



Universidade de Brasília – UnB  
Faculdade UnB Gama – FGA  
Engenharia de Software

# **Visualização de dados e classificação de perfil em uma plataforma de participação social**

Autor: Naiara Andrade Camelo  
Orientador: Titulação Acadêmica e Fábio Mendes

Brasília, DF  
2019



Naiara Andrade Camelo

## **Visualização de dados e classificação de perfil em uma plataforma de participação social**

Monografia submetida ao curso de graduação em Engenharia de Software da Universidade de Brasília, como requisito parcial para obtenção do Título de Bacharel em Engenharia de Software.

Universidade de Brasília – UnB

Faculdade UnB Gama – FGA

Orientador: Titulação Acadêmica e Fábio Mendes

Coorientador: quando houver, Titulação Acadêmica e Marília

Brasília, DF

2019

---

Naiara Andrade Camelo

Visualização de dados e classificação de perfil em uma plataforma de participação social/ Naiara Andrade Camelo. – Brasília, DF, 2019-  
24 p. : il. (algumas color.) ; 30 cm.

Orientador: Titulação Acadêmica e Fábio Mendes

Trabalho de Conclusão de Curso – Universidade de Brasília – UnB  
Faculdade UnB Gama – FGA , 2019.

1. Palavra-chave01. 2. Palavra-chave02. I. Titulação Acadêmica e Fábio Mendes. II. Universidade de Brasília. III. Faculdade UnB Gama. IV. Visualização de dados e classificação de perfil em uma plataforma de participação social

CDU 02:141:005.6

---

Naiara Andrade Camelo

## **Visualização de dados e classificação de perfil em uma plataforma de participação social**

Monografia submetida ao curso de graduação em Engenharia de Software da Universidade de Brasília, como requisito parcial para obtenção do Título de Bacharel em Engenharia de Software.

Trabalho aprovado. Brasília, DF, 01 de junho de 2013 – Data da aprovação do trabalho:

---

**Titulação Acadêmica e Fábio Mendes**  
Orientador

---

**Titulação e Nome do Professor**  
**Convidado 01**  
Convidado 1

---

**Titulação e Nome do Professor**  
**Convidado 02**  
Convidado 2

Brasília, DF  
2019

**A dedicatória é opcional. Caso não deseje uma, deixar todo este arquivo em  
branco.**

*Este trabalho é dedicado às crianças adultas que,  
quando pequenas, sonharam em se tornar cientistas.*

# Agradecimentos

A inclusão desta seção de agradecimentos é opcional, portanto, sua inclusão fica a critério do(s) autor(es), que caso deseje(em) fazê-lo deverá(ão) utilizar este espaço, seguindo a formatação de *espaço simples e fonte padrão do texto (sem negritos, aspas ou itálico)*.

**Caso não deseje utilizar os agradecimentos, deixar toda este arquivo em branco.**

A epígrafe é opcional. Caso não deseje uma, deixe todo este arquivo em  
branco.

*“Não vos amoldeis às estruturas deste mundo,  
mas transformai-vos pela renovação da mente,  
a fim de distinguir qual é a vontade de Deus:  
o que é bom, o que Lhe é agradável, o que é perfeito.  
(Bíblia Sagrada, Romanos 12, 2)*

# Resumo

O resumo deve ressaltar o objetivo, o método, os resultados e as conclusões do documento. A ordem e a extensão destes itens dependem do tipo de resumo (informativo ou indicativo) e do tratamento que cada item recebe no documento original. O resumo deve ser precedido da referência do documento, com exceção do resumo inserido no próprio documento. (...) As palavras-chave devem figurar logo abaixo do resumo, antecidas da expressão Palavras-chave:, separadas entre si por ponto e finalizadas também por ponto. O texto pode conter no mínimo 150 e no máximo 500 palavras, é aconselhável que sejam utilizadas 200 palavras. E não se separa o texto do resumo em parágrafos.

**Palavras-chaves:** latex. abntex. editoração de texto.



# Abstract

This is the english abstract.

