

UNIVERSIDADE DE SÃO PAULO  
INSTITUTO DE FÍSICA DE SÃO CARLOS

RENATO FABBRI

Estabilidade topológica e diferenciação textual  
em redes de interação humana:  
redes complexas para o participante  
e a física antropológica

São Carlos

2015



RENATO FABBRI

Estabilidade topológica e diferenciação textual  
em redes de interação humana: redes  
complexas para o participante e a física  
antropológica

Monografia apresentada ao Programa de Pós-Graduação em Física do Instituto de Física de São Carlos da Universidade de São Paulo, para o Exame de Qualificação como parte dos requisitos para obtenção do título de Doutor em Ciências.

Área de concentração: Física Aplicada  
Opção: Física Computacional  
Orientador: Prof. Dr. Osvaldo Novais de Oliveira Jr.

São Carlos

2015



# RESUMO

FABBRI, C. *Estabilidade topológica e diferenciação textual em redes de interação humana: redes complexas para o participante e a física antropológica*. Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2015.

As redes complexas compõem uma das áreas mais ativas da física recente. Há esforços consideráveis para apresentar estes avanços ao público não especialista, mas tudo indica que poucos ou nenhum são propostos para instrumentalizar o indivíduo que constitui estes sistemas a se beneficiarem. Ou seja, com um núcleo de conhecimento da área, e receitas para aproveitamento, fornece meios para o participante interagir e entender as redes nas quais ele se encontra. Este trabalho busca realizar tal tarefa por meio das redes sociais do participante. Verificamos que tais redes exibem uma estabilidade temporal de medidas topológicas e dos tamanhos relativos dos setores básicos (hubs, intermediários, periféricos). Observamos uma acentuada diferenciação da produção de texto de cada setor básico. Também formalizamos as conceitualizações vinculadas a estas redes como ontologias OWL onde foi possível, principalmente as instâncias de participação social previstas por lei e praticadas ou implementadas computacionalmente. Software e dados foram disponibilizados e usados. Protocolos escolhidos para facilitar a integração de estruturas de diferentes procedências, para reutilização dos dados em outros trabalhos e pesquisas, e para o benefício público. Consequências conceituais requerem considerações antropológicas e estão sendo redigidas. Próximos passos são: considerações tipológicas das propriedades físicas observadas nas redes de interação humana, com atenção aos outliers, às relações entre topologia do agente e texto produzido, e à ponte com a bagagem mais tradicional das ciências humanas no assunto; melhor documentação e desenvolvimento do aparato em software, ontologias e dados.

**Palavras-chave:** Redes complexas. Redes sociais. Complexidade. Física antropológica. Dados ligados. Web semântica. Participação social. Mineração de texto. Processamento de linguagem natural.



# Sumário

---

<b>1</b>	<b>Introdução</b>	<b>7</b>
1.1	Revisão de literatura . . . . .	8
1.1.1	Processamento de linguagem natural, dados ligados, participação social	9
1.1.2	Ambiguidades e sinônimos no jargão . . . . .	9
<b>2</b>	<b>Materiais</b>	<b>11</b>
2.1	O banco Gmane de dados públicos sobre listas de email (benchmark) . . . . .	11
2.2	Facebook, Twitter, Participa.br, Cidade Democrática, AA . . . . .	11
<b>3</b>	<b>Métodos</b>	<b>13</b>
3.1	Estatística temporal e circular . . . . .	14
3.2	Formação das redes de interação . . . . .	14
3.3	Setorialização de Erdös . . . . .	15
3.4	Médias e variâncias nas Análises de Componentes Principais de cada sistema .	18
3.4.1	Medidas consideradas e acrescentadas . . . . .	19
3.5	Teste de Kolmogorov-Smirnoff para os textos produzidos por cada setor . . . .	20
3.6	Audiovisualização de dados . . . . .	21
3.7	Considerações tipológicas e humanísticas . . . . .	21
3.8	Web semântica . . . . .	21
3.8.1	A construção de ontologias OWL e vocabulários SKOS . . . . .	22
3.8.2	A triplificação de dados relacionais . . . . .	22
<b>4</b>	<b>Resultados</b>	<b>23</b>
4.1	Estabilidade temporal e topológica; diferenciação textual em redes de interação humana . . . . .	23
4.2	Criação da nuvem brasileira de dados participativos ligados . . . . .	24
4.2.1	Síntese de ontologias e vocabulários de estruturas sociais . . . . .	25
4.2.2	Obtenção de dados ligados a partir de dados relacionais participativos .	26
4.2.3	Método de construção de ontologias orientado aos dados . . . . .	26
4.3	Aparato em software . . . . .	27
4.4	Benefício, utilidade e formalismo . . . . .	27
4.4.1	Sistemas de recomendação para o enriquecimento da navegação semântica de recursos . . . . .	27

4.4.2	Experimentos de percolação social e a física antropológica . . . . .	28
4.4.3	Entendimento sobre as estruturas sociais . . . . .	30
<b>5</b>	<b>Cronograma e afazeres</b>	<b>31</b>
<b>6</b>	<b>Conclusões e previsão</b>	<b>33</b>
	<b>Referências</b>	<b>34</b>



# 1 - Introdução

---

Os primeiros estudos sobre redes de interação humana datam do século XIX. Já a fundação da “Análise de Redes Sociais” /ARS (ou *Social Network Analysis*/SNA) é geralmente atribuída ao psiquiatra Jacob Moreno na metade do século vinte (1). Com a crescente disponibilidade de dados relacionados à interação humana, a pesquisa destas redes tem aumentado continuamente. Contribuições podem ser encontradas em uma variedade de áreas, de ciências sociais e humanidades (2) a ciências sociais (3) e física (4, 5), dada a natureza multidisciplinar do assunto. Uma das abordagens da perspectiva de uma ciência exata é representar a rede de interação como uma rede complexa (4, 5), com a qual algumas características foram reveladas. Por exemplo, a topologia das redes de interação humana exhibe um traço livre de escala, o que aponta para a existência de um pequeno número de hubs super conectados e um grande número de vértices pouco conectados.

Há um hiato de conhecimento e tecnologia entre o legado de redes complexas e o usufruto do participante. Este hiato é reativo, e há evidência de que conseguirá se manter como um ecossistema de conhecimento, tecnologia e empreendimento da sociedade em todas as suas escalas, acompanhando uma transição de fase histórica (6). Deve facilitar, por exemplo: elaboração e preparação de documentos, aquisição rápida de conhecimento, realização de empreitadas coletivas. Em geral: processos de coleta e difusão de informação (e bens) (7).

Este trabalho apresenta uma confirmação deste cenário e avanços. Algumas estratégias foram selecionadas para verificar a aplicabilidade de conceitos de redes complexas para o benefício do participante. Em especial, experimentos muito simples parecem capazes de modificar estruturas sociais. Neste contexto, verificamos estabilidades temporais nas redes de interação humana, e expomos que os setores básicos das redes (hubs, intermediários e periféricos) produzem textos bastante diferentes entre si. Este conhecimento é útil para uma tipologia não estigmatizante de participantes em redes de interação. A audiovisualização e interconexão de dados com arte e engenhocas em software deram suporte contínuo à pesquisa científica, e apresentam inovações. Aplicações foram complementadas em parceria com a Presidência da República e o Programa das Nações Unidas para o Desenvolvimento.

A próxima seção apresenta considerações gerais sobre a literatura de redes complexas. A Seção 1.1.1 faz observações pontuais sobre cada área secundária. A Seção 1.1.2 expõe a proliferação de ambiguidades e sinônimos no jargão deste trabalho. A Seção 2 é dedicada aos dados analisados. A Seção 3 contém os métodos usados para atingir os resultados, que

são explicitados na Seção 4. O cronograma de atividades e uma comparação entre afazeres planejados e finalizados estão na Seção 5. A monografia termina com as conclusões na Seção 6, seguida de agradecimentos e referências.

