



*Empoderando vidas.
Fortalecendo nações.*

Projeto BRA/12/018 - Desenvolvimento de Metodologias de Articulação e Gestão de Políticas Públicas para Promoção da Democracia Participativa

Produto 03 - Ferramentas assistidas de categorização de conteúdo

com processamento de linguagem natural e de redes complexas, adaptadas para o
ambiente do portal federal de participação social (Participa.br)

Renato Fabbri



Secretaria-Geral da Presidência da República

Produto 03 - Ferramentas assistidas de categorização de conteúdo

Contrato n. 2013/000566

Objeto da contratação: Aporte de conhecimentos e tecnologias para especificação de vocabulário e ferramentas assistidas que utilizam processamento de linguagem natural e análise de redes complexas para o conteúdo do portal da participação social.

Valor do produto: R\$ 10,800 (dez mil e oitocentos reais)

Data de entrega: 28 Julho de 2014

Nome do consultor: Renato Fabbri

Nome da supervisora: Gabriella Vieira Oliveira Gonçalves



Secretaria-Geral da Presidência da República

Fabbri, Renato

Ferramentas assistidas de categorização de conteúdo: com processamento de linguagem natural e de redes complexas, adaptadas para o ambiente do portal federal de participação social (Participa.br) / 2014.

Total de folhas: 37

Supervisora: Gabriella Vieira Oliveira Gonçalves

Secretaria: SNAS

Secretaria-Geral da Presidência da República

Palavras-chave: reconhecimento de padrões, redes complexas, processamento de linguagem natural, participação social.



Esta obra é licenciada sob uma licença Creative Commons - Atribuição-NãoComercial. 4.0 Internacional.



Empoderando vidas.
Fortalecendo nações.

Sumário

1	Introdução	7
1.1	Contexto e importância da consultoria	7
1.2	Contexto e importância do Produto	7
2	Desenvolvimento	8
2.1	Etapas de desenvolvimento	8
2.1.1	Estudo ontológico e triplicação dos dados para API de acesso	8
2.1.2	Instanciação de um Fuseki/Jena e um IPython Notebook	8
2.1.3	Classificação dos textos (mineração de texto / processamento de linguagem natural)	8
2.1.4	Classificação dos agentes pela conectividade (Redes Complexas)	9
2.1.5	Combinação de medidas de RC, PLN e outras	10
2.1.6	Aquisição de dados classificados	10
2.2	Justificativa do método	10
2.3	Justificativa das fontes	11
2.4	Confronto entre os resultados esperados e os alcançados	11
3	Usos dos resultados	11
4	Conclusão	12
4.1	Comentários, sugestões, recomendações	12
4.2	Impacto do Produto para a elaboração, gestão e/ou avaliação de políticas públicas de participação social	13
4.3	Como o Produto deverá impactar o público-alvo das políticas públicas a que se refere	13
	Anexos	16
A	Classificação via etiquetas e personas	16
B	Seleção por ranqueamento e limiar	20
C	Redes de interação e de amizade	23
C.1	Ordenação (ranking) por centralidade	23
C.2	Setores conectivos	32
C.3	Deteção de comunidades	32
D	Classificação de conteúdo via conectividade dos participantes	32



*Empoderando vidas.
Fortalecendo nações.*

E	Teste de conexão com o endpoint SparQL que distribui dados do Partipa.br	35
F	Instâncias online	37



Empoderando vidas.
Fortalecendo nações.

Resumo

Este documento descreve procedimentos selecionados para categorização de conteúdo do portal federal de participação social (Participa.br). O produto relacionado no termo de referência desta consultoria preve propostas de especificações e códigos. Dado o aspecto prático do trabalho, estão descritas também implementações e códigos operantes multiplataforma (linux/mac/windows). Parte deste trabalho é acessível online via HTTP, como os scripts no IPython Notebook e o endpoint SparQL que serve os dados do Participa.br por critérios semânticos.

Palavras-chave: reconhecimento de padrões, redes complexas, processamento de linguagem natural, participação social.



Empoderando vidas.
Fortalecendo nações.

1 Introdução

1.1 Contexto e importância da consultoria

Em confluência com o portal federal de participação social (Participa.br) e o Plano Nacional de Participação Social (PNPS), esta consultoria propõe métodos de classificação e priorização de conteúdo e formas de autorregulação para o portal. O presente produto apresenta uma seleção de métodos para classificação de conteúdo. Dadas a pertinência para o contexto participativo e a simplicidade, são apresentadas a classificação via 1) conectividade dos autores e 2) características dos documentos.

1.2 Contexto e importância do Produto

- Este Produto, através da classificação de conteúdos, visa 1) explicitar propriedades do sistema considerado; 2) permitir observação de conteúdos produzidos por nichos ou características em comum; 3) facilitar a assimilação das informações produzidas pelos participantes.
- Estão planejadas a incorporação destes métodos no funcionamento do Participa.br.
- Os participantes, através da vivência em suas redes, tendem a apropriar-se delas e buscarem processamentos para categorizar os conteúdos e os resumir, como ocorreu com os dashboards de IRC, como acontece com os apps do Facebook e os artigos sobre o Twitter.
- A especialização conectiva dos agentes sociais, e do texto produzido por indivíduos e grupos, é um fenômeno reconhecido. Há aproveitamento destas diferenciações. A entrega deste ferramental ao poder federal e à sociedade, via iniciativa conjunta (Participa.br), capacita a democracia participativa.
 - Permite análise dos dados dos processos, fortalecendo a transparência e desafiando falcatruas. Chega a legitimar o processo democrático online, através da entrega dos dados e dos métodos para analisá-los.
 - Permite a escolha, via critérios públicos e pré-estabelecidos, de articuladores de processos participativos. Esta seleção pode ser acompanhável online e ter como base os rastros digitais de dedicação, como respostas e horas despendidas.
 - Permite a remuneração de horas de dedicação de participantes escolhidos.
 - Permite aproveitamento aberto e conjunto de previsões estatísticas e leis naturais. Por exemplo, para dinâmicas (trilhas) participativas para elaboração conjunta de documentos, mídia, deliberações, coleta de informação (e bens), difusão de informação (e bens).



Empoderando vidas.
Fortalecendo nações.

2 Desenvolvimento

2.1 Etapas de desenvolvimento

2.1.1 Estudo ontológico e triplificação dos dados para API de acesso

Para viabilizar a classificação de conteúdos do portal participativo, em confluência com as propostas de web semântica desta consultoria e do Participa.br, foi necessária uma abordagem ontológica dos aspectos envolvidos no Participa.br, assim como a triplificação dos dados. Para isso, a OPS (Ontologia de Participação Social) foi revisada, com melhoras substanciais e compatibilização com a BFO, também a OPa (Ontologia do Participa.br) foi criada [1, 2]. Já a representação dos dados do Participa.br em triplas RDF envolveu o uso destas e diversas outras ontologias (e.g. FOAF, Dublin Core, SKOS) [3].

2.1.2 Instanciação de um Fuseki/Jena e um IPython Notebook

Os dados triplificados podem ser usados diretamente em algum aplicativo ou programa. A forma padrão de disponibilizar dados em RDF, porém, é através de um endpoint, que prepara os dados na RAM para buscas especificadas via SparQL. Está online um endpoint Jena para consultas SparQL via HTTP. Também uma seleção de scripts em Python, que acessam o endpoint e realizam procedimentos de interesse com os dados, está disponível através de navegadores comuns, como o Firefox ou o Chrom(e,um). **Veja os Anexos A-F.**

2.1.3 Classificação dos textos (mineração de texto / processamento de linguagem natural)

As possibilidades de classificação de conteúdo com base nos textos são inúmeras. Nesta subsubseção, são apontados alguns dos caminhos considerados.

- Com textos previamente classificados, pode-se treinar classificadores automatizados. Este é o “aprendizado supervisionado” de máquina, o tipo de classificação usual para etiquetagem de mensagens e personas. As técnicas em uso são inúmeras (redes neurais, algoritmos genéticos, etc). Para exemplificação, foi implementada uma aprendizagem Bayesiana. No Anexo A consta uma implementação e é o método recomendado para o caso, até mesmo para comparação com outros classificadores que vierem a ser usados.
- A classificação de objetos sem classes previamente definidas, com base somente nas propriedades dos objetos, é conhecido como “aprendizado não supervisionado”. Pode-se impor a existencia de 2 classes (com base no balanço estrutural [4]), ou mais classes, de forma a maximizar a dispersão inter-classe e diminuir a dispersão intra-classe. Este processo pode ser



Empoderando vidas.
Fortalecendo nações.

útil para observar nichos nas atividades, mesmo sem um conjunto de mensagens classificadas de antemão.