**Key-words:** latex. abntex. text editoration.

# Lista de ilustrações

Figura 1 – Análise de grafos do Pol.is . . . . .	19
Figura 2 – Participantes e vetores do PCA do Pol.is . . . . .	19
Figura 3 – Plataforma do Pol.is . . . . .	20
Figura 4 – ConsiderIt . . . . .	21
Figura 5 – PolitEcho . . . . .	22

## Lista de tabelas

# Lista de abreviaturas e siglas

Fig.	Area of the $i^{th}$ component
TIC	Tecnologia de Informação e Comunicação
LDA	Latent Dirichlet Allocation
EJ	Empurrando Juntas
PCA	Principal Component Analysis ou Análise de Componentes Principais

# Lista de símbolos

$\Gamma$	Letra grega Gama
$\Lambda$	Lambda
$\zeta$	Letra grega minúscula zeta
$\in$	Pertence

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>14</b>
<b>1.1</b>	<b>Justificativa</b>	<b>15</b>
<b>1.2</b>	<b>Objetivos</b>	<b>16</b>
1.2.1	Objetivo Geral	16
1.2.2	Objetivos Específicos	16
<b>1.3</b>	<b>Metodologia de Trabalho</b>	<b>16</b>
<b>1.4</b>	<b>Organização do Trabalho</b>	<b>17</b>
<b>2</b>	<b>VISUALIZAÇÃO DE DADOS DE PLATAFORMAS DE PARTICIPAÇÃO</b>	<b>18</b>
<b>2.1</b>	<b>Pol.is</b>	<b>18</b>
2.1.1	Evolução da visualização do Pol.is	18
<b>2.2</b>	<b>ConsiderIt</b>	<b>20</b>
<b>2.3</b>	<b>PolitEcho</b>	<b>21</b>
<b>3</b>	<b>METODOLOGIA</b>	<b>23</b>
<b>3.1</b>	<b>Distribuição Dirichlet</b>	<b>23</b>
<b>3.2</b>	<b>Naive Bayes</b>	<b>23</b>
<b>3.3</b>	<b>Latent Dirichlet Allocation</b>	<b>23</b>
	<b>REFERÊNCIAS</b>	<b>24</b>

# 1 Introdução

O avanço das tecnologias de informação e comunicação (TIC) vêm mudando a forma de viver das pessoas. As TICs têm tomado diversas áreas das atividades humanas (COLL; MONEREO, 2010) e vêm moldando e marcando a sociedade da informação. O uso de dispositivos, dos mais diversos tamanhos, surpreendem com suas funcionalidades auxiliando nas tarefas do dia-a-dia. (MARTINS, 2005) afirma que distintas práticas sociais estão cada vez mais orientadas por e para essas tecnologias.

As TICs também são reconhecidas como agentes de mudança no setor público e como instrumentos que viabilizam a implementação de processos inovadores na gestão. Se tornando cada vez mais necessária a adaptação do governo ao ambiente digital que tem ganhado espaço entre cidadãos e empresas. Exemplos desses benefícios são as possibilidades de melhorar a comunicação entre governo e cidadãos com mecanismos de participação democrática, colaboração na definição de políticas públicas, entrega de serviço e acesso aos dados públicos. (BARBOSA, 2016)

Já o acesso a internet transformou principalmente a forma de buscar conhecimento e se relacionar. A quantidade de informações encontradas na internet gerou o que se considera, atualmente, o maior acervo de todos os tempos com conteúdo de todos os países, inúmeros textos, imagens e vídeos. E com a evolução dos meios de comunicação em massa, se antes já foi preciso dias até a veiculação de notícias ou informações, hoje são disponibilizadas em menos de segundos devido à tecnologia.

Ainda no século XX uma nova economia surgiu em escala global. (CASTELLS, 1999 apud JAMIL; NEVES, 2000) a chama de informacional, global e em rede. É informacional porque a produtividade e a competitividade dependem da capacidade de gerar, processar e aplicar de forma eficiente a informação baseada em conhecimentos. E desde então a criação e obtenção de dados nos últimos anos supera todos os anos anteriores desde o início da nossa história.

Desde então dados têm se tornado o novo minério de ouro e é impossível não notar sua obtenção por meio de formulários, redes sociais, aplicativos, compras, entre outros. Muitos deles são obtidos por meio de dispositivos com acesso à internet, como celulares. Assim, empresas conseguem obter informações de localização, tipo de dispositivo, o tempo que utilizam e quais serviços e dados pessoais.