## 1.1 Revisão de literatura

A área das redes complexas é relativamente nova ( $\approx 25$  anos) e a literatura apresenta definições divergentes da área em si. Uma definição que tem recebido aceitação crescente é da rede complexa como “um grafo grande com características topológicas não triviais”. Esta definição é enganosa ao menos em três pontos. Primeiro, há redes de interesse com características topológicas triviais, como as redes de Erdős-Rényi e a Geográfica (1), ou as redes simples usadas para exemplos. Segundo, a definição falha ao não emitir a mensagem fundamental de que uma rede complexa não é somente uma estrutura matemática, um grafo isolado: as redes complexas de interesse são redes reais ou modelos idealizados para as entender. Além disso, não só grafos grandes são de interesse, mas grafos pequenos são comumente usados como extensão das estruturas maiores e como exemplos de propriedades. Uma definição, ainda longe de perfeita, mas preferida neste trabalho, é considerar a área das redes complexas como interessada em “redes usualmente grandes, consideradas no, ou para consideração do, meio em que residem”. Esta definição resolve ambos os pontos.

Os livros em geral apresentam um comum e poderoso repertório para a caracterização de sistemas complexos através de grafos. Talvez as mais notáveis características deste repertório sejam:

- O arsenal de medidas: grau, força, betweenness centrality, coeficiente de clusterização, etc. As medidas costumam se referir a um vértice, aresta, rede ou comunidade.
- Os paradigmas básicos de redes: Erdős-Rényi, geográfica, de mundo pequeno e livre de escala.
- A abordagem transdisciplinar para considerar o meio no qual a rede está inserida, ou que implica na rede.

A literatura sobre análise de redes sociais, por exemplo, pode ser frequentemente compreendida como redes complexas em sistemas sociais humanos. Uma consideração cuidadosa dos livros e artigos lidos para esta pesquisa está na Seção 5.

### 1.1.1 Processamento de linguagem natural, dados ligados, participação social

Diversos títulos foram lidos sobre processamento de linguagem natural, mineração de texto, visualização de dados e web semântica. Estas áreas têm impacto sobre o que está feito, e sendo feito, e foram cursadas formalmente uma disciplina sobre cada uma para o doutorado (veja Seção 5). Seguem informações pontuais sobre cada área.

Os termos processamento de linguagem natural (PLN) e mineração de texto (MT) podem em geral ser substituídos um pelo outro. O termo PLN é preferido nesta pesquisa pois o intuito é mais confluyente: compreender como a linguagem verbal está sendo usada para significar.

Os termos web semântica e dados ligados em geral também podem ser substituídos um pelo outro. O primeiro salienta a rede de referenciamento dos dados, o segundo os dados referenciando-se. Principalmente na esfera acadêmica, a área é, salvo segunda ordem, sinônimo de dados em RDF via XML ou Turtle, ontologias OWL e máquinas de inferência.

A visualização de dados de grafos em evolução temporal é bastante incipiente. Os poucos casos da literatura foram visitados. As animações abstratas de redes em evolução, e as “audio-visualizações” das redes, que disponibilizamos como parte desta pesquisa, são potencialmente contribuições na fronteira da visualização. Vídeo, porém, não é o formato mais apreciado pela literatura de visualização de dados, que tende a qualificar as figuras bidimensionais como as mais apropriadas para a pesquisa científica.

A participação social é a incorporação da própria sociedade nos processos de governança da sociedade. Quase toda a participação social atual é indireta e presencial, com a população fornecendo diretrizes, indicadores e acompanhamento para o setor público. A participação social tem sido fortalecida no mundo todo, e conceitos como transparência, participação direta (participação direta da sociedade civil na tomada de decisões pelo Estado) e democracia líquida (atribuição recursiva de competência para tomada de decisão), se estabelecendo a passos firmes como diretrizes para governos, acadêmicos e sociedade civil.

### 1.1.2 Ambiguidades e sinônimos no jargão

Além de recente, a área de redes complexas conflui diversas correntes científicas, como a física, a biologia e a sociologia. Portanto, possui termos ambíguos e sinônimos.

Exemplos de ambiguidade, sinônimos e delimitações adotadas:

- Os vértices mais conectados são, por definição, chamados hubs da rede. O vértice mais conectado é chamado hub da rede. No contexto do algoritmo HITS, o que é bem

comum, estes significados mudam: os hubs são os que possuem mais arestas saindo (grau de saída); as autoridades recebem as arestas, ou são referenciados por vários hubs e outras entidades.

- Há uma definição de centro e periferia com relação ao raio e diâmetro da rede (1, 8). Por extensão os intermediários podem ser considerados os que não são centro nem periferia. Esta setorialização centro, intermediários e periferia gera frações que diferem do previsto pela literatura para as frações de hubs, intermediários e periféricos. Um método apropriado para realizar esta setorialização da rede, com resultados estáveis e significativos, consta na Seção 3.3.
- *Aresta* e *ligação* são usadas como sinônimos. *Nó* e *vértice* também. É comum o uso de outros termos, em geral coerentes com a aplicação, como *agente*, *ator* ou *participante* para vértices de redes observadas em sistemas humanos.
- *Laço*, *loop*, *selfloop*, *autoloop*, *buckle* são termos usados para designar uma aresta de um vértice para ele próprio.

Neste trabalho, muitas outras questões sobre a nomenclatura merecem exposição para evitar entendimentos errados. Por falta de espaço, esta discussão ficará de fora desta monografia, mas dever á constar em (9).

## 2 - Materiais

---

### 2.1 O banco Gmane de dados públicos sobre listas de email (benchmark)

Mensagens de listas de email foram obtidas através do arquivo Gmane (10), que consiste em mais de 20 mil listas de email e mais de 130 milhões de mensagens públicas (11). Estas listas cobrem uma variedade de assuntos, em especial relacionados à tecnologia. O arquivo pode ser descrito como um corpus com metadados de emails, que incluem hora e lugar de envio, nome e email do remetente. O uso do Gmane para pesquisa científica é incidente no estudo de listas isoladas e de inovações lexicais (3, 12).

### 2.2 Facebook, Twitter, Participa.br, Cidade Democrática, AA

Embora as redes de email tenham sido usadas como referência na observação de propriedades gerais, outras fontes foram analisadas:

- Redes de amizade e interação do Facebook: 8 são usadas como referência em (13), mas dezenas, talvez algumas centenas, foram observadas nos experimentos da Seção 4.4.2.
- Milhares de tweets (talvez alguns milhões), geralmente vinculados a alguma *hashtag*. Em especial, a rede de *retweets* de 22 mil *tweets* com a *hashtag* #arenaNETmundial, foi analisada em (13).
- Mecanismos participativos como o Participa.br, Cidade Democrática e o AA. As redes de amizade e de interação do Participa.br foram analisadas em (13).



