Neste contexto, a disseminação de informações e ideias e o exercício da influência de alguns usuários sobre outros se mostra como um elemento fundamental para o resultado do processo em si. O objetivo deste sub-projeto é determinar quais usuários são influentes e quais os conteúdos relevantes, ou seja, ordenar tanto grupos de usuários quanto o conteúdo por eles disseminado.

1.2 Motivação

Cada processo de participação social tem uma dinâmica própria em razão tanto do próprio tópico objeto do processo quanto dos interesses envolvidos no mesmo. A transparência que é inerente a esses processos permite que os usuários interajam através das suas propostas e comentários, os quais configuram padrões de disseminação de informações e criam a oportunidade para o surgimento de formadores de opinião.

1.3 Entrada

A entrada do processo compreende as unidades de informação da consulta, as quais contem duas informações de interesse: usuários e conteúdo. Os usuários são definidos unicamente pela plataforma e isso facilita o processo. O conteúdo é sempre gerado por um dado usuário, e pode ser encadeado com base na unidade de informação referenciada e na sua evolução cronológica.

1.4 Saída

A saída do processo ordena os usuários e seus conteúdos de acordo com a sua capacidade de formação de opinião e relevância, respectivamente. Essas ordenações podem também ser apresentadas por usuário, por exemplo um usuário de referência que se queira entender o contexto.

1.5 Estratégia

Propomos a adaptação e utilização de uma técnica que se baseia em uma definição intuitiva e circular de relevância e influência [5]:

Um conteúdo é considerado relevante se ele é criado e propagado por usuários influentes, e usuários influentes criam conteúdo relevante.

Além disso, podemos reformular esse princípio de relevância global a fim de suportar funções de relevância e influência personalizadas da seguinte forma:

Um conteúdo c é considerado relevante para um dado usuário u se ele é criado e propagado por usuários que são influentes para u, e um usuário v é considerado influente para u se esse usuário v cria conteúdo que é relevante para u.

A personalização pode ser útil como ferramenta de recomendação por contexto do consulta em curso, auxiliando avaliadores a identificar usuários e conteúdos de interesse para uma dada tarefa de síntese.

Em ambos os casos podemos modelar a determinação os usuários influentes e os conteúdos relevantes de forma circular, a qual pode ser resolvida de forma interativa por caminhamento randomizado.

2 Solução Proposta

Esta seção provê definições para os principais conceitos empregados neste projeto.

2.1 Relevância e influência

Esta seção apresenta alguns conceitos importantes relacionados à identificação de conteúdos relevantes e usuários influentes com base em dados de difusão de informação. A idéia é associar os usuários influentes e conteúdo relevante através de uma definição circular, seguindo o seguinte princípio:

Um conteúdo é considerado relevante se ele é criado e propagado por usuários influentes, e usuários influentes criam conteúdo relevante.

Além disso, podemos reformular esse princípio de relevância global a fim de suportar funções de relevância e influência personalizadas da seguinte forma:

Um conteúdo c é considerado relevante para um dado usuário u se ele é criado e propagado por usuários que são influentes para u, e um usuário v é considerado influente para u se esse usuário v cria conteúdo que é relevante para u.

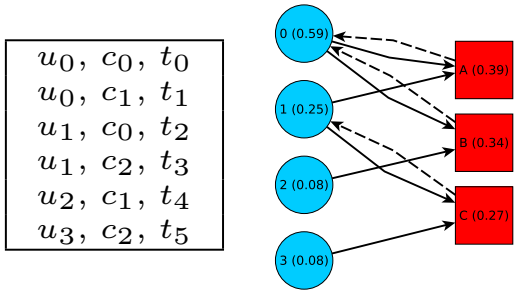
Seja C o conjunto de conteúdos e U o conjunto de usuários. Nós definimos a relevância global de um conteúdo $c \in C$ como uma função $r(c)$. Além disso, definimos a influência global de um usuário $u \in U$ como uma função $p(u)$. Como se trata de uma métrica orientada ao comportamento do usuário, a importância global de $r(c)$ depende da relevância personalizada $r(c, u)$, que dá a relevância do conteúdo c para o usuário específico u. No entanto, $r(c)$ também é afetada pela influencia dos usuários, ou seja quanto mais influentes forem os usuários para os quais c é relevante, mais relevante será c. Portanto, a relevância do conteúdo é baseada na influência dos usuários. Da mesma forma, definimos a influência $p(u)$ de um usuário u com base na relevância do conteúdo que ele produz. A função de influência personalizada $p(u_i, u_j)$ dá a influência de um usuário u_i para um usuário u_j. Essas definições circulares são formalizadas na Seção 2.3.