Com o aumento contínuo de pessoas usando essas tecnologias, os fornecedores de serviços se depararam com um enorme volume de clientes. Com o desafio de melhorar essa experiência e mantê-los conectados, essas empresas investem na personalização de serviços, tendo como auxílio o uso de algoritmos e matemática, visando o melhor atendimento para

cada usuário.

Serviços como Google e o Facebook são exemplos que evidenciam os diferentes filtros aplicados de forma particularizada, levando em consideração a quantidade de usuários que possuem em todo o mundo. E o uso de algoritmos personalizados, por exemplo, a pesquisa de “células troncos” pode gerar resultados diferentes para um ambientalista e para um executivo. A ampla maioria das pessoas imagina que os mecanismos de busca sejam imparciais. Mas essa percepção talvez se deva ao fato de que esses mecanismos são cada vez mais parciais, adequando-se a visão de mundo de cada um. Cada vez mais, o monitor do computador é um espécie de espelho que reflete os próprios interesses de cada um, baseando-se na análise de cliques feita por observadores algorítmicos. (PARISER, 2012)

O anúncio do Google representou um marco em uma revolução importante, porém quase invisível, no modo como são consumidas as informações. Segundo (PARISER, 2012), em dezembro de 2009, começou a era da personalização. Para ele a internet iria democratizar o planeta, conectando informações e traria uma espécie de utopia global libertadora. Entretanto, os algoritmos se tornaram os curadores da entrega de resultados seguindo essa personalização, por meio de filtros, aumentando o tempo de permanência de um usuário na rede e fazendo com que criadores de conteúdo invistam em conteúdo relevante dentro da rede social para conseguir a atenção das pessoas.

Todos esses comportamentos da sociedade influenciam no surgimento dos filtros bolhas que são as informações que os algoritmos direcionam a pessoas com perfil de interesse parecido. Isso gera no usuário uma sensação de estar cercado de pessoas de opiniões parecidas, distanciando-o assim de “bolhas” diferentes, informações diferentes e pessoas diferentes, bloqueando conhecimentos e evitando discussões.

## 1.1 Justificativa

Este trabalho faz parte do desenvolvimento da plataforma “Empurrando Juntos”, que foi idealizado e desenvolvido inicialmente pelo Laboratório Avançado de Produção Pesquisa e Inovação em Software (LAPPIS), da Universidade de Brasília em conjunto com o Instituto Cidade Democrática, em comum acordo com o Hacklab. Empurrando Juntos é uma plataforma de serviços que possibilita a criação de conversas, e que por meio da utilização de gamificação identifica perfis específicos. Existem diferentes tipos de perfis de utilização, e possibilita poderes temporários com o objetivo de manter a diversidade nos debates.

As conversas são criadas por qualquer pessoa e as demais a recebem em sua conta por meio de notificação push, com três opções disponíveis, sendo elas concordar com a conversa; discordar da conversa; e pular a conversa. Também é permitido comentários,



assim como concordar ou não com comentários já feitos. Essas são as informações de entrada que são utilizadas para a definição de perfis, além de informações opcionais de registro pessoal, como sexo e idade.

A arquitetura se resume em um aplicativo e uma plataforma de serviços em Software Livre que se conecta com aplicativos crowdfsource de participação e utiliza as notificações push para potencializar aspectos como o debate informado, a diversidade de opinião e a ação coletiva.

Este trabalho tem o objetivo de criar os perfis de usuários, com algoritmos que aproximem pessoas com pensamentos próximos, e tenha uma visualização desses dados de forma simples e clara para os usuários. A transparência dessas informações é de grande importância para que as pessoas possam se identificar nas suas bolhas e ter a compreensão do todo e manter debates saudáveis.

## 1.2 Objetivos

### 1.2.1 Objetivo Geral

Este trabalho tem o propósito de classificar os perfis de usuários da plataforma e transparecer a melhor forma de visualização das bolhas de opinião para os usuários do Empurrando Juntos, a fim de fomentar debates e discussões.