## 3 - Métodos

---

Para realização desta pesquisa, foram necessários métodos consagrados, adequações e variantes. Esta seção expõe uma seleção destes métodos, para organizar o conhecimento e exemplificar esta diversidade:

- A Seção 3.1 expõe medidas simples de estatística circular, ou direcional. A contribuição neste caso é unicamente nos padrões encontrados, o método é bastante estabelecido.
- A Seção 3.2 expõe a síntese de redes de interação. Talvez haja contribuição na síntese do conceito de redes de interação, pois não encontramos (ainda) na literatura tal exposição concisa. De qualquer forma, o conceito e o procedimento para obtenção das redes a partir de dados é usual, a exposição neste texto e no artigo (13) serve principalmente ao intuito de formalização do processo.
- A Seção 3.3 é dedicada à “Setorialização de Erdös”, para obtenção dos três setores básicos da rede, compostos por: hubs, intermediários e periféricos. O método parece não ter sido aplicado antes para este fim, e é resultado imediato da observação das caudas longas de dados reais contrastadas com o modelo de Erdös-Rényi (14).
- A Seção 3.4 apresenta o uso mais recorrente da Análise de Componentes Principais (PCA) neste trabalho. Várias redes são observadas, ou a mesma rede é observada em vários momentos, e a concentração de dispersão das componentes principais, e das medidas nas componentes principais, são observadas através de médias e desvios padrão.
- A Seção 3.4.1 apresenta as medidas utilizadas nas análises, com exposição formal das medidas de simetria potencialmente novas (não encontramos ainda na literatura), mas bastante relevantes para os resultados.
- A Seção 3.5 apresenta o uso que fazemos do teste de Kolmogorov-Smirnov de amostragem dupla. O método é bem estabelecido, e a contribuição está nos resultados alcançados com ele sobre diferenciação da produção de texto nas redes de interação.
- A Seção 3.6 expõe sobre a utilização dos dados de redes sociais para geração de imagem, música, e animação abstrata.
- A Seção 3.7 explicita a pertinente recorrência nesta pesquisa de considerações qualitativas e do cânone das ciências humanas.
- A Seção 3.8 delinea muito brevemente as abordagens utilizadas para registrar conceitualizações e vinculá-las aos dados.

### 3.1 Estatística temporal e circular

Para observação de padrões temporais, foram consideradas escalas diferentes. Em cada escala, de segundos e meses, foram construídos histogramas de atividade: cada unidade de tempo foi considerado um intervalo em que foram contabilizadas as atividades (e.g. mensagens de email). Também foram feitas algumas medidas de estatística circular (15), conforme exposto a seguir.

Considere cada *medida* (dato pontual) como um número complexo de módulo 1,  $z = e^{i\theta} = \cos(\theta) + i\sin(\theta)$ , onde  $\theta = medida \frac{2\pi}{periodo}$ . Os momentos  $m_n$ , tamanhos dos momentos  $R_n$ , ângulo médio  $\theta_\mu$ , e o ângulo médio reescalado  $\theta'_\mu$  são definidos assim:

$$\begin{aligned} m_n &= \frac{1}{N} \sum_{i=1}^N z_i^n \\ R_n &= |m_n| \\ \theta_\mu &= Arg(m_1) \\ \theta'_\mu &= \frac{period}{2\pi} \theta_\mu \end{aligned} \tag{3.1}$$

$\theta'_\mu$  é usado como medida de localização. A dispersão é medida usando a variância circular  $Var(z)$ , o desvio padrão circular  $S(z)$ , e a dispersão circular  $\delta(z)$ :

$$\begin{aligned} Var(z) &= 1 - R_1 \\ S(z) &= \sqrt{-2 \ln(R_1)} \\ \delta(z) &= \frac{1 - R_2}{2R_1^2} \end{aligned} \tag{3.2}$$

Como esperado, e pode ser notado nas informações de suporte de (13), há uma correlação positiva entre  $Var(z)$ ,  $S(z)$  e  $\delta(z)$ . A medida  $\delta(z)$  foi preferida na discussão dos resultados. A fração  $\frac{b_h}{b_l}$  entre a maior  $b_h$  e a menor  $b_l$  incidência nos histogramas também serviram como pista sobre quão próximas à distribuição uniforme são as distribuições observadas.

### 3.2 Formação das redes de interação

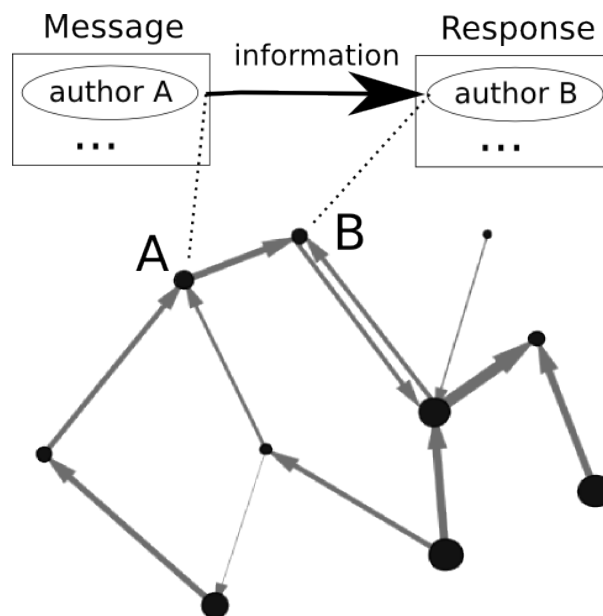
Redes de interação podem ser modeladas tanto com quanto sem peso, tanto dirigida quando não dirigida (1,3,16,17). Neste trabalho, quando possível, consideramos redes dirigidas e com peso, a mais informativa das possibilidades. Nestes casos, desconsideramos as versões



dirigidas sem peso, não dirigidas com peso e não dirigidas sem peso.

Em geral, as redes de interação são obtidas da seguinte forma: uma reação direta do participante B a uma mensagem do participante A implica em uma aresta de A para B, representando a informação que foi de A para B. O raciocínio é: se B reagiu a uma mensagem de A, ele/ela leu o que A escreveu e formulou uma reação, portanto B assimilou informação de A, assim  $A \rightarrow B$ . A inversão da direção da aresta produz a rede de status: B leu a mensagem e considerou o que A escreveu digno de resposta, dando status para A, portanto  $B \rightarrow A$ . Neste trabalho, as redes de interação são dirigidas conforme o fluxo de informação,  $A \rightarrow B$ . A Figura 3.1 expõe esta formação. Maiores detalhes são: arestas em ambas as direções são consideradas distintas; laços são considerados não informativos (para os interesses atuais) e descartados; a primeira interação  $A \rightarrow B$  cria a aresta com peso 1; a cada nova interação  $A \rightarrow B$  é adicionado 1 ao peso da aresta. Estas redes de interação humana constam na literatura como portadoras de propriedades livres de escala (e pequeno mundo), como esperado para (algumas) redes sociais (1, 3).

Figura 3.1 – A formação da rede de interação a partir de mensagens e respostas. Cada vértice representa um participante. Uma resposta do participante B a uma mensagem do participante A é considerada evidência de que B recebeu informação de A, representada então por uma aresta dirigida. Múltiplas mensagens adicionam “peso” à aresta dirigida. Maiores detalhes estão na Seção 3.2

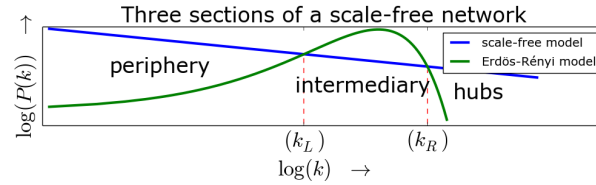


### 3.3 Setorialização de Erdős

Em uma rede livre de escala, os setores periféricos, intermediários e de hubs podem ser observados através de uma comparação com uma rede de Erdős-Rényi com o mesmo número

de arestas e vértices (14), como na Figura 3.2. Referiremos-nos a este procedimento como *setorialização de Erdős*, com os setores resultantes chamados *setores de Erdős* (ou *setores primitivos*, *setores básicos* da rede).

Figura 3.2 – As distribuições de grau de modelos ideais de redes livres de escala e Erdős-Rényi. A segunda possui mais vértices intermediários, enquanto a primeira possui mais vértices periféricos e hubs. As bordas dos setores são definidas pelas duas intersecções  $k_L$  e  $k_R$  das distribuições de conectividade. Os graus característicos estão nos intervalos compactos:  $[0, k_L]$ ,  $(k_L, k_R]$ ,  $(k_R, k_{max}]$  para os setores de Erdős (periferia, intermediários e hubs).