- Classificação de mensagens por similaridade às escolhidas. Esta distância pode ser euclidiana no espaço de contagem de palavras, ou calculada via redes semânticas (e.g. wordnet).
- Classificação via contexto similar da palavra ou via simples incidência da palavra. Assemelha-se aos buscadores usuais, com capacidades para lidar com contexto (outras palavras, tipo de autor, classificação da mensagem).
- Ranqueamentos para mensagens, autores e palavras:
 - Mais adjetivos, mais substantivos, mais pontuações, etc. Estes podem ser indicativos de especialização dos agentes, como apontado em [5, 6].
 - Maior tamanho médio das palavras ou variedade de tamanhos (desvio padrão). O Anexo B apresenta uma implementação.
 - Frases mais longas em caracteres ou em palavras, variedade de tamanhos (desvio padrão).
 - Uso de limiares para o ranqueamento, p.ex.: os participantes que mais usam adjetivos (ou escrevem mensagens de mobilização) dentre os que possuem mais de 10 mensagens. O Anexo B apresenta uma implementação.

2.1.4 Classificação dos agentes pela conectividade (Redes Complexas)

- Em geral, as redes formadas com os rastros de atividade digital são: redes de interação ou redes de relações. No participa, há, em especial, a rede de amizades entre os usuários (relações) e as redes de interação: quem responde quem, etc. O Anexo C exibe a formação destas redes.
- Pode-se classificar os usuários por comunidades detectadas nas redes. Veja o Anexo C.3.
- Ranqueamento por centralidade é um dos recursos mais comuns. Há medidas com base na conectividade (grau), na intermediação (betweenness), na proximidade (closeness) e ainda outras medidas de centralidade. O Anexo C.1 exibe ranqueamentos dos participantes por estas medidas.
- As redes sociais, por serem em geral “livres de escala”, possuem especialização dos agentes, canonicamente pensados em “hubs”, “intermediários” e “periféricos”. Estes setores podem ser obtidos com mais propriedade comparando-se o histograma de conectividade da rede real



com uma Erdős-Renyi com o mesmo número de vértices e arestas [5, 6]. No Anexo C.2 está uma implementação em Python para obtenção dos integrantes destes setores com os dados do Participa.br.

2.1.5 Combinação de medidas de RC, PLN e outras

- As medidas de redes e de texto podem ser combinadas para melhorar a qualidade dos classificadores de mensagem e personas. As estabilidades nestas medidas sugerem que haja outliers e uma tipologia pertinente para os agentes [5, 6].
- Medidas de uso do sítio e do perfil do participante podem enriquecer os classificadores.
- Um exemplo de classificação de conteúdo (trecho de texto), com base na classificação conectiva do autor (rede), está no Anexo D.

2.1.6 Aquisição de dados classificados

Para o aprendizado supervisionado (e.g. etiquetação automática de personas e mensagens), é utilizado um conjunto de dados etiquetados de antemão, para “treinar” o classificador. Nas áreas de comunicação e monitoramento, são etiquetadas à mão as mensagens como positivas, negativas e neutras e em outras classes de interesse (e.g. geolocalizações, assuntos). Os autores são classificados em personas (autor masculino, feminino, ativista, militante, curioso, etc). Esta classificação manual pode servir para treinar um classificador público, talvez do Participa.br mesmo, especialmente se revisada por uma ou mais pessoas.

Com alguma frequência, estes classificadores chegam a 80% de acerto, enquanto podem alcançar 95% ou mais com bons dados e investimento em pesquisa e comunidade.

2.2 Justificativa do método

- Classificações mais fundamentais: os métodos utilizados (bag-of-words, aprendizado bayesiano, medidas de grau e betweenness) são as mais usuais, além de facilitar a comparação e estabelecimento de benchmarks, possuem eficiência conhecida e significados mais facilmente compartilhados.
- Amadurecimento com equipe do Participa.br: há outros consultores e integrantes da SG/PR, e da sociedade civil, que compõem ou se comunicam com a equipe do Participa.br. Neste contexto, foram propostas e amadurecidas diversas possibilidades de classificação de conteúdos, na qual foi decantada esta seleção de métodos, apresentada neste Produto.



Empoderando vidas.
Fortalecendo nações.

2.3 Justificativa das fontes

- Pesquisa científica: o autor é pesquisador nas áreas relacionadas com produção bibliográfica em revistas internacionais e em circulação nacional.
- Os frameworks computacionais utilizados (nltk, networkx, rdflib, jena, etc) são amadurecidos no mundo todo, em desenvolvimento aberto, com comunidades dedicadas e pública discussão.
- A equipe do Participa.br é uma equipe da SG/PR voltada para a participação social. Desta equipe provém boa parte dos avanços na participação social.

2.4 Confronto entre os resultados esperados e os alcançados

Este Produto preve “propostas de especificações e códigos” de classificação de conteúdo do Participa.br. Este Produto compreende estas propostas. Há, além disso, alguns resultados alcançados a mais:

- Propostas operantes em códigos online, já integrado aos dados semânticos e disponibilização via endpoint SparQL. Veja IPython e Jena no Anexo F.
- Interfaces/frontends já estudadas para gráficos, reatividade e streaming (meteor+d3). Veja MMISSA, MM, Telões e MyNSA no Anexo F.
- Entrega, através dos resultados dos scripts, de uma breve análise do Participa.br em termos dos rastros digitais, dos conteúdos e dos usuários. Veja os Anexos A-E.

Este documento e os scripts foram reunidos em um repositório git público usual [7].

3 Usos dos resultados

O próximo Produto desta mesma consultoria possui foco na utilização destas classificações. Exemplos de usos estão topificados abaixo.

- Navegação dos conteúdos do portal: facilitar a aquisição das informações de interesse; permite observar o conteúdo com base em características dos participantes (e.g. hub, periférico, intermediário) ou dos conteúdos (e.g. fração de adjetivos ou classificada com rótulos X ou Y).
- Enriquecimento do legado semântico do Participa.br e outros portais: boa parte dos cálculos, necessários para obtenção das estatísticas e classificações, requerem recursos computacionais poderosos e técnicas nada triviais. Assim, os resultados podem ser disponibilizados junto aos dados, em RDF para acesso via consultas SparQL.



Empoderando vidas.
Fortalecendo nações.

- Atribuição de função: através das estatísticas dos grupos, pode-se recompensar atores ou convidá-los para atividades ou funções especiais.
- Resumos: usualmente dashboards, redes ou relações de palavras, visões gerais da entidade de interesse. A entidade pode ser um portal, uma comunidade, um usuário, uma trilha ou uma etapa participativa. Estes resumos são bastante úteis para valorizar as instâncias e orientar os participantes.
- Coleta destas informações para usos/ações: difusão de mídia, consultas, propostas, estudos, deliberações, etc.

4 Conclusão

A categorização de conteúdo do Participa.br pode ser feita de forma distribuída. Os dados, servidos por um endpoint SparQL, podem ser analisados por frontends, como um IPython Notebook, um ScrapperWiki ou um Meteor+d3 para visualizações interativas. Foi testado um intermediário em Flask para servir os dados já formatados para o frontend. Funcionou com serviços gratuitos do Heroku, Meteor e Mongo Labs, embora com limitações e alguns impasses para desenvolvimento em nuvem. Mesmo assim, as consultas podem envolver muitos dados e os processamentos são por vezes intensos, de forma que frontends de análise e aproveitamento devem ser disponibilizados junto aos dados (e.g. IPython Notebook, Meteor+d3 juntos ao endpoint Fuseki/Jena). Os Anexos A-F ao final deste documento, e o repositório git [7], são os resumos principais do Produto.

4.1 Comentários, sugestões, recomendações

- Para boas aplicações de classificadores de conteúdo, é necessária uma quantidade grande de conteúdo classificado previamente, geralmente à mão. Assim, é pertinente a etiquetagem das mensagens com os parceiros da comunicação, para liberação junto aos dados semânticos e treino de classificadores.
- Como pode-se notar no Anexo B, há bastante mensagem repetida. Embora este problema tenha sido tratado na utilização dos dados, é recomendado que seja corrigido o problema nos dados do Participa.br. Uma boa solução, embora não ideal, é a correção destes dados na triplificação.



Empoderando vidas.
Fortalecendo nações.

4.2 Impacto do Produto para a elaboração, gestão e/ou avaliação de políticas públicas de participação social

- Este produto facilita a apropriação dos processos participativos através da categorização de conteúdos e observação de suas características.
- Este produto explicita a entrega das informações para a população, para observação distribuída.
- Este produto introduz na participação social brasileira tecnologias livres de redes sociais, redes complexas, processamento de linguagem natural e web semântica.
- Este produto disponibiliza estes algoritmos de forma executável em navegadores HTTP comuns, como Firefox ou Chrome(um).
- Este produto entrega uma instância operante de acesso aos dados do Participa.br, em formato RDF e enriquecidos semanticamente.