### 1.2.2 Objetivos Específicos

- Visualização de dados e classificação de perfil em uma plataforma de participação social;
- Realizar estudo técnico sobre estrutura de dados;
- Realizar estudo técnico sobre algoritmos de personalização;
- Explorar modelos de visualização de bolhas;
- Realizar estudo técnico comparativo das visualizações existentes;
- Formalizar e analisar resultados.

## 1.3 Metodologia de Trabalho

Para o desenvolvimento deste trabalho será necessário mockar dados mais próximos possíveis da realidade para que a visualização atenda aos dados reais.

Para isso será necessário:

- A criação da base de dados por distribuição estatística: distribuição Dirichlet.
- Análise de modelos: Naive Bayes, LDA.
- Análise de visualizações

## 1.4 Organização do Trabalho

Este trabalho está organizado como ...

## 2 Visualização de dados de plataformas de participação

### 2.1 Pol.is

Pol.is é uma das referências deste trabalho. Segundo (FILHO; POPPI, 2017) foi utilizado como uma das referências também de Software Livre para a construção do Empurrando Juntos por abordar as dimensões de governança digital, inclusão e manipulação. É uma plataforma que ajuda as organizações a se entenderem visualizando o que seus membros pensam. Para que tenham uma visão clara de todos os pontos de vista para ajudar a levar a conversa adiante. <sup>1</sup>

Para os criados do Pol.is, o problema era simples: a ineficiência de grandes grupos de pessoas tentando se comunicar efetivamente sobre um determinado tópico on-line. E por isso desenvolveram uma maneira de combinar dados de pesquisa de centenas de pessoas com aprendizado de máquina e visualização interativa de dados. O resultado final é uma maneira simples e limpa para qualquer um, enquanto permite que os usuários estimulem conversas com base em todas as informações. <sup>2</sup>

#### 2.1.1 Evolução da visualização do Pol.is

O objetivo dos criados do Pol.is sempre foi mostrar os grupos de opinião. E no processo de concepção dessa forma de visualização, no início queriam mostrar a “distância” entre os participantes com base em padrões de votação semelhantes e diferentes na conversa como elos, e esse agrupamento emergiria naturalmente disso, entretanto não foi isso que aconteceu. A primeira tentativa foi a visualização por meio de rede de grafos na Fig. 1, entretanto perceberam que tinha muitas informações à mostra, e decidiram não mostrar algumas informações utilizando a matemática. <sup>3</sup>

A utilização de Análise de Componentes Principais (ou PCA em inglês), o que hoje é cerne do Pol.is mostra os dois primeiros principais componentes [img 2] nos eixos x e y. Ocorre a perda de alguns dados na compactação para duas dimensões, mas preserva as maiores diferenças de opinião.

Na Fig. 2 é possível observar que cada participante é mostrado como um ponto preto e os vetores do PCA são expostos na visualização. Clicar nos círculos nos eixos traria

<sup>1</sup> <https://pol.is/home>

<sup>2</sup> <https://www.geekwire.com/2014/startup-spotlight-polis/>

<sup>3</sup> <https://blog.pol.is/the-evolution-of-the-pol-is-user-interface-9b7dccf54b2f>

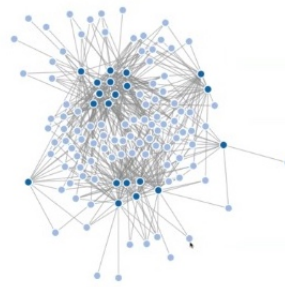


Figura 1 – Análise de grafos do Pol.is

Fonte: <<https://blog.pol.is/the-evolution-of-the-pol-is-user-interface-9b7dccb54b2f>>

os comentários associados àquele vetor - pessoas mais à esquerda, por exemplo, teriam maior probabilidade de concordar com alguma coleção de comentários.