A distribuição de grau  $\tilde{P}(k)$  de uma rede livre de escala ideal  $\mathcal{N}_f$  com  $N$  vértices e  $z$  arestas possui menos vértices com grau médio do que a distribuição  $P(k)$  de uma rede Erdős-Rényi com o mesmo número de vértices e arestas. De fato, definimos (neste trabalho) o setor intermediário de uma rede como sendo o conjunto de todos os vértices cujo grau é menos abundante em uma rede real do que no modelo de Erdős-Rényi. Para assegurar a validade estatística dos histogramas, os intervalos podem ser escolhidos de forma que contenham ao menos  $\eta$  vértices da rede real. Assim, cada intervalo, começando no grau  $k_i$ , estende-se por  $\Delta_i = [k_i, k_j]$ , onde  $j$  é o menor inteiro tal que há ao menos  $\eta$  vértices com grau maior que ou igual a  $k_i$ , e menos que  $k_j$ . Assim, podemos escrever que:

$$\sum_{x=k_i}^{k_j} \tilde{P}(x) < \sum_{x=k_i}^{k_j} P(x) \Rightarrow i \text{ é intermediário} \quad (3.3)$$

Se  $\mathcal{N}_f$  for dirigida e não possuir laço (aresta de um vértice para ele próprio), a probabilidade de existência de uma aresta entre dois vértices arbitrários é  $p_e = \frac{z}{N(N-1)}$ . Um vértice em um dígrafo de Erdős-Rényi com o mesmo número de vértices e arestas, portanto mesma probabilidade  $p_e$  para existência de aresta, terá grau  $k$  com probabilidade  $P(k)$  ditada pela distribuição binomial:

$$P(k) = \binom{2(N-1)}{k} p_e^k (1-p_e)^{2(N-1)-k} \quad (3.4)$$

A cauda longa de graus baixos consiste nos vértices de borda, i.e. o setor periférico ou periferia, onde  $\tilde{P}(k) > P(k)$  e  $k$  é mais baixo que qualquer valor intermediário de  $k$ . A cauda longa de grau alto é o setor dos hubs, i.e.  $\tilde{P}(k) > P(k)$  e  $k$  é maior que qualquer valor de  $k$  do setor intermediário. O raciocínio para esta classificação é: os vértices tão conectados que são virtualmente inexistentes em redes conectadas por puro acaso (i.e. sem ligação preferencial)

são corretamente associadas aos hubs. Vértices com pouquíssimas conexões, e muito mais abundantes do que esperado por puro acaso, são atribuídos à periferia. Vértices com valores de grau previstos como os mais abundantes caso as conexões sejam fruto de puro acaso, valores próximos da média, e menos abundantes em nas redes reais, são classificados como intermediários. Se a força  $s$  for usada para comparação,  $P$  permanece a mesma, mas  $P(\kappa_i)$  com  $\kappa_i = \frac{s_i}{\bar{w}}$  deve ser usado na comparação, com  $\bar{w} = 2 \frac{\sum_i s_i}{N}$  o peso médio da aresta e  $s_i$  o peso do vértice  $i$ . Para graus de entrada e saída  $(k^{in}, k^{out})$  a comparação com a rede real deve ser feita com:

$$\hat{P}(k^{way}) = \binom{N-1}{k^{way}} p_e^k (1-p_e)^{N-1-k^{way}} \quad (3.5)$$

onde  $way$  (sentido) pode ser  $in$  ou  $out$  (entrada e saída). Forças de entrada e saída  $(s^{in}, s^{out})$  são divididas por  $\bar{w}$  e comparadas também usando  $\hat{P}$ . Note que  $p_e$  permanece a mesma, pois cada aresta é uma aresta de entrada (ou de saída), e há no máximo  $N(N-1)$  arestas entrando (ou saindo), portanto  $p_e = \frac{z}{N(N-1)}$  assim como no caso do grau total.

Em outras palavras, sejam  $\gamma$  e  $\phi$  inteiros nos intervalos  $1 \leq \gamma \leq 6$ ,  $1 \leq \phi \leq 3$ . Cada uma das seis possibilidades de setorialização de Erdős  $\{E_\gamma\}$  possui três setores de Erdős  $E_\gamma = \{e_{\gamma,\phi}\}$  definidos como:

$$\begin{aligned} e_{\gamma,1} &= \{ i \mid \bar{k}_{\gamma,L} \geq \bar{k}_{\gamma,i} \} \\ e_{\gamma,2} &= \{ i \mid \bar{k}_{\gamma,L} < \bar{k}_{\gamma,i} \leq \bar{k}_{\gamma,R} \} \\ e_{\gamma,3} &= \{ i \mid \bar{k}_{\gamma,i} < \bar{k}_{\gamma,R} \} \end{aligned} \quad (3.6)$$

onde  $\bar{k}_{\gamma,i}$  é a medida  $\gamma$  no vértice  $i$ , convencionada:

$$\begin{aligned} \bar{k}_{1,i} &= k_i \\ \bar{k}_{2,i} &= k_i^{in} \\ \bar{k}_{3,i} &= k_i^{out} \\ \bar{k}_{4,i} &= \frac{s_i}{\bar{w}} \\ \bar{k}_{5,i} &= \frac{s_i^{in}}{\bar{w}} \\ \bar{k}_{6,i} &= \frac{s_i^{out}}{\bar{w}} \end{aligned} \quad (3.7)$$

e ambos  $\bar{k}_{\gamma,L}$  e  $\bar{k}_{\gamma,R}$  são encontrados usando  $P(\bar{k})$  ou  $\hat{P}(\bar{k})$  como descrito.

Como métricas diferentes podem ser usadas para identificar os três tipos de vértices, critérios compostos podem ser definidos. Após uma inspeção cuidadosa das possibilidades, os critérios compostos foram reduzidos a 6. Utilizando as Equações 3.6, estes critérios compostos  $C_\delta$ , com  $\delta$  inteiro no intervalo  $1 \leq \delta < 6$  podem ser descritos como:

$$\begin{aligned}
 C_1 &= \{c_{1,\phi} = \{i \mid i \in e_{\gamma,\phi}, \forall \gamma\}\} \\
 C_2 &= \{c_{2,\phi} = \{i \mid \exists \gamma : i \in e_{\gamma,\phi}\}\} \\
 C_3 &= \{c_{3,\phi} = \{i \mid i \in e_{\gamma,\phi'}, \forall \gamma, \forall \phi' \geq \phi\}\} \\
 C_4 &= \{c_{4,\phi} = \{i \mid i \in e_{\gamma,\phi'}, \forall \gamma, \forall \phi' \leq \phi\}\} \\
 C_5 &= \{c_{5,\phi} = \{i \mid i \in e_{\gamma,\phi'}, \forall \gamma, \\
 &\quad \forall (\phi' + 1) \% 4 \leq (\phi + 1) \% 4\}\} \\
 C_6 &= \{c_{6,\phi} = \{i \mid i \in e_{\gamma,\phi'}, \forall \gamma, \\
 &\quad \forall (\phi' + 1) \% 4 \geq (\phi + 1) \% 4\}\}
 \end{aligned} \tag{3.8}$$

No artigo (13), os critérios  $C_1$ ,  $C_3$  e  $C_5$  foram chamados exclusivistas, os critérios  $C_2$ ,  $C_4$  e  $C_6$  foram chamados inclusivistas, os critérios  $C_3$  e  $C_4$  de cascata e os critérios  $C_5$  e  $C_6$  de externos. Note que uma cascata exclusivista  $C_3$  é a mesma classificação que uma cascata invertida (considera-se dos periféricos aos hubs) e inclusivista. Estes critérios compostos são especialmente úteis para observar estruturas com poucos participantes ou fruto de pouca atividade (veja as figuras do documento de Supporting Information de (13)).