4.3 Como o Produto deverá impactar o público-alvo das políticas públicas a que se refere

- Permitindo navegação seletiva pelos conteúdos disponíveis.
- Valorizando as instâncias e as tornando mais informativas, com resumos estatísticos e visuais.
- Explicitando propriedades dos processos participativos e usos destas propriedades.
- Integrando o portal federal de participação social (Participa.br) ao legado humano de dados linkados (via critérios semânticos).
- Permitindo critérios funcionais para atribuição de papéis para participantes. Por exemplo, a construção de um manifesto ou resumo final pode ser feito requisitando: de periféricos, os substantivos; de hubs, os adjetivos; e de intermediários, que formem o texto com aquelas palavras. Outra possibilidade é a remuneração de hubs pela participação efetuada ou a convocação de periféricos para oxigenar o processo participativo.
- Aproximando perfis técnicos pela qualidade das tecnologias utilizadas, pela da relevância dos dados sobre participação social e pela pertinência dos métodos.



Empoderando vidas.
Fortalecendo nações.

Referências

- [1] “Ontologia de participação social,” <http://tinyurl.com/p2doueu>.
- [2] “Ontologia do participa,” <http://tinyurl.com/lcccwop>.
- [3] “Especificação da triplificação dos dados do participa.br,” <http://tinyurl.com/k74z3yl>.
- [4] D. Easley and J. Kleinberg, *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press, 2010.
- [5] “Stability in human interaction networks: primitive typology of vertex, prominence of measures and activity statistics,” <http://tinyurl.com/o25dcb7>.
- [6] “Elementary observations of textual production in interaction networks,” <http://tinyurl.com/owsmzfv>.
- [7] *Repositório Git do produto 3: documento e scripts*. <https://github.com/ttm/pnud3>.



Empoderando vidas.
Fortalecendo nações.

Abreviações e jargão

RC: Redes Complexas

PLN: Processamento de Linguagem Natural

OPS: Ontologia de participação Social

OPA: Ontologia do Participa.br

MMISSA: Monitoramento Massivo e Interativo da Sociedade pela Sociedade para Aproveitamento

AARS: A Análise de Redes Sociais

MyNSA: Monitoring yields Natural Streaming and Analysis

PNPS: Plano Nacional de Participação Social

RDF: Resource Description Framework

HTTP: Hypertext Transfer Protocol

SPARQL: Simple Protocol and RDF Query Language

endpoint SPARQL: ponto de acesso, geralmente HTTP, a dados em RDF via buscas em SPARQL.

Participa.br: Portal federal de participação social.

IPython Notebook: instância online para rodar scripts Python

Mateor: arcabouço para páginas reativas e com funcionamento distribuído.

D3js: biblioteca de visualização de dados.



*Empoderando vidas.
Fortalecendo nações.*

Anexos

A seguir estão trechos de código Python que executam funções básicas de classificação de conteúdo do Participa.br ou úteis para isso. Estes códigos estão executáveis online, acessíveis nos browsers web usuais (firefox, chrome(um)), através de um IPython Notebook. Mais informações no Anexo F.

Anexo A: Classificação de conteúdo via mensagens rotuladas e personas

Importação das bibliotecas para acesso aos dados enriquecidos semanticamente:

```
In [70]: from SPARQLWrapper import SPARQLWrapper, JSON
import time
```

Definição de prefixos úteis para as consultas SPARQL:

```
In [71]: PREFIX=""
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX ops: <http://purl.org/socialparticipation/ops#>
PREFIX opa: <http://purl.org/socialparticipation/opa#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX tsioc: <http://rdfs.org/sioc/types#>
PREFIX schema: <http://schema.org/>""
```

Buscando todos os comentários no endpoint SparQL Fuseki/Jena:

```
In [72]: NOW=time.time()
q="SELECT ?comentario ?titulo ?texto WHERE \
      {?comentario dc:type tsioc:Comment.\
      OPTIONAL {?comentario dc:title ?titulo . }\
      OPTIONAL {?comentario schema:text ?texto .}}\"
sparql3 = SPARQLWrapper("http://localhost:82/participabr/query")
sparql3.setQuery(PREFIX+q)
sparql3.setReturnFormat(JSON)
results4 = sparql3.query().convert()
print("%.2f segundos para puxar todos os comentários do Participa.br"%
      (time.time()-NOW,))
```

2.43 segundos para puxar todos os comentários do Participa.br

Removendo pontuação e fazendo lista de palavras. São muitas mensagens repetidas, dois tipos:

- uma possui "teste de stress" no título, outra possui
- "comunidade de desenvolvedores e nesse caso, quanto mais" no corpo da mensagem.

Ambas as mensagens são removidas.

```
In [87]: msgs_=results4["results"]["bindings"]
msgs=[mm for mm in msgs_ if ("titulo" not in mm.keys()) or
      (("teste de stress" not in mm["titulo"]["value"].lower())
      and ("comunidade de desenvolvedores e nesse caso, quanto mais"
      not in mm["texto"]["value"].lower()))]
NOW=time.time()
import string, nltk as k
exclude = set(string.punctuation+u'\u201c'+u'\u2018'+u'\u201d'+u'\u2022'+u'\u2013')
palavras=string.join([i["texto"]["value"].lower() for i in msgs])
palavras = ''.join(ch for ch in palavras if ch not in exclude)
palavras_=palavras.split()
print(u"feita lista de todas as palavras de todos os comentários em %.2fs"%
      (time.time()-NOW,))
```

feita lista de todas as palavras de todos os comentários em 0.19s

Removendo stopwords e fazendo contagem das palavras restantes:

```
In [83]: NOW=time.time()
stopwords = set(k.corpus.stopwords.words('portuguese'))
palavras__=[pp for pp in palavras_ if pp not in stopwords]
fdist_=k.FreqDist(palavras__)
print("retiradas stopwords e feita contagem das palavras em %.2fs"%
      (time.time()-NOW,))
for fd,ii in [(fdist_[i],i) for i in fdist_.keys()[:14]]: print fd, ii

retiradas stopwords e feita contagem das palavras em 0.29s
1277 é
1256 não
762 ser
717 participação
548 social
526 sociedade
468 à
459 sobre
367 governo
357 são
337 forma
327 políticas
310 públicas
302 brasil
```

```
In [75]: print(u"são %i palavras em %i palavras diferentes"%(len(palavras__),len(fdist_)))

são 91361 palavras em 14653 palavras diferentes
```

```
In [76]: # para radicalizar (lematização é similar)
# NOW=time.time()
#stemmer = k.stem.RSLPStemmer()
#palavras__=[stemmer.stem(pp) for pp in palavras_]
#fdist__=k.FreqDist(palavras__)
#print("feita freq dist (radicalizada) em %.2f"%(time.time()-NOW,))
```

Escolhendo as palavras mais frequentes para fazer caracterização das mensagens:

```
In [77]: # escolhendo as 200 palavras mais frequentes
palavras_escolhidas=fdist_.keys()[:200]
```

Extraindo atributos (contagem das palavras) e fazendo classificação bayesiana ingênua. Note que os rótulos "pos" e "neg" estão sendo atribuídos ao acaso. Para aproveitamento, é necessário que sejam usados os dados rotulados, provavelmente rotulados pelo pessoal da comunicação.

```
In [78]: def document_features(documento):
          features={}
          for palavra in palavras_escolhidas:
              features["contains(%s)"%(palavra,)]=(palavra in documento)
          return features
msgsP= [(rr["texto"]["value"],"pos") for rr in msgs[:500]]
msgsN=[(rr["texto"]["value"],"neg") for rr in msgs[500:1000]]
msgsT=msgsP+msgsN
random.shuffle(msgsT)
feature_sets=[(document_features(msg[0]),msg[1]) for msg in msgsT]
train_set, test_set = feature_sets[:500], feature_sets[500:]
```

```
classifier = k.NaiveBayesClassifier.train(train_set)
```

Mostrando as características mais informativas:

```
In [79]: classifier.show_most_informative_features(5)
```

Most Informative Features

contains(comitê) = True	pos : neg	=	4.1 : 1.0
contains(hoje) = True	pos : neg	=	4.0 : 1.0
contains(saúde) = True	neg : pos	=	3.1 : 1.0
contains(sugiro) = True	neg : pos	=	2.7 : 1.0
contains(grupo) = True	pos : neg	=	2.4 : 1.0

Precisão nos dados de teste dummy ≈ 0.5 , pois os rótulos não foram atribuídos com coerência (classificações de antemão de um conjunto de mensagens por conhecedores):

```
In [80]: k.classify.accuracy(classifier, test_set)
```

```
Out[80]: 0.526
```

Classificação de um documento:

```
In [86]: classifier.classify(document_features(msgsT[12][0]))
```

```
Out[86]: 'neg'
```

|||--- FIM ---|||

Anexo B: Seleção de conteúdo por ranqueamento (e.g. por tamanho de palavras) e limiar (e.g. número mínimo de palavras)