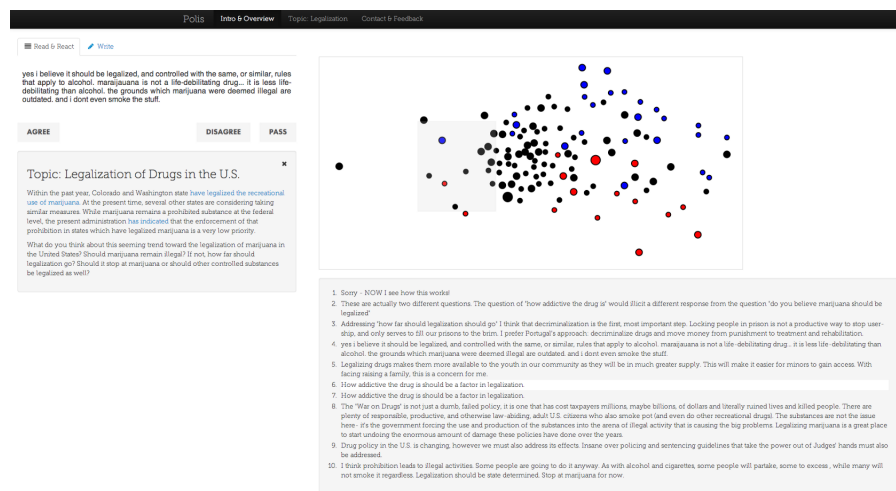


Figura 2 – Participantes e vetores do PCA do Pol.is

Fonte: <<https://blog.pol.is/the-evolution-of-the-pol-is-user-interface-9b7dccb54b2f>>

Para primeiro protótipo, Pol.is utilizou *k-means* aos pontos e eliminou os pontos que tinham menos de um certo número de votos (eles tendiam a se agrupar no centro). Isso melhorou a sensação e começou a transmitir a ideia principal - há grupos de participantes que votaram de maneira semelhante e são um grupo porque compartilham um certo número de perspectivas, não apenas uma.

Dividiu os usuários em forma de seta, dimensionado proporcionalmente e um marcador no mapa, círculo azul, para enfatizar o aspecto espacial. Seguido da tarefa de criar uma correlação mais forte entre o comentário selecionado e o estado da visualização.

A suposição levantada pelos criados sobre o anonimato era muito restritiva. E colocar as pessoas na visualização resolveria todos os tipos de problemas, inclusive tornando a visualização muito mais concreta. O resultado deste trabalho pode ser visto na Fig 3. este foi o resultado

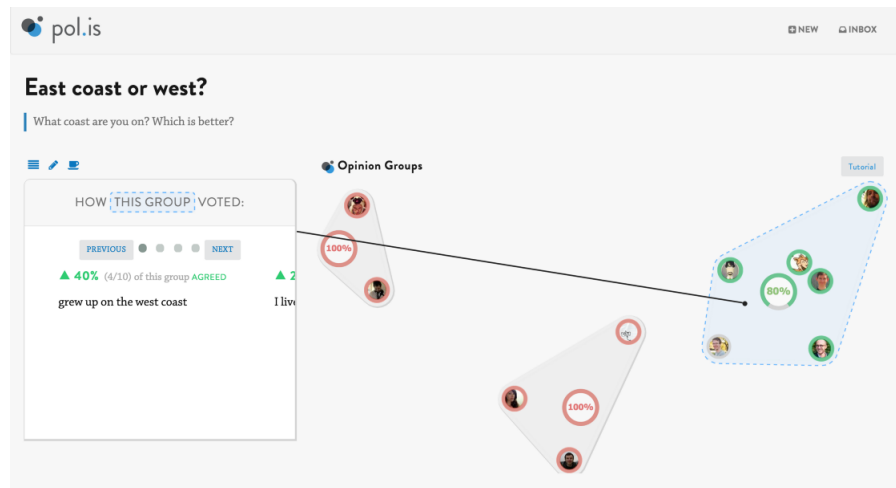


Figura 3 – Plataforma do Pol.is

Fonte: <<https://blog.pol.is/the-evolution-of-the-pol-is-user-interface-9b7dccf54b2f>>

## 2.2 ConsiderIt

ConsiderIt foi criado na Universidade de Washington, como parte da pesquisa de doutorado financiada pela National Science Foundation, com o objetivo de criar um método pelo qual grandes grupos de pessoas pudessem deliberar juntos e encontrar um terreno comum, mesmo em tópicos controversos.<sup>4</sup>

Segundo (KRIPLEAN et al., 2012) a plataforma ConsiderIt pode ajudar a construir a confiança do público por meio de interfaces que encorajam as pessoas a considerar questões e refletir sobre as diversas perspectivas, enquanto é aprimorada a capacidade coletiva de tomar ações mais eficazes como reforma financeira e mudança climática.