### 3.4 Médias e variâncias nas Análises de Componentes Principais de cada sistema

A Análise de Componentes Principais (PCA é a sigla consagrada, do inglês Principal Component Analysis) é bastante estabelecida e bem documentada e foi usada para saber: 1) quais as medidas que contribuem para cada componente principal e em que proporção; 2) quanto da dispersão está concentrada em cada componente. 3) médias e desvios padrão destas quantidades para vários sistemas diferentes. Em geral, os diferentes sistemas eram evolução temporal um do outro.

Ou seja, foram analisados os autovetores e autovalores das matrizes de vértices e suas medidas da seguinte forma: seja  $\mathbf{X} = \{X[i, j]\}$  a matriz de todos os vértices  $i$  e respectivos

valores de cada medida  $j$ ,  $\mu_X[j] = \frac{\sum_j X[j]}{J}$  a média da métrica  $j$ ,  $\sigma_X[j] = \sqrt{\frac{(X[j] - \mu_X[j])^2}{J}}$  o desvio padrão da métrica  $j$ , e  $\mathbf{X}' = \frac{X[i,j] - \mu_X[j]}{\sigma_X[j]}$  a matriz com *z-score* de cada métrica  $j$  de  $\mathbf{X}$  em cada coluna. Seja  $\mathbf{V} = \{V[j, k]\}$  a matriz  $J \times J$  de autovetores da matriz  $\mathbf{C}$  de covariância de  $\mathbf{X}'$ , um autovetor por coluna. Cada autovetor combina as medidas originais em uma componente principal, portanto,  $V'[j, k] = 100 * \frac{|V[j, k]|}{\sum_{j'} |V[j', k]|}$  dá a percentagem da componente principal  $k$  que é contribuição da medida  $j$ . Com o vetor de  $k$  autovalores  $D[k]$ , basta observar  $D'[k] = 100 * \frac{D[k]}{\sum_{k'} D[k']}$  para saber a percentagem da dispersão pela qual a componente principal é responsável. Com os autovalores  $k$  ordenados de forma decrescente, em geral observam-se os primeiros três autovalores e respectivos autovetores em percentagens  $\{(V'[j, k], D'[k])\}$ , pois em geral já revelam padrões suficientes para uma boa análise e somam entre 60 e 95% da dispersão de todo o sistema. Em (13), em especial, foram feitas médias e desvios das contribuições de cada componente para a dispersão e das medidas em cada componente. Ou seja, dadas  $L$  observações  $l$ , cada uma com  $k$  pares de autovalores e autovetores, são observadas, para cada medida, a média  $\mu_{V'}[j, k]$  e desvio  $\sigma_{V'}[j, k]$  da medida  $j$  na componente principal  $k$ , e a média  $\mu_{D'}[k]$  e desvio  $\sigma_{D'}[k]$  da contribuição da componente  $k$  na dispersão do sistema:

$$\begin{aligned}
 \mu_{V'}[j, k] &= \frac{\sum_l^L V'[j, k, l]}{L} \\
 \sigma_{V'}[j, k] &= \sqrt{\frac{(\mu_{V'} - V'[j, k, l])^2}{L}} \\
 \mu_{D'}[k] &= \frac{\sum_l^L D'[k, l]}{L} \\
 \sigma_{D'}[k] &= \sqrt{\frac{(\mu_{D'} - D'[k, l])^2}{L}}
 \end{aligned} \tag{3.9}$$

A matriz de covariância  $\mathbf{C}$  também é observada diretamente para uma primeira pista sobre os padrões. Isso é feito com associações simples: valores absolutos pequenos indicam baixa correlação (a princípio independência); valores altos indicam correlação positiva (diretamente proporcional); valores negativos com módulo grande indicam correlação negativa (inversamente proporcional).

### 3.4.1 Medidas consideradas e acrescentadas

A topologia das redes foi estudada utilizando PCA (18) com uma pequena seleção das medidas mais básicas e fundamentais de cada vértice.

As seguintes medidas bastante conhecidas foram usadas: grau, grau de entrada, grau de saída, força, força de entrada, força de saída, coeficiente de clusterização, centralidade de intermediação (*betweenness centrality*) (1). Além disso, para apreender as simetrias das atividades dos participantes, as seguintes métricas foram introduzidas para cada vértice  $i$ :

- Assimetria:  $asy_i = \frac{k_i^{in} - k_i^{out}}{k_i}$ .
- Média da assimetria das arestas:  $\mu_i^{asy} = \frac{\sum_{j \in J_i} e_{ji} - e_{ij}}{|J_i| = k_i}$ , onde  $e_{xy}$  é 1 se houver aresta de  $x$  para  $y$ , e 0 caso contrário.  $J_i$  é o conjunto de vizinhos do vértice  $i$ , e  $|J_i| = k_i$  é o número de vizinhos do vértice  $i$ .
- Desvio padrão da assimetria das arestas:  $\sigma_i^{asy} = \sqrt{\frac{\sum_{j \in J_i} [\mu_i^{asy} - (e_{ji} - e_{ij})]^2}{k_i}}$ .
- Desequilíbrio:  $dis_i = \frac{s_i^{in} - s_i^{out}}{s_i}$ .
- Média do desequilíbrio das arestas:  $\mu_i^{dis} = \frac{\sum_{j \in J_i} \frac{w_{ji} - w_{ij}}{s_i}}{k_i}$ , onde  $w_{xy}$  é o peso da aresta  $x \rightarrow y$  e zero se não houver tal aresta.
- Desvio padrão do desequilíbrio das arestas:  $\sigma_i^{dis} = \sqrt{\frac{\sum_{j \in J_i} [\mu_i^{dis} - \frac{(w_{ji} - w_{ij})}{s_i}]^2}{k_i}}$ .

### 3.5 Teste de Kolmogorov-Smirnoff para os textos produzidos por cada setor

Sejam  $F_{1,n}$  e  $F_{2,n'}$  duas distribuições cumulativas empíricas onde  $n$  e  $n'$  contam as observações em cada amostragem. O teste de Kolmogorov-Smirnov de amostragem dupla rejeita a hipótese nula (rejeita que  $F_{1,n}$  seja fruto da mesma distribuição que  $F_{2,n'}$ ) se:

$$D_{n,n'} > c(\alpha) \sqrt{\frac{n+n'}{nn'}} \quad (3.10)$$

onde  $D_{n,n'} = \sup_x [F_{1,n} - F_{2,n'}]$  (a maior diferença entre as duas cumulativas) e  $c(\alpha)$  é tabelado para cada região crítica  $\alpha$  (probabilidade da hipótese nula ser verdadeira).

São calculados  $D_{n,n'}$ , enquanto  $n$  e  $n'$  são dados. Todos os termos da Equação 3.10 são positivos e  $c(\alpha)$  pode ser isolado:

$$c(\alpha) < \frac{D_{n,n'}}{\sqrt{\frac{n+n'}{nn'}}} = c'(\alpha) \quad (3.11)$$

Utilizamos  $c'(\alpha)$  como distância entre pares de distribuições empíricas, o que é coerente com a teoria (19).

## 3.6 Audiovisualização de dados

Redes foram visualizadas com imagens, vídeos e engenhocas online para esta pesquisa (20–22). Redes também foram sonificadas, em especial como faceta sonora de animações abstratas (10, 23–25). Tais “audiovisualizações” foram cruciais para guiar a pesquisa para características relevantes das redes de interação. Além disso, os tamanhos relativos dos três setores de Erdős foram visualizados como linhas temporais. A visualização da estrutura em rede foi especialmente útil na inspeção dos dados e estruturas das redes de email (20).