Importanto bibliotecas úteis:

```
In [68]: from SPARQLWrapper import SPARQLWrapper, JSON
import time, string, numpy as n, nltk as k
```

Definição de prefixos úteis para as consultas SparQL:

```
In [55]: PREFIX=""PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX ops: <http://purl.org/socialparticipation/ops#>
PREFIX opa: <http://purl.org/socialparticipation/opa#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX tsioc: <http://rdfs.org/sioc/types#>
PREFIX schema: <http://schema.org/>""
```

Buscando todos os comentários no endpoint SparQL Fuseki/Jena com os dados do Participa.br:

```
In [56]: NOW=time.time()
q="SELECT ?comentario ?titulo ?texto WHERE \
      {?comentario dc:type tsioc:Comment.\
      OPTIONAL {?comentario dc:title ?titulo . }\
      OPTIONAL {?comentario schema:text ?texto .}}"
sparql3 = SPARQLWrapper("http://localhost:82/participabr/query")
sparql3.setQuery(PREFIX+q)
sparql3.setReturnFormat(JSON)
results4 = sparql3.query().convert()
print("%.2f segundos para puxar todos os comentários do Participa.br"%(time.time()-NOW,))
```

2.34 segundos para puxar todos os comentários do Participa.br

Limpando mensagens das sugeiras do BD:

```
In [57]: msgs=results4["results"]["bindings"]
msgs=[mm for mm in msgs_ if ("titulo" not in mm.keys() or (("teste de stress" not in mm["titulo"]["value"].lower())
and ("comunidade de desenvolvedores e nesse caso, quanto mais" not in mm["texto"]["value"].lower())))]
```

Fazendo função para extrair atributos das mensagens:

```
In [58]: import string, numpy as n
exclude = set(string.punctuation+u'\u201c'+u'\u2018'+u'\u201d'+u'\u2022'+u'\u2013')
def atributos(_msg):
    texto=_msg["texto"]["value"]
    texto_ = ''.join(ch for ch in texto if ch not in exclude)
    palavras=texto_.split()
    tams=[]
    for palavra in palavras:
        tams.append(len(palavra))
    return len(tams), n.mean(tams), n.std(tams)
```

Criando vetor de atributos de cada mensagem:

```
In [66]: atrs=[atributos(msg) for msg in msgs]
atrs_=n.array(atrs)
```

Fazendo seleção das mensagens que possuem entre 110 e 115 palavras:

```
In [67]: max_palavras=115
min_palavras=110
n_msgs=((atrs[:,0]>min_palavras)*(atrs[:,0]<max_palavras)).sum()
print(u"são %i mensagens com mais de %i palavras e menos de %i"%
      (n_msgs, min_palavras, max_palavras) )
```

são 26 mensagens com mais de 110 palavras e menos de 115

Selecionando mensagens com média do tamanho da palavra maior que 6 caracteres:

```
In [61]: msgs_i=((atrs[:,0]>min_palavras)*(atrs[:,0]<max_palavras)).nonzero()[0]
textos=[msgs[i]["texto"]["value"] for i in msgs_i][j]
for j in (atrs[msgs_i][:,1]>6.).nonzero()[0]
for texto in textos: print texto+"\n=====
```

1. Auxiliar na divulgação das informações jurídicas para estrangeiros por meio de projetos de longo prazo dentro no âmbito

1. Realizar na divulgação das informações jurídicas para estrangeiros por meio de projetos de longo prazo dentro no âmbito das próprias universidades (ex. immigration/refugee law clinics), estimulando-se a parceria com as instituições, organizações e programas/projetos relacionados à temática migratória;

2. Inserir na nova política migratória a capacitação/formação específica para agentes públicos (obrigatória) que atuam ou possam vir a atuar em razão de suas atribuições institucionais com migrantes, refugiados e apátridas. Nesse sentido, a experiência e contribuição da academia será fundamental do ponto de vista da expertise científica (ensino e pesquisa - colaboração para o conteúdo programático da capacitação), além de reforçar a importância do ensino do Direito Internacional e Direitos Humanos nas Universidades.

=====

ROSAS, Agostinho da Silva; MELO NETO, José Francisco de. Educação Popular: Enunciados Teóricos. João Pessoa: Editora Universitária da UFPB, 2008.

Os autores buscam delimitar um campo da educação em que seja possível delinear-se características que apenas à educação popular façam parte. Trata-se de um esforço de apresentação de constituintes com dimensões formadoras, tendo como aspectos centrais as experiências de vários educadores populares e as reflexões em desenvolvimento na época. Nessa perspectiva, são abordadas a experiência histórica, a cultura, o popular, a realidade, o trabalho, a autonomia, a liberdade e a igualdade como componentes fundantes para a realização de práticas em educação popular, lastreados pela dimensão ética do diálogo.

Disponível em:

http://www.prac.ufpb.br/copac/extelar/producao_academica/livros/pa_l_2004_educacao_popular_enunciados_teoricos.pdf

=====

Selecionando mensagens com média do tamanho da palavra menor que 5 caracteres:

```
In [62]: textos=[msgs[i]["texto"]["value"] for i in msgs_i][j]
          for j in (attrs[msgs_i][:,1]<5.).nonzero()[0]
          for texto in textos: print texto+"\n====="
```

Seria suave aos ouvidos ver que a população participando e contribuindo pode notar que seus desejos desde que na medida do possível estão sendo ouvidos e colocado em prática.

A população cansou, pois até coloca sua cara pra bater, porém na hora de colocar em prática o que tanto desejam nada acontece.

O que resta, a saber, se quando a população participar da construção das políticas públicas, deixara suas vontades o campo da teoria para o campo da prática, muito difícil isso ocorrer ou talvez impossível, pois o desejo da população praticamente em todos as vezes e momentos que é aferida suas vontades, são deixados de lado, isso é uma pena.

=====

O §1º é a atividade ou ação natural, o principio de ofício ou por assim dizer a o fato singular do GT.

O §2º é a preposições facultadas que podem ou não advir, em conformidade das necessidades observadas na decorrência do §1º, ou seja, dependente do §1º.

O §3º é mais fácil, pois não há registros do que não aconteceu, então a referencia, por lógica, deve ser a ultima.

Não me ative em pensar em mesmo método com os incisos, por serem muitos e não existe espaço para estabelecer a desfiguração, observando que o contexto das determinações estarem em elevação de importância, mesmo que a ordenação ou 'cronologia' traga maior facilidade de compreensão.

=====

Laura, essa alteração legislativa é muito importante. É muito desproporcional exigir que o migrante tenha que indicar uma repartição consular em outro país para obter o visto. O fundamento está no fato de que o Estatuto do Estrangeiro não permite concessão de visto a quem está irregular. Mas é possível uma interpretação sistemática e razoável da norma, para que seja feita uma distinção para os estrangeiros indocumentados (que estão ainda sem o visto mas que têm o direito à regularização migratória, como por exemplo por motivo de reunião familiar). Exigir que o estrangeiro que já está no Brasil tenha que sair do país apenas para buscar o visto é muito oneroso e desproporcional.

=====

Fazendo contagem das palavras mais frequentes para seleção:

```
In [69]: NOW=time.time()
          palavras=string.join([i["texto"]["value"].lower() for i in msgs])
          exclude = set(string.punctuation+u'\u201c'+u'\u2018'+u'\u201d'+u'\u201e'+u'\u2022'+u'\u2013')
          palavras = ''.join(ch for ch in palavras if ch not in exclude)
          palavras_=palavras.split()
          print(u"feita lista de todas as palavras de todos os comentários em %.2f"%(time.time()-NOW,))
          stopwords = set(k.corpus.stopwords.words('portuguese'))
          palavras__=[pp for pp in palavras_ if pp not in stopwords]
          fdist=k.FreqDist(palavras__)
          print("retiradas stopwords feita contagem das palavras em %.2f"%(time.time()-NOW,))
```

feita lista de todas as palavras de todos os comentários em 0.18

retiradas stopwords feita contagem das palavras em 0.44

Fazendo seleção das 14 palavras mais incidentes nos comentários do Participa.br:

```
In [70]: for fd,ii in [(fdist_[i],i) for i in fdist_.keys()[:14]]: print fd, ii
```

1277 é
1256 não
762 ser
717 participação
548 social
526 sociedade
468 à
459 sobre
367 governo
357 são

7/30/2014

2. Seleção por ranqueamento e limiar

337 forma
327 políticas
310 públicas
302 brasil

|||-- FIM --|||

Anexo C: Síntese das redes de amizade e de interação do Participa.br

Importando bibliotecas:

```
In [89]: from SPARQLWrapper import SPARQLWrapper, JSON
import time, numpy as n, networkx as x
```