É interessante entender o problema de identificar os usuários influentes e conteúdo relevante sob uma perspectiva de recomendação. Um conteúdo que é relevante para alguns usuários deveria ser recomendado para tais usuários. Portanto, podemos aplicar as funções $r(c, u)$ e $p(u_i, u_j)$ em um contexto de recomendação. Ao avaliar a eficácia dessas funções, podemos avaliar a qualidade das funções $r(c)$ e $p(u)$. Esta abordagem torna-se especialmente útil quando não há informação sobre a relevância do conteúdo e a influência dos usuários - que é um cenário muito frequente. Por esse motivo utilizamos a tarefa de recomendação como uma forma de avaliar a eficiência da nossa nova métrica.

2.2 Dados de difusão de informação

Nós chamamos de dados de difusão de informação o conjunto de ocorrências de um item de informação. Cada ocorrência de um item é definida como uma tupla na forma $< u, c, t >$, onde u é um usuário do conjunto de usuários U, c é um conteúdo do conjunto de conteúdos

C , e t é um instante do tempo. Portanto, os dados de difusão de informação descrevem a associação entre usuários e conteúdo ao longo do tempo.



(a) Dados (b) Modelo de difusão

Figura 1: Modelagem de dados de difusão: Visão geral. Os dados em 1(a) representam usuários u que postaram conteúdos c nos tempos t. Esses dados são utilizados para gerar o grafo bipartido 1(b). Os usuários são representados por círculos e os conteúdos pelos quadrados. As setas contínuas ligam o usuário à cada conteúdo que ele criou ou propagou, e as setas tracejadas ligam o conteúdo de volta apenas ao usuário que o criou. Os números dentro de cada vértice indicam os valores de influência/relevância calculados pela nossa técnica.

A Figura 1(a) mostra um exemplo ilustrativo de um conjunto de dados de difusão de informação onde $U = \{u_0, u_1, u_2, u_3\}$, $C = \{c_0, c_1, c_2\}$ durante o intervalo de tempo $[t_0, t_5]$. Considerando-se o portal do MJ como exemplo, U representa o conjunto de usuários que fazem contribuições e comentários, C representa o conteúdo postado pelos usuários, e os instantes de tempo são definidos de acordo com o momento das postagens. Dados de difusão de informação aparecem em muitos outros cenários da vida real, especialmente em aplicações de mídia social. É importante notar que a nossa definição de dados de difusão de informação não leva os relacionamentos sociais entre os usuários da rede em consideração. Ou seja, não temos qualquer informação sobre amizades, seguidores ou qualquer outro tipo de relação que pode ser considerado como um meio para a difusão da informação.

2.3 Modelo de Difusão de Informação

Nosso modelo é baseado em um grafo bipartido $G(U, C, F, E)$ que associa os usuários à conteúdos através de dois conjuntos de arestas, F e E. Para cada usuário $u \in U$ e conteúdo $c \in C$, existe uma aresta direcionada $(u, c) \in F$ se o usuário u postou o conteúdo c, ou seja, se ele é o criador (primeira pessoa a postar aquele conteúdo) ou se ele propaga esse mesmo conteúdo posteriormente na rede. As arestas em F atribuem relevância ao conteúdo com base na influência do usuário. Além disso, existe uma aresta direcionada $(c, u) \in E$ que parte de um conteúdo para o usuário que o criou, e cada conteúdo tem apenas um criador. Arestas em E dão crédito aos usuários de acordo com a relevância do conteúdo que eles criam. A Figura 1(b) apresenta o grafo bipartido construído a partir dos dados mostrados na Figura 1(a). Definimos a relevância do conteúdo $r(c)$ como a frequência relativa em que um *random surfer* [1] que começa a partir de um nó de usuário arbitrário e, navegando através do grafo bipartido $G(U, C, F, E)$, atinge um determinado conteúdo c. Já a relevância personalizada $r(c, u)$ começa a partir de um determinado usuário u em vez de um usuário arbitrário. De um modo semelhante, a influência de um usuário $p(u)$ é a frequência relativa que o *random surfer*

visita um determinado usuário e pode ser personalizado, iniciando a partir de um usuário particular. Para dar uma visão mais realista sobre o nosso modelo, vamos considerar o Portal do MJ como um cenário de exemplo. O algoritmo seguido pela nosso *random surfer*, com base em dados do Portal, pode ser descrito como se segue:

1. Seleciona um perfil de usuário arbitrário;
2. Escolhe um conteúdo aleatório ou comentário do usuário atual;
3. Seleciona o perfil do autor do conteúdo dado; e
4. Volta para o passo 2.

O grafo bipartido $G(U, C, F, E)$ pode ser representado por duas matrizes, M e L. A matriz $M = (m_{i,j})$ é $|U| \times |C|$ e $m_{i,j} = 1/q_i$, onde q_i é a quantidade de conteúdo que u_i criou ou propagou. Além disso, $L = (l_{i,j})$ é $|C| \times |U|$ e $l_{i,j} = 1$ se o usuário u_j criou o conteúdo c_i ou $l_{i,j} = 0$, caso contrário. Com base em M e L, a função de relevância do conteúdo $r(c)$ e a função de influência do usuário $p(u)$ são definidas como:

$$r = pM$$

$$p = rL$$

onde r é um vetor de relevância do conteúdo (ou seja, r_i é a relevância do conteúdo c_i) e p é um vetor de relevância do usuário (ou seja, p_j é a influência do usuário u_j). Nessa definição, assumimos que já temos um dos vetores (r ou p), a fim de calcular um a partir do outro, o que não acontece na realidade. Contudo, r e p podem ser calculados recursivamente:

$$r^{(k)} = r^{(k-1)}LM$$

$$p^{(k)} = p^{(k-1)}ML$$

onde $k \geq 0$ e $r^{(0)}$ e $p^{(0)}$ são vetores uniformes[1]. Esse modelo apresenta dois problemas importantes: (1) A possível presença de usuários *dangling* e (2) a possível existência de *buckets*. Um usuário *dangling* é um usuário que nunca propaga conteúdo de outros usuários. Considerando a metáfora do *random surfer*, o *surfer* ficará preso sempre que um usuário *dangling* u é atingido pois ele sempre seguirá arestas para o conteúdo gerado por u e depois consequentemente voltará para u. O PageRank também precisa lidar com páginas *dangling* e nós aplicamos uma solução semelhante aqui. Criamos uma aresta (u, c) de cada usuário *dangling* para um conteúdo "fantasma"c e adicionamos uma aresta (c, u) a partir do conteúdo fantasma para cada usuário $u \in U$. Como consequência, garantimos que o *random surfer* conseguirá, a partir de um usuário *dangling*, chegar a qualquer outro usuário em U. No gráfico mostrado na Figura 1(b), u_0 é um usuário *dangling*. Um *bucket* é um subgrafo fortemente conexo do grafo bipartido. Quando o *random surfer* atinge um *bucket*, ele não é capaz de deixá-lo. Podemos ver um usuário *dangling* como se fosse um *bucket* de tamanho 1. A fim de evitar que o *random surfer* fique preso em *buckets*, podemos adicionar um mecanismo de amortecimento ao nosso modelo. Esse mecanismo determina uma pequena probabilidade d do *random surfer* pular do usuário atual para um conteúdo aleatório ou vice-versa. Nós adicionamos esse mecanismo na definição de r e p da seguinte forma:

$$r^{(k)} = cr^{(k-1)}LM + (1 - c)u$$

[1]Em um vetor uniforme, todos os valores são iguais e a soma deles é 1

$$p^{(k)} = cp^{(k-1)}ML + (1-c)u$$

em que u é um vetor uniforme. Podemos reformular as equações acima algebricamente a fim de obter as suas soluções exatas de uma forma não recursiva:

$$r = (1-d)u(I - dLM)^{-1} \quad (1)$$

$$p = (1-d)u(I - dML)^{-1} \quad (2)$$

Na próxima seção, vamos discutir por que essa formulação algébrica não é computacionalmente eficiente. Duas questões mais importantes neste momento são: (1) Será que essas equações têm uma solução? e (2) Essas soluções são únicas? A resposta afirmativa para a primeira pergunta vem do fato de que as matrizes ML e LM são estocásticas. De fato, sabe-se que o produto de duas matrizes estocásticas é sempre uma matriz estocástica. Além disso, uma combinação linear de duas matrizes estocásticas é também estocástica. Em relação à pergunta 2, podemos mostrar que as nossas equações têm uma solução única, baseada no *teorema de Perron-Frobenius* [3, 2]. O teorema de Perron-Frobenius diz que se uma matriz A é irredutível (ou seja, se seu gráfico associado é fortemente conectado) e também quadrada não negativa, então a equação $xA = rx$, onde $x > 0$ e $\sum_i x_i = 1$, tem uma única solução. Como M, L, e u são não-negativos, nossas equações tem matrizes não negativas. Além disso, a remoção de usuários *dangling* e de *buckets* garante que ML e LM são irredutíveis. Na Figura 1(b), calculamos os valores da influência do usuário e da relevância do conteúdo usando d igual a 0,85. Podemos notar que o usuário mais influente é u_0 ($p(u_0) = 0.59$), pois os dois conteúdos produzidos por u_0 (c_0 e c_1) são propagados por dois usuários (u_1 e u_2). Os conteúdos produzidos por u_1 também são propagados por dois usuários, mas esses usuários são menos influentes do que os usuários que propagam o conteúdo a partir de u_0. Portanto, u_1 é menos influente que u_0. O conteúdo mais relevante é c_0 porque ele foi difundido por dois usuários influentes (u_0 e u_1). Embora c_2 também seja difundido por dois usuários, esses usuários não são tão influentes como os associados a c_0. A elaboração de valores personalizados de relevância de conteúdo ($r(c, u)$) e influência de usuário ($p(u_i, u_j)$) em nosso modelo é simples. Nestes cenários, em vez de iniciar a partir de um usuário arbitrário, vamos supor que o *random surfer* começa a partir de um usuário específico para qual o modelo está sendo personalizado. Da mesma forma, em vez de saltar para um conteúdo aleatório com uma probabilidade não-zero, o *random surfer* sempre salta de volta para esse nó específico. Esse comportamento pode ser induzido substituindo o vetor uniforme u por um vetor 1_i, que é um vetor com todos os elementos iguais a 0, com exceção da posição i igual a 1, onde u_i é o usuário para o qual o modelo está sendo personalizado.

2.4 Solução Eficiente

Na seção anterior, descrevemos as equações que definem influência do usuário e relevância do conteúdo no nosso modelo. Para aplicar esse modelo em cenários reais, com potencialmente grande volume de usuários e conteúdo, precisamos resolver tais equações de forma eficiente. Em situações reais, as matrizes M e L tendem a ser muito grandes e esparsas. Portanto, uma solução eficiente para nosso modelo deve levar em consideração essas propriedades. Como mostrado nas Equações 1 e 2, podemos calcular os vetores r e p invertendo uma matriz $|U| \times |U|$ e uma matriz $|C| \times |C|$. Como a inversão de uma matriz $n \times n$ tem custo $O(n^3)$, calcular os valores exatos de r e p não é viável em situações reais. No entanto, o método das potências [4, 2], que é um método de iteração rápido para calcular o autovalor e autovetor

dominante de uma matriz, pode ser aplicado no cálculo de r e p. O Algoritmo 1 descreve o método da potências. Ele recebe duas matrizes (Z_1 e Z_2) e repetidamente itera sobre a solução g, que é iniciada como uniforme, até que um determinado número de iterações k seja atingido. Se $Z_1 = M$ e $Z_2 = L$, o método nos dá o vetor de influência (p). Por outro lado, se $Z_1 = L$ e $Z_2 = M$, ele nos dá o vetor de relevância (r). Conforme descrito na seção anterior, podemos calcular os valores personalizados de influência e relevância para um usuário u_i, substituindo o vetor uniforme U por um vetor 1_i que tem 1 na i-ésima posição e 0 nas posições restantes. Além de aplicar o método de potências para calcular a influência e relevância, fazemos uso de representações esparsas das matrizes M e L, com o objetivo de reduzir a quantidade de memória e o tempo de execução para computar r e p. Mais especificamente, representamos matrizes no *formato de coordenadas*. Valores são armazenados numa lista de tuplas (linha, coluna, valor), onde apenas tuplas com valores diferentes de zero são inseridas.