## 3.7 Considerações tipológicas e humanísticas

As redes estudadas neste trabalho são constituídas por seres humanos. Quando há classificação envolvida, seja dos agentes ou dos sistemas em si, reflexões humanísticas são pertinentes, como as disparadas pelas perguntas: qual o potencial estigmatizante da classificação? O que mais sabemos sobre o indivíduo ou a rede que é classificada? Quais dados posso usar e que procedimentos posso realizar sem desviar a atenção da pesquisa para leis e processos de comitês de ética? Qual a melhor forma de proceder com os dados e conhecimentos frutos da pesquisa?

Todas estas questões, e muitas outras, estão em constante amadurecimento com grupos de pesquisa (26), leituras (27), escrita (28, 29), e contatos individuais com outros pesquisadores.

## 3.8 Web semântica

As estruturas sociais são muitas vezes ditadas por estruturas pré-concebidas, fruto de tradições e esforços especialistas. Para a formalização de conceitualizações, e associados formatos de dados apropriados para armazenamento compartilhamento e referência, foram adotadas as recomendações de dados ligados / web semântica da W3C (30, 31). De forma bastante resumida, o arcabouço utilizado pode ser visto como uma maneira de formalizar conceitos (classes), relações entre conceitos (propriedades) e instâncias dos conceitos (indivíduos). As informações são expressas de forma semi-estruturada em RDF: triplas “sujeito predicado objeto”, com o sujeito sempre uma classe, o predicado sempre uma propriedade, e o objeto sempre uma classe ou dado. As propriedades podem ter especificidades, chamadas “axiomas de propriedade”. As classes podem ser restritas a possuírem certas relações, chamadas “restrições de classe”. É uma recomendação da W3C e o padrão acadêmico para dados ligados,

i.e. para representação na web semântica.

Utilidades da tecnologia incluem: inferência por máquina através de especificações ontológicas; interconexão de dados de fontes diferentes; organização ontológica de conhecimento específico para consideração cuidadosa, seja individual ou em grupo.

As ontologias são chave dentre as tecnologias de web semântica. Uma ontologia é geralmente definida como uma “especificação de uma conceitualização”, e a recomendação é o uso do padrão OWL (31). Os vocabulários são coleções de termos e metadados, como definição, e a recomendação é o uso do padrão SKOS (32). A web semântica tem apresentado avanços: as inferências, por exemplo, têm se tornado mais ágeis e úteis, especialmente para buscas. Ao mesmo tempo, é uma tecnologia complicada e com algumas dificuldades de implementação. Por exemplo, um conceito SKOS é um indivíduo (instância de uma classe), e uma classe OWL, se identificada com um conceito SKOS é, por consequência, um indivíduo. Neste caso, quando um indivíduo (instância de uma classe) é também uma classe, dada a complexidade, os recursos de inferência por máquina ficam limitados e lentificados.

### 3.8.1 A construção de ontologias OWL e vocabulários SKOS

Para formalizar conceitualizações referentes às estruturas sociais, mais especificamente relacionadas à participação social, foram construídas ontologias OWL e vocabulários SKOS a partir de entrevistas com especialistas acadêmicos e gestores públicos. Também foram feitas ontologias e vocabulários a partir de bancos de dados, decretos presidenciais e outras documentações. O processo consistiu sempre que possível na coleta de informações, formalização dos conceitos e devolutiva aos entrevistados, com figuras e outras documentações, até que não tivessem mais contribuições (33).

### 3.8.2 A triplificação de dados relacionais

Para disponibilização e uso de dados de diferentes fontes, foram feitos pequenos programas de computador (*scripts*) para acessar dados relacionais e escrever triplas RDF com os dados semanticamente enriquecidos. Estes *scripts* formalizam conceitos e os vinculam aos dados. Na sequência, acessam as ontologias pertinentes, salvam uma versão com os dados e ontologias, e uma versão com os dados, as ontologias e as triplas resultantes da inferência sobre os dados com a ontologia.



## 4 - Resultados

---

### 4.1 Estabilidade temporal e topológica; diferenciação textual em redes de interação humana

Explicitados cuidadosamente em (13), os principais resultados da estabilidade temporal e topológica em redes de interação humana são:

- A atividade ao longo do tempo é praticamente a mesma para todas as listas de email analisadas, e em todas as escalas. A maior dispersão foi encontrada nos segundos e minutos, seguida pela dispersão encontrada nos dias do mês, meses, dias da semana e horas do dia. Padrões estáveis foram apreciados em todas estas escalas: segundos, minutos e dias do mês apresentaram uniformidade; meses parecem seguir calendários acadêmicos e escolares; dias da semana apresentam redução para dois ou um terço das atividades nos finais de semana; nas horas do dia, há concentração de atividades das 12-18h, mas o pico ocorre pouco antes das 12h.
- A fração de participantes em cada setor de Erdős é estável ao longo do tempo e esta estrutura já desponta na rede mesmo com poucas mensagens.
- As métricas topológicas se combinam nas componentes principais do PCA praticamente da mesma forma para todas as listas e todos os *snapshots*.
- As medidas de simetria da topologia, como definidas na Seção 3.4.1, são responsáveis por mais dispersão do que o coeficiente de clusterização. Resultado menor: o coeficiente de clusterização se combina com os desvios padrões de assimetria e desequilíbrio para a formação da terceira componente.
- Estes comportamentos são muito estáveis para redes de interação de email. Nas outras redes analisadas, Twitter e Participa.br apresentaram redes bastante similares às de email. Nas redes do Facebook foram encontradas algumas redes que diferiam do modelo apresentado pelas redes de listas públicas de email em dois aspectos: algumas proporções e combinações de medidas das componentes principais; frações de participantes em cada setor de Erdős.
- Para um mesmo número de mensagens (sejam 20 mil) e diferentes listas, há uma correlação negativa entre número de participantes e número de *threads* quando os participantes são poucos (até  $\approx 2$  mil participantes quando são 20 mil mensagens). Para uma quantidade maior de participantes, há uma correlação positiva entre o número de

participantes e o número de *threads*. Este fato deve estar relacionado a outras características topológicas e textuais da rede e pode servir para uma tipologia das próprias redes.

- A setorialização de Erdős implica em uma tipologia de agentes em redes humanas de interação. Esta tipologia é, a princípio, não estigmatizante pois os agentes mudam de setor constantemente. Além disso, um mesmo agente pertence a todos os setores ao mesmo tempo, mas em redes diferentes. Maiores qualificações desta tipologia, decorrente do pertencimento a um setor de Erdős, estão no final dos resultados do artigo (13).

Com base nestes resultados, foi investigada a produção de texto na rede, com foco na potencial relação entre topologia, setor de Erdős e texto produzido (34). As principais conclusões são:

- O texto produzido por cada setor de Erdős é bastante diferente um do outro: os  $c(\alpha)$  fruto do teste de Kolmogorov-Smirnov entre histogramas de uso de recursos textuais (pontuação, adjetivos, etc) de cada setor são tão grandes que as tabelas não registram os valores (veja Seção 3.5). Além disso, as diferenças entre  $c(\alpha)$  de setores iguais de redes diferentes são, na grande maioria das vezes, menores que as encontradas entre setores diferentes de uma mesma rede. Isso decorre de uma maior discrepância de massa probabilística entre os histogramas de setores diferentes de uma mesma rede do que entre setores iguais de redes diferentes.
- As características topológicas e textuais de cada agente apresentam correlações não triviais (como entre centralidade de intermediação e uso de advérbios) e triviais (como entre grau e número de caracteres escritos). Mesmo assim, são muito menos correlacionadas entre si do que separadamente. Ou seja, as componentes principais possuem tendência à prevalência de medidas topológicas **ou** textuais, mas a combinação de medidas de ambos os tipos é incidente.