Definição de prefixos úteis para as consultas SparQL:

```
In [90]: PREFIX=""PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX ops: <http://purl.org/socialparticipation/ops#>
PREFIX opa: <http://purl.org/socialparticipation/opa#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX tsioc: <http://rdfs.org/sioc/types#>
PREFIX sioc: <http://rdfs.org/sioc/ns#>
PREFIX schema: <http://schema.org/>""
```

Buscando amizades do Participa.br:

```
In [91]: NOW=time.time()
q=""SELECT DISTINCT ?aname ?bname
WHERE {
    ?a foaf:knows ?b .
    ?a foaf:name ?aname .
    ?b foaf:name ?bname .
}""
sparql3 = SPARQLWrapper("http://localhost:82/participabr/query")
sparql3.setQuery(PREFIX+q)
sparql3.setReturnFormat(JSON)
results4 = sparql3.query().convert()
print("%.2f segundos para puxar todas as amizades do Participa.br"%
      (time.time()-NOW,))
```

0.07 segundos para puxar todas as amizades do Participa.br

Erigindo rede de amizades como um grafo não direcionado:

```
In [92]: g=x.Graph()
for amizade in results4["results"]["bindings"]:
    nome1=amizade["aname"]["value"]
    nome2=amizade["bname"]["value"]
    g.add_edge(nome1,nome2)
```

```
In [93]: print(u"são %i amizades entre %i pessoas no Participa.br"%
      (g.number_of_edges(), g.number_of_nodes()))
```

são 910 amizades entre 443 pessoas no Participa.br

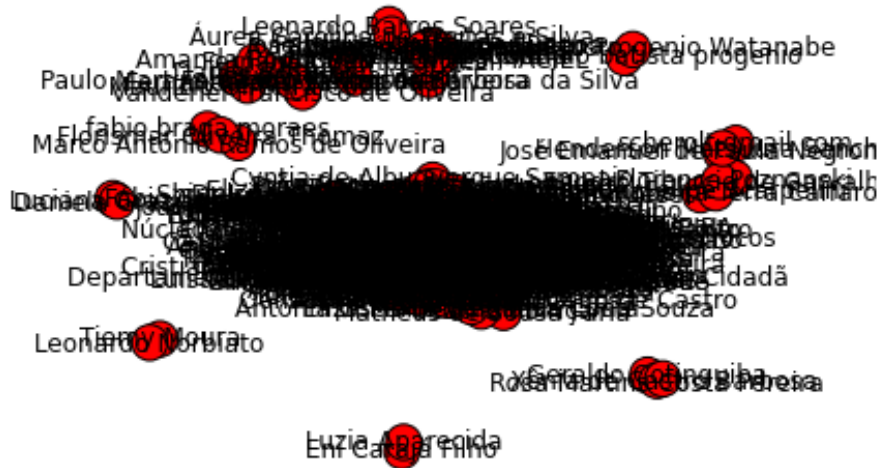
```
In [141]: x.write_graphml(g,"amizadesParticipa.graphml")
```

Para visualizar e analisar, a estrutura de amizades está pronta neste link:

<https://github.com/ttm/pnud3/blob/master/amizadesParticipa.graphml>.

Usos são imediatos, p.ex:

```
In [94]: x.draw(g,pos=x.layout.fruchterman_reingold_layout(g))
```



Puxando as interações de comentários do Participa.br:

```
In [95]: NOW=time.time()
q2="""SELECT DISTINCT ?aname ?bname
WHERE {
    ?comentario dc:type tsioc:Comment.
    ?participante1 ops:performsParticipation ?comentario.
    ?participante1 foaf:name ?aname.
    ?artigo sioc:has_reply ?comentario.
    ?participante2 ops:performsParticipation ?artigo.
    ?participante2 foaf:name ?bname.
}"""
sparql3.setQuery(PREFIX+q2)
sparql3.setReturnFormat(JSON)
results = sparql3.query().convert()
print("%.2f segundos para puxar as interações do Participa.br"%
      (time.time()-NOW,))
```

15.58 segundos para puxar as interações do Participa.br

Sintetizando rede direcionada de interação:

```
In [96]: d=x.DiGraph()
for interacao in results["results"]["bindings"]:
    nome_chegada=interacao["aname"]["value"]
    nome_partida=interacao["bname"]["value"]
    if (nome_partida,nome_chegada) in d.edges():
        d[nome_partida][nome_chegada]["weight"]+=1
```



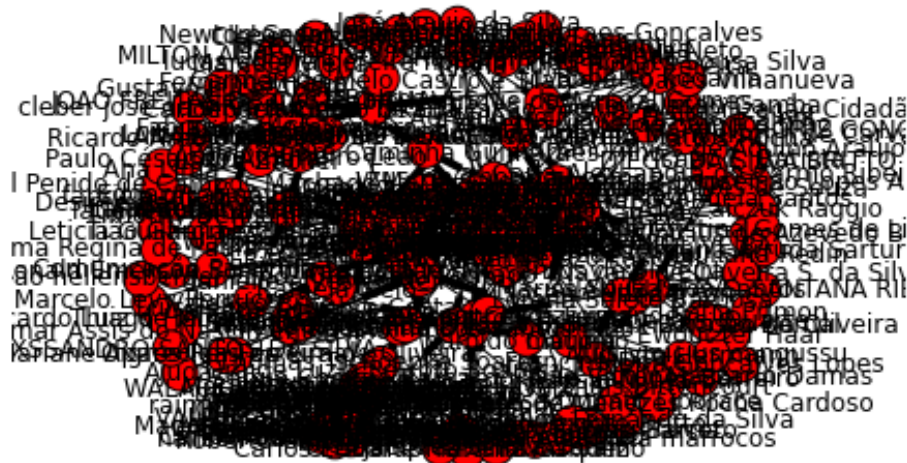
```
else:
    d.add_edge(nome_partida,nome_chegada,weight=1.)
```

```
In [142]: x.write_graphml(d,"interacoesParticipa.graphml")
```

Estrutura de interação pronta para análises e visualizações neste link:

<https://github.com/ttm/pnud3/blob/master/interacoesParticipa.graphml>.

```
In [97]: x.draw(d,pos=x.layout.fruchterman_reingold_layout(d))
```



C.1. ordenação (ranking) por centralidade

Elencando as 15 pessoas mais conectadas via **interação**:

```
In [143]: import operator
sorted_d = sorted(d.degree().iteritems(), key=operator.itemgetter(1))
sorted_d[::-1][:15]
```

```
Out[143]: [(u'Portal', 46),
            (u'Renato Fabbri', 26),
            (u'Gabriela Valle', 19),
            (u'Hylton Sarcinelli Luz', 15),
            (u'F\xeldia Rebou\xe7as', 13),
            (u'Grazielle Machado', 12),
            (u'Jose mendon\xe7a Furtado Neto', 11),
            (u'andre luiz da silva', 11),
            (u'Daniel Pitangueira de Avelino', 10),
            (u'Henrique Parra Parra Filho', 9),
            (u'Juliano Geraldi', 8),
            (u'Roberto Kodama', 7),
            (u'Marcelo Rodrigues Saldanha da Silva', 7),
            (u'L\xeddia Maria Alves Pereira', 7),
            (u'Frank Lane', 7)]
```

Elencando as 15 pessoas mais conectadas via **amizades**:

```
In [100]: sorted_g = sorted(g.degree().iteritems(), key=operator.itemgetter(1))
sorted_g[::-1][:15]
```

```
Out[100]: [(u'Marcelo Branco', 122),
(u'Maria Jos\xea9lia Amaral de Menezes', 60),
(u'Ana C\xea9lia Costa', 36),
(u'Vicente Aguiar', 35),
(u'Laura Zacher', 33),
(u'Luis Felipe Coimbra Costa', 32),
(u'Ricardo Poppi', 31),
(u'LUCAS MOREIRA DE SOUSA', 30),
(u'Ronald Emerson Scherolt da Costa', 29),
(u'Valessio Brito', 27),
(u'Grazielle Machado', 20),
(u'Andr\xea9 Filipe de Assun\xea7\xea3o e Brito', 20),
(u'Renato Fabbri', 20),
(u'Daniela Feitosa', 18),
(u'Daniel Pitangueira de Avelino', 18)]
```

Para medida de centralidade, as listagens acima usam quantidade de conexão, ou grau.