Algoritmo 1: Método das Potências. O método recebe duas matrizes Z_1 e Z_2 e repetidamente itera sobre a solução g, que é iniciada como uniforme, até que um número de iterações k seja atingido.

Input: Z_1, Z_2, k, d
Output: g
1 $u \leftarrow$ vetor uniforme;
2 $g \leftarrow u$;
3 $i \leftarrow 0$;
4 **while** $i < k$ **do**
5 $\quad g \leftarrow dgZ_1Z_2 + (1.0 - d)u$;
6 $\quad i \leftarrow i + 1$;

3 Estudo de Caso

3.1 Preparação dos dados

A base utilizada para a avaliação apresentada a seguir, tem 232 items comentáveis, dos quais 188 foram de fato comentados por 1020 comentários enviados por 227 usuários.

Foi realizado um pré-processamento da base, identificando duas informações relevantes em relação aos endossos, o seu destino e o nível de concordância. Distinguimos três tipos de concordância: positiva, negativa e neutra. Concordância positiva é representada pela ocorrência de “concordo", enquanto concordância negativa é caracterizada pela ocorrência dos termos “não concordo"e “discordo". Como discutido posteriormente, essa identificação possui várias oportunidades de melhoria. Durante este mesmo processo, distinguimos dois tipos de comentários: comentários sobre itens do anteprojeto e comentários sobre comentários, que servem para identificar formadores de opinião, enquanto todos os comentários provem informações sobre conteúdo relevante. Assim, os identificadores que se iniciam por “i_"se referem a itens do projeto, enquanto aqueles iniciados por “c_"se referem a comentários.

3.2 Resultados Preliminares

Nesta seção apresentamos resultados preliminares após a execução do algoritmo discutido na Seção 2.

Id	Escore	Conteúdo
i_5	108994.728384	Art 2, caput
i_70	62553.2634291	Art 10, Parag 1o
i_15	55939.1752377	Art 4, caput
i_231	35340.1816057	Art 51
i_20	35340.1816057	Art 5, Inciso 3
(...)		
c_1017	26949.2383413	Art 2, Parag 2, Inciso 2
c_76	26949.2383413	Art 2, Parag 2, Inciso 2
c_260	26949.2383413	Art 5, Inciso I
c_599	26949.2383413	Art 2, Parag 2, Inciso 2

Tabela 1: Conteúdos relevantes positivos

3.2.1 Conteúdo relevante positivo

A Tabela 1 apresenta dois conjuntos de conteúdos relevantes positivos. No topo da tabela estão itens do anteprojeto, que são em muito maior número que os comentários, apresentados na parte inferior da tabela. Podemos perceber que os primeiros do ranking são itens do anteprojeto, mas na nona posição já temos quatro comentários que surgem com uma certa relevância (empatados com o mesmo index). O interessante é que, ao olharmos os dados, vemos que a relevância dos comentários é muito significativa. Isso acontece porque muita gente, ao comentar um item, está na verdade comentando um comentário ("Concordo com Fulano"), enquanto que o contrário não acontece. A seguir apresentamos alguns exemplos de conteúdos relevantes positivos:

i_70 (Art. 10, inciso VII, item c, parágrafo 1)

Texto: *Considera-se nulo o consentimento caso as informações tenham conteúdo enganoso ou não tenham sido apresentadas de forma clara, adequada e ostensiva.*

5 comentários, dentre os quais:

1. gleison melo:
   Concordo, é um meio de assegurar os direitos das pessoas que foram enganadas.
   Avaliação: concorda com o item: verdadeiro positivo
2. rafaela16:
   Concordo com a sugestão trazida pela Gabriela. (...)
   Avaliação: concorda com o comentário de outro usuário: falso positivo

c_1017 (comentário de Wellington Cremasco no Art. 2, parágrafo 2, inciso II)