Estes resultados permearam várias outras frentes de pesquisa e desenvolvimento tecnológico (33, 35–38).

## 4.2 Criação da nuvem brasileira de dados participativos ligados

Iniciada para formalizar as redes e participantes, estabelecer *benchmarks* (valores de referência), e observar os aspectos mandatórios, e relativamente estáveis, das conceitualizações

sobre as estruturas sociais. Esta frente rapidamente se voltou para as formalizações de conceitualizações referentes às estruturas e sistemáticas já em prática e previstas em lei. Dados também foram associados às ontologias feitas. Estes dados e ontologias foram em grande parte já publicados e estão em uso, mas a grande maioria não recebeu artigo científico ainda (28, 33, 35–37). Foi publicado no arXiv somente um artigo sobre a OPS (37). Este escrito aguarda confluência com orientador para publicação em revista, potencialmente na revista PLOS ONE. Foram escritos também os produtos PNUD/ONU, publicados em instâncias governamentais e em repositórios públicos (33, 35, 36). Foram publicados dados em RDF do Participa.br, Cidade Democrática e AA no Datahub.io (39). Ontologias e vocabulários foram publicadas junto ao ministério do planejamento e em repositórios públicos (33). *Scripts* para síntese das ontologias e triplificação de dados estão também publicamente acessíveis e junto aos produtos PNUD da bibliografia.

#### 4.2.1 Síntese de ontologias e vocabulários de estruturas sociais

Ontologias OWL feitas neste trabalho:

- OPS (Ontologia de Participação Social, fruto de diversos esforços da América Latina, principalmente do Brasil): nesta pesquisa, revisamos a ontologia e disponibilizamos a versão em uso por instâncias diferentes da academia, Estado e sociedade civil (37).
- OPa (Ontologia do Participa.br): esta é uma ontologia feita para e a partir dos dados do Participa.br (Portal Federal de Participação Social, SG-PR).
- OPP (Ontologia de Portais Participativos): pensada com a equipe do Participa.br e outros especialistas como esquema geral de portais participativos. Ontologia relativamente complexa, centrada em 3 classes: Participante, Comunidade, Mecanismo Participativo.
- Ontologiaa (Ontologia do AA): uma pequena ontologia para o minimalista AA (Autor-regulação Algorítmica), um software para registrar e compartilhar processos intelectuais como para pesquisa e arte (33, 40, 41).
- OCD (Ontologia do Cidade Democrática): é uma ontologia extensa para o portal participativo Cidade Democrática, da sociedade civil. Dado o tamanho da ontologia, o processo de sua construção deu origem ao método de construção de ontologia OWL a partir dos dados, descrito na Seção 4.2.3 e utilizado também para a construção da OPa (acima).
- OBS (Ontologia da Biblioteca Social): uma coleção de ontologias, uma para cada conceito que precisasse, e uma para cada mecanismo ou instância de participação social

prevista no Decreto Presidencial nº 8.243, conhecido como decreto da PNPS ou da Política Nacional de Participação Social. Esta ontologia contou com entrevistas feitas diretamente para construí-la, e uma atividade especial na Secretaria-Geral da Presidência da República, para explicitar a utilidade destas formalizações semânticas e coletar informações sobre diversos mecanismos e instâncias de participação social previstos em lei e praticados (33).

O VBS (Vocabulário da Biblioteca Social) é uma adaptação (com complementos) da OBS no formato de vocabulário SKOS, principalmente para facilitar usos junto ao DSPACE.

As ontologias e vocabulários são todas construídas através de scripts, com exceção da OPP, feita no Protegé (33).

#### **4.2.2 Obtenção de dados ligados a partir de dados relacionais participativos**

Roteiros para conversão de dados relacionais em dados RDF enriquecidos semanticamente (33):

- Triplificação do Participa.br: dados originalmente em PostgreSQL. São usados, através de buscas SparQL, para auxiliar na construção da OPa.
- Triplificação dos dados do Cidade Democrática. Estes dados são utilizados para auxiliar na construção da OCD.
- Triplificação dos dados do AA: dados do AA encontrados em bancos de dados MySQL e MongoDB, e em *logs* de IRC (33, 41). Esta foi a única triplificação feita depois da ontologia e não aproveitada para a construção da ontologia.

#### **4.2.3 Método de construção de ontologias orientado aos dados**

Um método de levantamento de ontologia orientado aos dados surgiu, potencialmente útil a todos os portais e software em necessidade de ontologias, e foi responsável por 2 ontologias (OPa e OCD). Resumidamente, o método consiste em: representar os dados de interesse como RDF; realizar buscas SparQL para construir ontologia trivial com as classes e propriedades encontradas; realizar buscas SparQL para inferir restrições de classe e axiomas de propriedade (33).

## 4.3 Aparato em software

*Scripts* para verificar as estabilidades topológicas e diferenciações textuais em redes humanas estão reunidos em um pacote oficial da linguagem Python (10). Estão sendo feitos pacotes para organizar os numerosos *scripts* de triplificação de dados, construção de ontologias e vocabulários e mineração das estruturas (42, 43). Os dados, classes e propriedades das ontologias e triplificações estão também disponíveis (em parte) através das próprias URIs, redirecionadas via [purl.org](http://purl.org) para um servidor de pesquisa. Ou seja, caso você acesse <http://purl.org/socialparticipation/opa/Participant>, o servidor em <http://purl.org> redireciona seu navegador para um servidor de pesquisa com várias entidades do conceito “Participante” da ontologia “opa”. Os dados estão em um *endpoint* SparQL, e *scripts* para a mineração destes dados estão disponíveis em interfaces web via um IPython Notebook. As ontologias estão também disponíveis na instalação do Webprotegé da Stanford (33). Muitas engenhocas foram criadas para gerar figuras, vídeos e inspecionar estruturas sociais de emails, Facebook, Twitter, Participa.br, AA, IRC e outras fontes (22, 25, 28, 41). Outras engenhocas foram criadas para experimentações estéticas e informacionais (10, 44–46).

## 4.4 Benefício, utilidade e formalismo

Esta seção complementa a monografia neste aspecto: registra andamentos fronteiriços do trabalho, difíceis de formalizar e até inconclusivos, mas cuja utilidade para o participante é latente.

### 4.4.1 Sistemas de recomendação para o enriquecimento da navegação semântica de recursos

O relacionamento semântico de dados e conceitualizações via tecnologias de web semântica torna os recursos navegáveis à semelhança do que fazemos com os navegadores Web ao abrir páginas HTML (por isso a área chama-se **web** semântica). Ao invés de páginas HTML, os recursos são formatados em RDF e os links são consequência de critérios semânticos. No decorrer desta pesquisa, surgiram possibilidades de enriquecimento da navegação semântica através de recomendações de recursos com métodos abertos e propostas de aproveitamento pelo usuário. Esta versão recebeu prova de conceito (36):

- São geradas estruturas auxiliares: rede de amizades, rede de interação, histograma de

radicais (morfemas do texto), seleção dos 400 radicais mais incidentes para caracterizar o domínio, histograma de radicais de cada recurso (postagem, comentário, participante, etc.).

- O solicitante pode requerer recomendação de recursos de qualquer tipo a partir de um recurso de qualquer tipo. Pode optar pelo método de recomendação topológico (utilizando as redes de amizade e interação), textual (utilizando os histogramas de radicais) ou híbrido (utilizando ambos). Pode optar por polaridade de similaridade (recomenda recursos similares), dissimilaridade (recomenda recursos dissimilares) ou mista (mistura de recursos similares e dissimilares ou recomendação na qual essa classificação não se aplica).
- Os métodos são todos explicitados em texto e código para o participante. Cada método conta também com um registro de potenciais utilidades para o participante, assim como cada recomendação.