Medidas de interesse imediato são: de atravessador (presente em mais geodésicas do grafo, *betweenness*), de mais "perto de todos os outros" (*closeness*) e de auto-vetor:

```
In [105]: bc=x.betweenness_centrality(g)
cc=x.closeness_centrality(g)
```

```
In [106]: sorted_bc = sorted(bc.iteritems(), key=operator.itemgetter(1))
sorted_bc[::-1][:15]
```

```
Out[106]: [(u'Marcelo Branco', 0.42882163196866385),
(u'Maria Jos\xea9lia Amaral de Menezes', 0.1500460662666134),
(u'LUCAS MOREIRA DE SOUSA', 0.09201604347662772),
(u'Ana C\xea9lia Costa', 0.07270058962510031),
(u'Grazielle Machado', 0.060658560472243744),
(u'Laura Zacher', 0.05514913696577906),
(u'Valessio Brito', 0.05235162430751249),
(u'Nilzete Pacheco ', 0.049895256955160554),
(u'C\xelssio N. Martin', 0.04578534023146533),
(u'Jesulino Alves de Souza', 0.04285934737795151),
(u'Vicente Aguiar', 0.042734420429139115),
(u'Daniel Pitangueira de Avelino', 0.03954188318259412),
(u'Luis Felipe Coimbra Costa', 0.036605448228684445),
(u'kelly cristina alves', 0.035831717510820746),
(u'Ricardo Poppi', 0.034899619183799586)]
```

```
In [107]: sorted_cc = sorted(cc.iteritems(), key=operator.itemgetter(1))
sorted_cc[::-1][:15]
```

```
Out[107]: [(u'Marcelo Branco', 0.44299145920610333),
(u'Maria Jos\xea9lia Amaral de Menezes', 0.3820683267566498),
(u'Vicente Aguiar', 0.36163480373131945),
```

```
(u'Ronald Emerson Scherolt da Costa', 0.36090643554656354),
(u'Grazielle Machado', 0.35873882932706463),
(u'Ana C\xe9lia Costa', 0.35695228137224855),
(u'Valessio Brito', 0.3562426346896),
(u'Luis Felipe Coimbra Costa', 0.34895821859565485),
(u'Ricardo Poppi', 0.34895821859565485),
(u'Jesulino Alves de Souza', 0.3469313557577324),
(u'Paulo Henrique de Lima Santana', 0.3442652166164626),
(u'Daniela Feitosa', 0.3384136831895539),
(u'Paulo Meirelles', 0.3384136831895539),
(u'Laura Zacher', 0.33524797988562915),
(u'Andr\xe9 Filipe de Assun\xe7\xe3o e Brito', 0.3346219332378502)]
```

Para as redes de interação:

```
In [108]: bcd=x.betweenness centrality(d)
          ccd=x.closeness centrality(d)
```

```
In [109]: sorted_bcd = sorted(bcd.iteritems(), key=operator.itemgetter(1))
          sorted_bcd[::-1][:15]
```

```
Out[109]: [(u'Renato Fabbri', 0.027219251336898394),
           (u'Grazielle Machado', 0.019025092554504318),
           (u'Ricardo Poppi', 0.013204442616207322),
           (u'andre luiz da silva', 0.008021390374331552),
           (u'Hylton Sarcinelli Luz', 0.007853078294254766),
           (u'JO\x30 BATISTA CAVALCANTI', 0.007260386672151378),
           (u'Daniel Pitangueira de Avelino', 0.007198683669271905),
           (u'Jose mendon\xe7a Furtado Neto', 0.004669546140134375),
           (u'Jorge', 0.004668860551213492),
           (u'Caroline Burle', 0.0046482928835870005),
           (u'Gabriela Valle', 0.004324009324009324),
           (u'Tiago Henrique Ferreira Zoroastro', 0.0033730974907445494),
           (u'Vicente Aguiar', 0.003331962155491567),
           (u'Frank Lane', 0.0031262854792266552),
           (u'edison cardoso teixeira', 0.0030234471410941995)]
```

```
In [110]: sorted_ccd = sorted(ccd.iteritems(), key=operator.itemgetter(1))
          sorted_ccd[::-1][:15]
```

```
Out[110]: [(u'Portal', 0.18563072508553138),
           (u'Renato Fabbri', 0.0896842747470639),
           (u'andre luiz da silva', 0.08298272075361858),
           (u'Henrique Parra Parra Filho', 0.07487334184166763),
           (u'Marcelo Rodrigues Saldanha da Silva', 0.07365123564218587),
           (u'Jose mendon\xe7a Furtado Neto', 0.07349614053766303),
           (u'Cleber Possani Jr', 0.07341788286278758),
           (u'Hylton Sarcinelli Luz', 0.07064662093125092),
           (u'Daniel Pitangueira de Avelino', 0.06848317922486721),
           (u'Juliano Geraldi', 0.06619519942465014),
           (u'Roberto Kodama', 0.06435394670688788),
           (u'Josenildo Fernando da Silva', 0.06338269573563692),
           (u'Maira Gon\xe7alves Lopes', 0.062471719457013575),
           (u'Ricardo Poppi', 0.06187519730611386),

           (u'Ronald Emerson Scherolt da Costa', 0.05675816292745991)]
```

C.2. Setores conectivos

Na caracterização qualitativa das redes livres de escala, há uma distinção que salta aos olhos: uma multidão de vértices pouco conectados (periféricos), pouquíssimos vértices muito conectados (hubs), e alguns intermediários (intermediários mesmo).

Esta divisão é usualmente concebida na proporção 5% hubs, 15% intermediários e 80% periféricos.

- $\approx 5 \pm 4\%$ dos vértices mais conectados (entidades operantes) possuem $\approx 70 \pm 10\%$ das arestas (recursos). São os hubs.
- $\approx 70 \pm 10\%$ dos vértices menos conectados possuem $\approx 20 \pm 12\%$ das arestas. São os periféricos.
- $\approx 25 \pm 14\%$ dos vértices restantes possuem $\approx 16 \pm 10\%$ das arestas. São os intermediários.

```
In [144]: hi=int(len(sorted_g)*0.05)
          ii=int(len(sorted_g)*0.15)
          pi=int(len(sorted_g)*0.80)
          sorted_g=sorted_g[::-1]
          hubs=sorted_g[:hi]
          intermediarios=sorted_g[hi:ii]
          perifericos=sorted_g[ii:]
```

```
In [117]: hubs
```

```
Out[117]: [(u'Marcelo Branco', 122),
            (u'Maria Jos\x9lia Amaral de Menezes', 60),
            (u'Ana C\x9lia Costa', 36),
            (u'Vicente Aguiar', 35),
            (u'Laura Zacher', 33),
            (u'Luis Felipe Coimbra Costa', 32),
            (u'Ricardo Poppi', 31),
            (u'LUCAS MOREIRA DE SOUSA', 30),
            (u'Ronald Emerson Scherolt da Costa', 29),
            (u'Valessio Brito', 27),
            (u'Grazielle Machado', 20),
            (u'Andr\x9 Filipe de Assun\x7\xe3o e Brito', 20),
            (u'Renato Fabbri', 20),
            (u'Daniela Feitosa', 18),
            (u'Daniel Pitangueira de Avelino', 18),
            (u'Nilzete Pacheco ', 18),
            (u'Carolina de Abreu Batista Claro', 17),
            (u'Jesulino Alves de Souza', 17),
            (u'Paulo Meirelles', 17),
            (u'Paulo Henrique de Lima Santana', 15),
            (u'kelly cristina alves', 14),
            (u'Daniel Chiaretti', 14)]
```

Descomentar para acessar participantes nos setores:

```
In [ ]: # intermediarios
        # perifericos
```

E para a rede de interação:

```
In [123]: hi=int(len(sorted_d)*0.05)
          ii=int(len(sorted_d)*0.15)
          pi=int(len(sorted_d)*0.80)
          sorted_d=sorted_d[::-1]
          hubs_d=sorted_d[:hi]
          intermediarios_d=sorted_d[hi:ii]
          perifericos_d=sorted_d[ii:]
```

```
In [125]: hubs_d
          # intermediarios_d
          # perifericos_d
```

```
Out[125]: [(u'Portal', 46),
           (u'Renato Fabbri', 26),
           (u'Gabriela Valle', 19),
           (u'Hylton Sarcinelli Luz', 15),
           (u'F\xeldia Rebou\xe7as', 13),
           (u'Grazielle Machado', 12),
           (u'Jose mendon\xe7a Furtado Neto', 11),
           (u'andre luiz da silva', 11),
           (u'Daniel Pitangueira de Avelino', 10),
           (u'Henrique Parra Parra Filho', 9),
           (u'Juliano Geraldi', 8)]
```

Há correlação com a atividade de cada participante, principalmente na rede de interação.

Além disso, é recomendada a comparação com a rede de Erdős-Renyi com o mesmo número de vértices e arestas, como feito neste trabalho: <http://sourceforge.net/p/labmacambira/fimDoMundo/ci/master/tree/textos/evolutionSN/paper.pdf?format=raw>.