Texto: *Este inciso preciso ser mais detalhado. Citar apenas que podem ser realizados para fins jornalísticos deixa o inciso aberto a interpretações que podem não condizer com o objetivo da Lei*

1. Renata Oliveira
   Concordo
   Avaliação: concorda com o comentário: verdadeiro positivo.

Id	Escore	Conteúdo
i_20	243447.512673	Art 5, Inciso 3
i_10	145511.343317	Art 2, Parag 2, Inciso 1
i_11	145511.343317	Art 2, Parag 2, Inciso 2
c_90	104792.791603	Art 5, Inciso 3
i_81	72147.4018178	Art 11, Inciso 7

Tabela 2: Conteúdos relevantes negativos

3.2.2 Conteúdo relevante negativo

A Tabela 2 apresenta os conteúdos mais relevantes que representam discordância em relação ao item e ao anteprojeto. Novamente um destaque para os itens do anteprojeto, mas já na quarta posição temos um comentário. Um exemplo de conteúdo relevante negativo é:

i_20 (Art. 5, inciso III)

> Texto: *dados sensíveis: dados pessoais que revelem a origem racial ou étnica, as convicções religiosas, filosóficas ou morais, as opiniões políticas, a filiação a sindicatos ou organizações de caráter religioso, filosófico ou político, dados referentes à saúde ou à vida sexual, bem como dados genéticos;*
>
> 27 comentários, dentre os quais:
>
> 1. andre malveira
>
>   Texto: *Informar dados do tipo, religião, etnia, filosofia, moral, política e vida sexual, não concordo com isso, estaríamos de certa forma cometendo um certo tipo de discriminação (...)*
>
>   Avaliação: discorda do item: verdadeiro positivo. O problema, neste caso, é que esse usuário inseriu esse comentário três vezes, enviesando a base.

3.2.3 Usuários Influentes

A determinação de usuários influentes, como mencionado, depende dos comentários de comentários, os quais materializam algum tipo de formação de opinião. Como o número de tais comentários ainda é pequeno, podemos visualizar o grafo resultante e discutir os resultados encontrados pelo algoritmo. O grafo é apresentado na Figura 2 e a referência dos identificadores de usuários na Tabela 3.

As fontes das arestas direcionadas são os autores dos comentários de comentários; os alvos indicam os autores dos comentários que causaram reação. A Tabela 4 apresenta o ranking dos usuários influentes. Alguns usuários cujos comentários foram referenciados não aparecem como influentes, o que pode ser explicado pela natureza aleatória do algoritmo. Em particular, o fato de que é sorteado o vértice de início do "random surfer"não garante que todos os componentes conectados sejam avaliados, ou mesmo que alguns componentes sejam avaliados mais frequentemente. Como medida paleativa imediata podemos aumentar o número de caminhamentos. Como medida secundária podemos avaliar mecanismos que reduzam a fragmentação do grafo e portanto as distorções no resultado.