ConsiderIt foi construído a partir do básico da deliberação pessoal para promover uma deliberação pública mais eficaz. É focado em fazer as pessoas pensarem sobre as compensações de uma ação proposta, como uma medida em uma eleição, convidando-os a criar uma lista de prós e contras como mostra a Fig 4. Em vez de apenas ter a opção binária de concorda ou não, existe a possibilidade de proporcionalidade de opinião, e a criação das listas com prós e contras.

ConsiderIt reaproveita essas deliberações pessoais para oferecer um guia em evolução para o pensamento público e apresenta as considerações mais notáveis pró e contra baseadas na frequência com que são incluídas e se são incluídas por pessoas com diferentes posições sobre o assunto. Também permite aprofundar os pontos relevantes para diferentes segmentos da população, podendo assim gerar *insights* sobre as considerações de pessoas com diferentes perspectivas, podendo ajudar os usuários a identificar áreas comuns inesperadas.

<sup>4</sup> <https://consider.it/tour?feature=moderation#research>

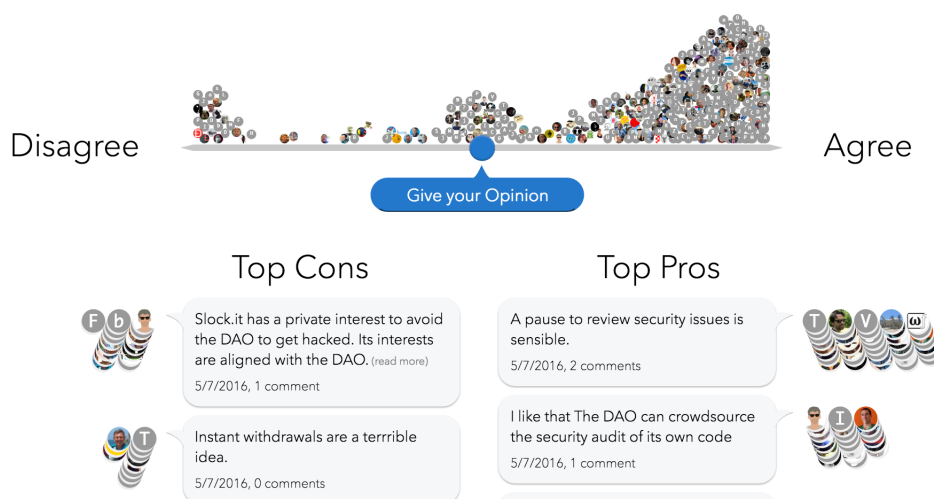


Figura 4 – ConsiderIt

Fonte: <<https://consider.it/examples>>

Também contribui com uma métrica de classificação de pró/contra feita para destacar pontos que ressoam com um público diverso, para promover pontos persuasivos e, ao mesmo tempo, incentivar uma diversidade de pontos de vista e, com sorte, resistir à manipulação estratégica.

## 2.3 PolitEcho

PolitEcho mostra o enviesamento político de amigos do Facebook e *feed* de notícias de um usuário. É uma extensão do Google Chrome que conecta com o Facebook e atribui a cada amigo uma pontuação baseada em uma previsão de tendências políticas e exibe um gráfico da lista de amigos. Em seguida, calcula o viés político no conteúdo do feed de notícias e compara-o com o viés da lista de amigos para destacar possíveis diferenças entre os dois. As cores azul e vermelho representam viés liberal e conservador respectivamente como pode ser visto na Fig 5.