C.3. Detecção de comunidades

```
In [145]: [aa for aa in x.community.k_clique_communities(g,5)]
```

```
Out[145]: [frozenset([u'Daniel Pitangueira de Avelino',
                      u'Francisco Marcelo de Ara\xfajo Lima J\xfanior',
                      u'Grazielle Machado',
                      u'Arthur de Moura Del Esposte',
                      u'Carolina Fonseca',
                      u'Fabiano Rangel Cidade',
                      u'Marina Arilha Silva',
                      u'Valessio Brito',
                      u'Paulo Meirelles',
                      u'Ana C\xe9lia Costa',
                      u'Ronald Emerson Scherolt da Costa',
                      u'Ricardo Poppi',
                      u'Bertoni',
                      u'Daniela Feitosa',
                      u'Vicente de Paula Blum de Oliveira',
                      u'Gustavo Buzzatti Pacheco',
                      u'Paulo Henrique de Lima Santana',
                      u'Juliana Hil\xelrio',
                      u'Mariel Zasso',
```

```

u'Everton dos Santos Silva Rodrigues',
u'Vicente Aguiar',
u'Luis Felipe Coimbra Costa',
u'Aur\x9lio A. Heckert',
u'Nitai Bezerra da Silva',
u'Renato Fabbri',
u'Maria Jos\x9lia Amaral de Menezes',
u'Luciano Santa Brigida',
u'Christian Miranda',
u'Jesulino Alves de Souza',
u'Marcelo Branco',
u'Andr\x9 Filipe de Assun\x7\x3o e Brito',
u'Sady Jacques']],
frozenet([u'Daniela Corr\xea Jacques Brauner',
u'Daniel Chiaretti',
u'Nara de Souza Rivitti',
u'Laura Zacher',
u'JOAO FREITAS DE CASTRO CHAVES',
u'Fernando de Souza Carvalho',
u'Alexandra Rodrigues Lazzarini',
u'Bruno Vinicius Batista Arruda',
u'Fernanda Hahn',
u'aidio junior mariano gon\x7alves',
u'Sabrina Piccoli Marques',
u'ANA LUISA ZAGO DE MORAES',
u'Isabel Penido de Campos Machado',
u'Maria Jos\x9lia Amaral de Menezes',
u'MARIA L\xdaCIA NIDBALLA DOS SANTOS',
u'Marcelo Branco',
u'Eraldo Silva J\xfanior']],
frozenet([u'Andrea Pacheco Pac\xedfico',
u'Carolina de Abreu Batista Claro',
u'Helisane Mahlke',
u'La\xeds Azeredo Alves',
u'erika pires ramos',
u'Shirley Cardoso Gon\x7alves de Aquino Alves']])

```

As comunidades acima foram definidas pelo número de amizades mínimo em comum no grupo. Há bons métodos espectrais (Newman) e iterativos.

|||--- FIM ---|||

Anexo D: Classificação de conteúdo via conectividade dos participantes

Importadas as bibliotecas:

```
In [10]: from SPARQLWrapper import SPARQLWrapper, JSON
import time, operator, numpy as n, networkx as x
```

Guardados os prefixos mais úteis no domínio:

```
In [31]: PREFIX=""PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX ops: <http://purl.org/socialparticipation/ops#>
PREFIX opa: <http://purl.org/socialparticipation/opa#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX tsioc: <http://rdfs.org/sioc/types#>
PREFIX sioc: <http://rdfs.org/sioc/ns#>
PREFIX schema: <http://schema.org/>""
```

Baixadas todas as amizades do participa

```
In [51]: NOW=time.time()
q="""SELECT DISTINCT ?a ?b ?aname ?bname
      WHERE {
        ?a foaf:knows ?b .
        ?a foaf:name ?aname .
        ?b foaf:name ?bname .
      }"""
sparql3 = SPARQLWrapper("http://localhost:82/participabr/query")
sparql3.setQuery(PREFIX+q)
sparql3.setReturnFormat(JSON)
results4 = sparql3.query().convert()
print("%.2f segundos para puxar todas as amizades do Participa.br"%
      (time.time()-NOW,))
```

0.12 segundos para puxar todas as amizades do Participa.br

Feita rede de amizades do Participa.br:

```
In [52]: g=x.Graph()
for amizade in results4["results"]["bindings"]:
    nome1=amizade["a"]["value"]
    nome2=amizade["b"]["value"]
    g.add_edge(nome1,nome2)
```

Baixadas interações no Participa.br:

```
In [17]: NOW=time.time()
q2="""SELECT ?participante1 ?participante2 ?aname ?bname
      WHERE {
        ?comentario dc:type tsioc:Comment.
        ?participante1 ops:performsParticipation ?comentario.
        ?participante1 foaf:name ?aname.
        ?artigo sioc:has_reply ?comentario.
        ?participante2 ops:performsParticipation ?artigo.
        ?participante2 foaf:name ?bname.
      }"""
sparql3.setQuery(PREFIX+q2)
sparql3.setReturnFormat(JSON)
results = sparql3.query().convert()
print("%.2f segundos para puxar as interações do Participa.br"%
      (time.time()-NOW,))
```

16.06 segundos para puxar as interações do Participa.br

Feita rede de interações:

```
In [18]: d=x.DiGraph()
for interacao in results["results"]["bindings"]:
    nome_chegada=interacao["participante1"]["value"]
    nome_partida=interacao["participante2"]["value"]
    if (nome_partida,nome_chegada) in d.edges():
        d[nome_partida][nome_chegada]["weight"]+=1
    else:
        d.add_edge(nome_partida,nome_chegada,weight=1.)
```

Listando os mais conectados para ver suas URIs e verificar bom andamento das buscas SparQL:

```
In [53]: sorted_g = sorted(g.degree().iteritems(), key=operator.itemgetter(1))
```



```
sorted_d[::-1][:15]
```

```
Out[53]: [(u'http://participa.br/profile/marcelobranco', 122),
          (u'http://participa.br/profile/mjade', 60),
          (u'http://participa.br/profile/anita', 36),
          (u'http://participa.br/profile/vicentedeaguiar', 35),
          (u'http://participa.br/profile/laurazacher', 33),
          (u'http://participa.br/profile/lfelipe', 32),
          (u'http://participa.br/profile/ricardopoppi', 31),
          (u'http://participa.br/profile/lucasmoreira', 30),
          (u'http://participa.br/profile/ronald costa', 29),
          (u'http://participa.br/profile/valessiobrito', 27),
          (u'http://participa.br/profile/grazi_machado', 20),
          (u'http://participa.br/profile/decko', 20),
          (u'http://participa.br/profile/rfabbri', 20),
          (u'http://participa.br/profile/nilzetepacheco', 18),
          (u'http://participa.br/profile/cclaro', 18)]
```

```
In [54]: sorted_d = sorted(d.degree().iteritems(), key=operator.itemgetter(1))
sorted_d[::-1][:15]
```

```
Out[54]: [(u'http://participa.br/profile/portal', 46),
          (u'http://participa.br/profile/rfabbri', 26),
          (u'http://participa.br/profile/gabriela', 19),
          (u'http://participa.br/profile/hyltonsarcinelliluz', 15),
          (u'http://participa.br/profile/fadia', 13),
          (u'http://participa.br/profile/grazi_machado', 12),
          (u'http://participa.br/profile/josefurtado', 11),
          (u'http://participa.br/profile/andre61', 11),
          (u'http://participa.br/profile/davelino', 10),
          (u'http://participa.br/profile/parrahenri', 9),
          (u'http://participa.br/profile/julianogeraldi', 8),
          (u'http://participa.br/profile/capuano', 7),
          (u'http://participa.br/profile/kodama', 7),
          (u'http://participa.br/profile/ibebrazil', 7),
          (u'http://participa.br/profile/thiagozoroastro', 7)]
```

Pode-se selecionar mensagens de periféricos, hubs e intermediários. Por simplicidade, aqui estão 5 mensagens dos 5 mais conectados nas **atividades**:

```
In [55]: uris=[i[0] for i in sorted_d[::-1][:5]]
for uri in uris:
    q=""" SELECT ?texto
        WHERE {
            ?comentario dc:creator <%s> .
            ?comentario dc:type tsioc:Comment.
            ?comentario schema:text ?texto .
        } LIMIT 2"""%(uri,)
    #print q
    sparql3.setQuery(PREFIX+q)
    results4 = sparql3.query().convert()
    #print results4
    print "\n"
    try:
        print uri+"\n"+ results4["results"][0]["bindings"][0]["texto"]["value"]
    except:
        pass
```

<http://participa.br/profile/rfabbri>
que órgãos, de que programas?

<http://participa.br/profile/gabriela>
Não pode esquecer da gestão democrática nas universidades e instituições de ensino.
Vale a pena mencionnar isso espqecificamente, porque já tem lei prevendo, mas na prática não acontece.

<http://participa.br/profile/hyltonsarcinelliluz>
Apoio a sugestão de incluir o termo "Controle" ao nome da politica em pauta, com vista a uniformizar a conceituação, visto que já existe e é difundido o termo "controle social" para políticas públicas, no qual se insere a participação em debate.