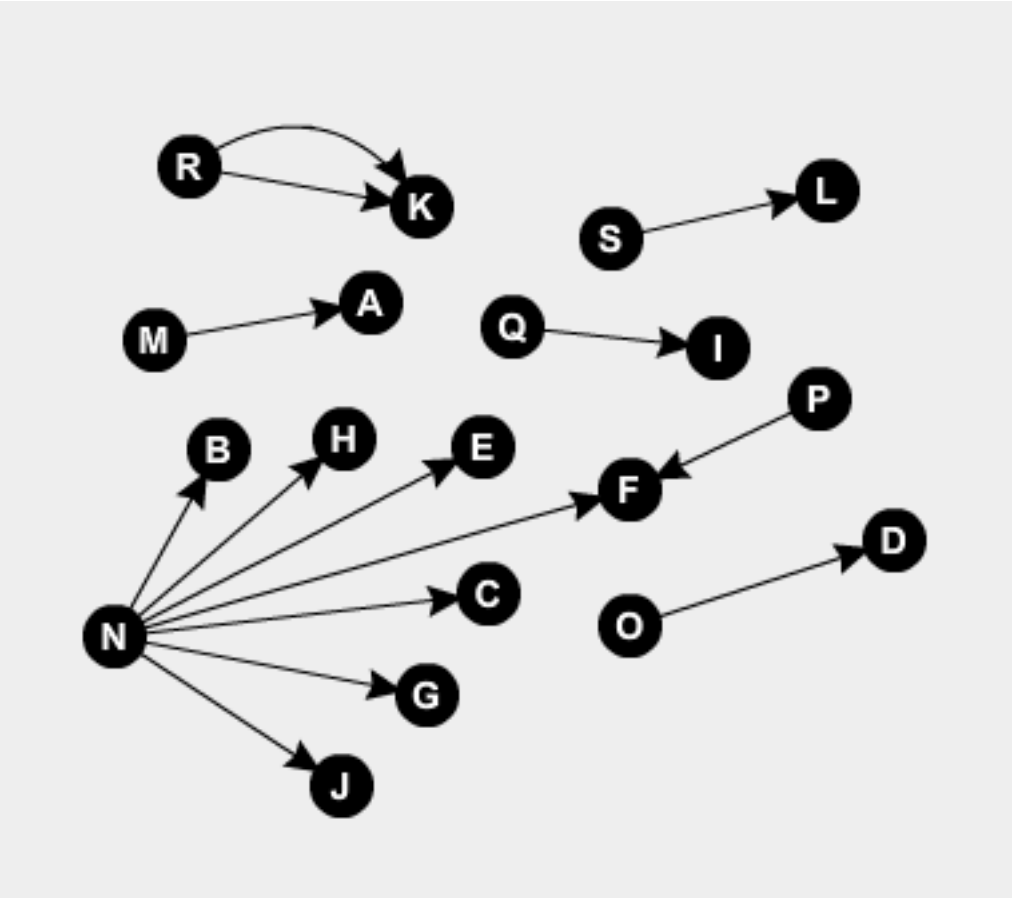


Figura 2: Grafo de Endossos da Consulta sobre Proteção de Dados

Id	Usuário
A	AME
B	Anderson
C	gabriela martins
D	Marcos Baldin
E	Flavio Costa
F	Elizane Gomes
G	clara campos
H	Bruno Diego
I	Lucas Zolet
J	adamir
K	Grupo de Pesquisa em Políticas Públicas para o Acesso à Informação/GPoPAI da USP
L	Wellington Cremasco
M	Eurico Matos
N	TV Aberta + Merchant = Peculato
O	MVianna
P	Marcone
Q	Prof. Marcos
R	joana varon
S	Renata Oliveira

Tabela 3: Indice dos usuários que aparecem no grafo

Usuário	Escore
AME	620689.651687
Anderson	34482.7589376
Lucas Zolet	34482.7589376
Elizane Gomes	34482.7589376
gabriela martins	34482.7589376
clara campos	34482.7589376
Bruno Diego	34482.7589376
Marcos Baldin	34482.7589376
adamir	34482.7589376
Flavio Costa	34482.7589376
Wellington Cremasco	34482.7589376

Tabela 4: Usuários relevantes

4 Próximos Passos

Há três próximos passos imediatos. O primeiro é trabalhar com a base completa, que recentemente foi expandida com dados dos PDFs. O segundo ponto é discutir o que fazer com comentários redundantes, ou seja, o mesmo comentário enviado pelo mesmo usuário mais de uma vez, o que distorce a amostra. O terceiro passo é a melhoria da identificação dos vários tipos de comentários de comentários, aumentando o espectro de comentários e, portanto, usuários que podem ser avaliados.

Referências

[1] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.*, 30(1-7):107–117, 1998.

[2] Massimo Franceschet. Pagerank: standing on the shoulders of giants. *Commun. ACM*, 54(6):92–101, June 2011.

[3] G. Frobenius. Über Matrizen aus nicht Negativen Elementen. 1912.

[4] R. V. Mises and H. Pollaczek-Geiringer. Praktische Verfahren der Gleichungsauflösung. *Zamm-zeitschrift Fur Angewandte Mathematik Und Mechanik*, 9:58–77, 1929.

[5] Arlei Silva, Sara Guimarães, Wagner Meira Jr., and Mohammed Javeed Zaki. Profilerank: finding relevant content and influential users based on information diffusion. In *Proceedings of the 7th Workshop on Social Network Mining and Analysis, SNAKDD 2013, Chicago, IL, USA, August 11, 2013*, pages 2:1–2:9, 2013.