As avaliações políticas dos amigos são baseadas nas páginas políticas do Facebook que eles gostam. É comparada as páginas que os amigos gostaram em um banco de dados de páginas do Facebook que foram classificadas por seu viés liberal/conservador e, é computado uma pontuação com base em quaisquer correspondências.<sup>5</sup>

<sup>5</sup> <https://politecho.org/>

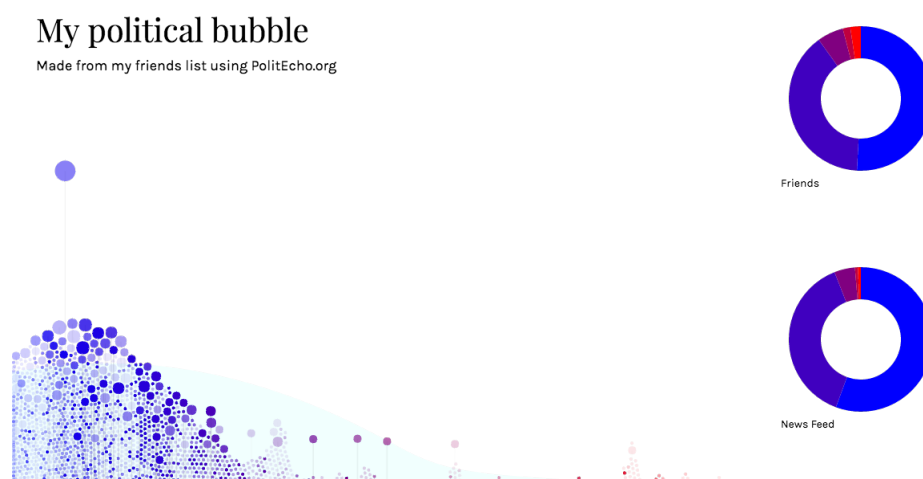


Figura 5 – PolitEcho

Fonte: <<https://politecho.org/>>

## 3 Metodologia

### 3.1 Distribuição Dirichlet

### 3.2 Naive Bayes

### 3.3 Latent Dirichlet Allocation



## Referências

- BARBOSA, A. F. *TIC Governo Eletrônico 2015*. São Paulo: [s.n.], 2016. Disponível em: <[https://cetic.br/media/docs/publicacoes/2/TIC\\_eGOV\\_2015\\_LIVRO\\_ELETRONICO.pdf](https://cetic.br/media/docs/publicacoes/2/TIC_eGOV_2015_LIVRO_ELETRONICO.pdf)>. Citado na página 14.
- CASTELLS, M. *A sociedade em rede*. São Paulo: Paz e Terra, 1999. Citado na página 14.
- COLL, C.; MONEREO, C. Psicologia da educação virtual: aprender e ensinar com as tecnologias da informação e da comunicação. In: \_\_\_\_\_. [S.l.]: Artmed, 2010. cap. Educação e aprendizagem no século XXI: novas ferramentas, novos cenários, novas finalidades, p. 15–46. Citado na página 14.
- FILHO, H. C. P.; POPPI, R. A. Governança digital como vetor para uma nova geração de tecnologias de participação social no Brasil. In: *Liinc*. [s.n.], 2017. v. 13. Disponível em: <[http://www.brapci.inf.br/\\_repositorio/2010/11/pdf\\_d9bd5b50ed\\_0012703.pdf](http://www.brapci.inf.br/_repositorio/2010/11/pdf_d9bd5b50ed_0012703.pdf)>. Citado na página 18.
- JAMIL, G. L.; NEVES, J. T. R. A era da informação: considerações sobre o desenvolvimento das tecnologias da informação. 2000. Disponível em: <[http://www.brapci.inf.br/\\_repositorio/2010/11/pdf\\_d9bd5b50ed\\_0012703.pdf](http://www.brapci.inf.br/_repositorio/2010/11/pdf_d9bd5b50ed_0012703.pdf)>. Citado na página 14.
- KRIPLEAN, T. et al. Supporting reflective public thought with considerit. In: *ACM Conference on Computer Supported Cooperative Work*. Seattle: [s.n.], 2012. Citado na página 20.
- MARTINS, R. X. Competências em tecnologia da informação no ambiente escolar. In: *Psicologia Escolar e Educacional*. [s.n.], 2005. v. 9. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1413-85572005000200016](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-85572005000200016)>. Citado na página 14.
- PARISER, E. *O filtro invisível: O que a internet está escondendo de você*. [S.l.]: Zahar, 2012. ISBN 8537808032. Citado na página 15.