<http://participa.br/profile/fadia>
O termo participação já está em ampla utilização no âmbito dos governos, apropriado inclusive pelo capital. O que se questiona é o que se entende por essa participação e o que se quer é a efetivação da participação. Concordo com a inserção do termo controle.

Política Nacional para Efetivação do Controle e Participação Social

Pode-se selecionar mensagens de periféricos, hubs e intermediários. Por simplicidade, aqui estão 5 mensagens dos 5 mais conectados nas **amizades**:

```
In [56]: uris=[i[0] for i in sorted_g[::-1][:5]]
for uri in uris:
    q=""" SELECT ?texto
```

```

WHERE {
  ?comentario dc:creator <%s> .
  ?comentario dc:type tsioc:Comment.
  ?comentario schema:text ?texto .
} LIMIT 2""%(uri,)
#print q
sparql3.setQuery(PREFIX+q)
results4 = sparql3.query().convert()
#print results4
print "\n"
try:
  print uri+"\n"+ results4["results"]["bindings"][0]["texto"]["value"]
except:
  pass

```

<http://participa.br/profile/marcelobranco>
 Uma foto ou figura em cada post ficaria muito melhor.

<http://participa.br/profile/anita>
 Oiiii Ronald!!!

<http://participa.br/profile/vicentedeaguiar>
 Por exemplo, essa sua resposta que você fez no meu comentário, eu não consigo visualizar no mural, só diretamente no artigo. Vo consegue ver no mural essa minha resposta agora? Se sim, pode ser problema na versão do CSS para Firefox 25.0.1.

<http://participa.br/profile/laurazacher>
 Sejam muito bem-vindos!

Iniciamos a Conferência Virtual da DPU sobre Migrações e Refúgio já com uma importante discussão!

Aproveito o início da discussão reiterando a manifestação da Dra. Ana Luisa e repassando algumas orientações para organizarmos processo de discussão coletiva:

- Inicialmente, a discussão será em sentido amplo; contudo, conforme a discussão avançar, é preciso que os tópicos postados indiquem (1) a qual dos eixos temáticos estão vinculados, (2) a qual bloco está contemplando; (3) qual temática aborda.

Exemplo: Discussão proposta sobre expulsão: incluir no título da mensagem:

Eixo Temático nº (I ou IV) - Bloco: Mudanças Legislativas - Tema: Expulsão

Relembro os eixos temáticos a serem discutidos na Conferência da DPU:

Eixo I - Igualdade de tratamento e acesso a serviços e direitos;

Eixo IV - Abordagem de violações de direitos e meios de prevenção e proteção, no qual também será discutido o Subtema - Enfrentamento ao Sequestro/ Subtração Internacional de Crianças e a aplicação da Convenção de Haia de 1980 sobre os Aspectos Cíveis do Sequestro Internacional de Crianças.

Em cada uma das mensagens postadas, favor indicar a qual dos blocos propostos está vinculada:

- (a) mudanças legislativas (lei em sentido amplo);
- (b) políticas públicas;
- (c) atuação e estruturação da DPU para assistência ao migrante.

- Estas orientações visam facilitar o processo de sistematização das propostas. Para ilustrar o produto final da sistematização deixo o link para o Caderno de Propostas da 1ª Conferência Nacional de Segurança Pública, cujo modelo de conferência é a base da atual Comigrar:

http://www.ipea.gov.br/participacao/images/pdfs/conferencias/Seguranca_Publica/caderno_propostas_1_conferencia_seguranca_publica.pdf

- Informações sobre os eixos temáticos e sobre o processo da Conferência podem ser obtidas através do Manual do Participante: http://www.dpu.gov.br/images/stories/arquivos/PDF/Manual_do_Participante_-_DPU_-_COMIGRAR.pdf

- Qualquer dúvida sobre o processo da Conferência, favor entrar em contato através do telefone (51) 3216-6961 e/ou do correio eletrônico comigrar@dpu.gov.br.

Vamos construir uma política migratória com Justiça e Igualdade para todos!

|||--- FIM ---|||

Anexo E: Teste de conexão com o endpoint SparQL que distribui dados do Participa.br

Este script faz buscas simples no endpoint SparQL com os dados do Participa.br, para testar a conexão e o tempo de resposta. Para os fins até agora utilizados, foi suficiente:

- I. Fazer uma busca simples na DBpedia, para testar conexão de onde está sendo executado o script e para testar o funcionamento dos recursos envolvidos.
- II. Uma busca de dados do Participa.br, leve (todos os participantes, com nomes e URIs)
- III. Uma busca pesada de dados do Participa.br (todos os comentários, com texto, título e URIs)

Basta clicar nas células e apertar SHIFT+ENTER para executar os comandos dela.

```
In [3]: from SPARQLWrapper import SPARQLWrapper, JSON
import time
```

```
In [35]: NOW=time.time()
sparql = SPARQLWrapper("http://dbpedia.org/sparql")
sparql.setQuery("""
    PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
    SELECT ?label
    WHERE { <http://dbpedia.org/resource/Love> rdfs:label ?label }
""")
sparql.setReturnFormat(JSON)
results = sparql.query().convert()
print(    "%.2f segundos para consultar a dbpedia os nomes:" +
        " %s"*len(results["results"]["bindings"]) )
% tuple([time.time()-NOW]+[i["label"]["value"]
                        for i in results["results"]["bindings"]]) )
```

0.49 segundos para consultar a dbpedia os nomes: 愛 Liebe Love Amor Amor Amour Amore
愛 Liefde Miłość Любoвь Kärlek

```
In [36]: PREFIX="""PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX ops: <http://purl.org/socialparticipation/ops#>
PREFIX opa: <http://purl.org/socialparticipation/opa#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX tsioc: <http://rdfs.org/sioc/types#>
PREFIX schema: <http://schema.org/>
"""
```

```
In [40]: q2="SELECT ?nome WHERE {?s rdf:type ops:Participant . ?s foaf:name ?nome .}"
NOW=time.time()
sparql3 = SPARQLWrapper("http://localhost:82/participabr/query")
# caso esteja rodando em instancia propria:
#sparql3 = SPARQLWrapper("http://200.144.255.210:8082/participabr/query")
sparql3.setQuery(PREFIX+q2)
sparql3.setReturnFormat(JSON)
results3 = sparql3.query().convert()
print("%.2f segundos para puxar todos os %i nomes dos participantes do Participa.br"
      %(time.time()-NOW,len(results3["results"]["bindings"])))
print( "com os seguintes campos: %s"
      %(str(results3["results"]["bindings"][0].keys())) )
```

0.07 segundos para puxar todos os 2624 nomes dos participantes do Participa.br

0.07 segundos para puxar todos os 3624 nomes dos participantes do Participa.br
com os seguintes campos: [u'nome']

```
In [43]: NOW=time.time()
q="""SELECT ?comentario ?titulo ?texto WHERE {?comentario dc:type tsioc:Comment.
OPTIONAL {?comentario dc:title ?titulo . }
OPTIONAL {?comentario schema:text ?texto .}}"""
sparql3.setQuery(PREFIX+q)
sparql3.setReturnFormat(JSON)
results4 = sparql3.query().convert()
print("%.2f segundos para puxar todos os %i comentários do Participa.br"%
      (time.time()-NOW,len(results4["results"]["bindings"])))
print("com campos: "+str(results4["results"]["bindings"][0].keys()))
```

2.44 segundos para puxar todos os 16253 comentários do Participa.br
com campos: [u'titulo', u'comentario', u'texto']

|||--- FIM ---|||



Empoderando vidas.
Fortalecendo nações.

F Instâncias online

Alguns recursos desenvolvidos neste trabalho foram disponibilizados online, via http, pela natureza do trabalho e por conveniências. Consistem em ferramentas de streaming, ferramentas de análise, experimentos artísticos e repositórios. Uma seleção pertinente a este escrito é:

- IPython Notebook, com os códigos dos Anexos A-E: <http://200.144.255.210:8003>
- Endpoint Fuseki/Jena, para acesso aos dados do Participa.br via critérios semânticos: <http://200.144.255.210:8082>.
- Telões para streaming de estruturas sociais (cria 3 redes de relacionamento (d3js) entre os agentes e algumas seleções de palavras e contextos): <http://ocupagov.meteor.com>
- BrServers, com o resumo “redes+filtros multidimensionais”: <http://gentle-mesa-5082.herokuapp.com/twitter/arenaNETmundial/>
- MyNSA, pra coleta de dados do Facebook e Twitter: <http://mynsa.meteor.com>
- MMISSA, disposições livres, iniciais e artísticas sobre dados das redes sociais: <http://mmissa.meteor.com>
- MM, sonificação de estruturas sociais, com propósitos artísticos e para acessibilidade: <http://mm.meteor.com>
- Repositório de códigos e com este documento: <https://github.com/ttm/pnud3>