#### 4.4.2 Experimentos de percolação social e a física antropológica

Foram realizados procedimentos cíclicos e procedimentos efêmeros de difusão de informação para observar as reações e testar hipóteses de modificação das estruturas sociais. Experimentos paradigmáticos e hipóteses serão expostas nesta seção. Em dezembro de 2012 foram iniciados ciclos de coleta e difusão de informação sobre as redes sociais e o potencial benéfico para o indivíduo civil. Duraram meses e redes diferentes foram usadas, todas redes das quais faço parte. Foram confirmadas as hipóteses: de modificação das estruturas sociais para comportar a pesquisa, com suporte humano, financeiro e institucional; de modificação do tratamento da sociedade sobre o tema; de que seriam verificáveis estes resultados em minhas interações cotidianas. Estas foram algumas consequências da “percolação do tecido social” (mudança abrupta das propriedades físicas do tecido social acompanhado de mudança gradual de conectividade). Em especial, a minha rede de amizades do Facebook foi utilizada (cada vértice é um amigo meu, cada aresta indica uma amizade entre eles), e amigo por amigo foi acionado, dos menos conectados aos mais conectados, três vezes (47).

Em outra ocasião, percebi que uma característica não intuitiva: em praticamente qualquer rede de amizades com mais de 500 pessoas, dentre as 50 pessoas com a maior intermediação (*betweenness centrality*, mais participa de geodésicas) havia sempre pouquíssimas que constavam também dentre as 50 com maior *closeness centrality* (mais perto de todos os outros agentes). Selecionamos estes dois grupos em minha rede e em redes de parceiros que fazem

experimentos semelhantes. Cada pessoa enviou uma mensagem diferente, cada grupo de cada pessoa recebeu uma cópia desta mensagem. O grupo com a maior intermediação reagiu sempre calorosamente, repassava a mensagem, os membros até interagiam entre si, mesmo sem se conhecerem ou serem próximos. Os grupos de maior *closeness* nunca reagiam, membros saíam rapidamente da interface. A hipótese mais plausível que surgiu para explicar esta diferença de reação é a de que os membros de maior intermediação tinham maior influência sobre a rede, enquanto os de maior *closeness* sofriam maior influência.

Em um evento grande em São Paulo sobre transparência e governança na internet (*#arenaNETmundial*), foi operacionalizado um telão de streaming de estruturas sociais, escrito no percurso desta pesquisa, que expunha em tempo real as três diferentes redes de Twitter formada por usuários relacionados por *retweet*, vocabulário e *#hashtag*. A tecnologia pode ser compreendida como “*streaming* de estruturas sociais”, e gerou bastante reflexão com as pessoas que foram ao evento, inclusive com os próprios comunicadores que constavam nas redes de *retweet* (45). Houve confirmação da hipótese de que as pessoas se interessariam e se instruíam. Houve alguns usos a mais da ferramenta, localizados e a pedido do meio, não por necessidade da pesquisa.

O segundo experimento foi feito por vários parceiros de pesquisa, por onde pôde ser verificado o comportamento constante. O primeiro experimento ainda não foi replicado. É comum após alguma apresentação ou reunião de pesquisa alguém se prontificar a fazê-lo, mas isso nunca aconteceu. Eu mesmo já me comprometi comigo a replicar o experimento, mas não aconteceu. Uma hipótese usual é que haja bloqueios mentais que nos impedem de realizar uma intervenção tão direta na nossa existência social, ou nosso eu-rede (2, 48).

Estes experimentos, e outras anotações de dados, são, no escopo deste trabalho, considerados questionáveis, potencialmente inapropriados, caso não sejam observadas algumas diretrizes: estudo das redes das quais o pesquisador faz parte, como um estudo de si; uso de anotações (de si) com a devida atenção para não expor as pessoas desnecessariamente e para quaisquer maiores cuidados sugeridos pelo contexto; abertura constante dos procedimentos, dados, códigos e literatura produzida.

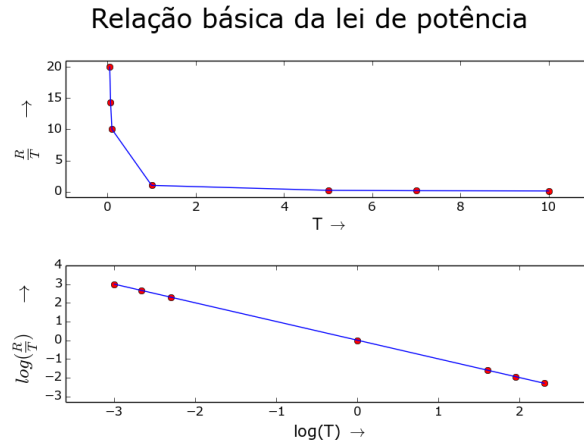
Estas diretrizes foram apreendidas em grande parte da tradição antropológica, e, portanto, configuram uma pesquisa com alguns aspectos “antropológicos”. O termo “física antropológica” começou a ser usado no Brasil principalmente por acadêmicos (físicos, cientistas da computação, filósofos, antropólogos) em 2013-14, no contexto dos experimentos de difusão de informação e das análises, ambos em minhas próprias redes. Considerações cuidadosas estão sendo feitas constantemente sobre o presente trabalho, sobre o termo, sobre o legado

antropológico, sobre a física e as redes complexas, e sobre termos relacionados, como física social (7) e sociofísica (49). Há resistência do meio científico, mas no geral o balanço aponta para uma pertinência do uso do termo para representar o que está sendo feito.

#### 4.4.3 Entendimento sobre as estruturas sociais

Há a intenção de disponibilizar um compêndio às redes complexas através da instrumentalização do leitor com estes conhecimentos e tecnologias para exploração de si próprio. Um esboço consta em (9). Um exemplo especial de fundamentação que parece não constar na literatura é a constatação de que a propriedade livre de escala apresenta uma distribuição uniforme de recursos com respeito à conectividade (e.g. grau ou força). Para apreender este fato, considere uma quantidade fixa  $R$  de recursos que será utilizada para a realização da rede em conectividade. Considere que, para cada quantidade de recursos  $T$ , são contadas  $f = \frac{R}{T}$  partes de tamanho  $T$ , como na Figura 4.1.

Figura 4.1 – A curva resultante da divisão de uma mesma quantidade  $R$  de recursos em  $\frac{R}{T}$  partes de tamanho  $T$ . Utilizada para expor uma potencial causa da ubiquidade da propriedade livre de escala.



Segue que  $\log(f) = -1\log(T) + C$ , com  $C = \log(R)$  uma constante arbitrária. Uma reta descreve a relação entre  $\log(f)$  e  $\log(T)$ , como na Figura 4.1. Os recursos são alocados pelo sistema de forma uniforme, pois  $T \frac{R}{T} = R = \text{constante}$ .

Considere que  $T = T_1 T_2$  (e.g. recursos da rede=agentes x tempo de cada agente). Neste caso,  $f = \frac{R}{T_1 T_2}$  e segue que  $\log(f) = -\log(T_1 T_2) + C$ . Se  $T_1 = T_2$ ,  $\log(f) = -2\log(T_1) + C$ , e  $\gamma = 2$  como previsto pela literatura. No exemplo, o tempo alocado é o tempo dos próprios agentes, portanto é razoável considerar  $T_1 = T_2$ . Possíveis causas para a distorção do valor exato  $\gamma = 2$  são: propriedades fractais, recursos em número diferente, associações entre os recursos.



## 5 - Cronograma e afazeres

Tabela 5.1 – Cronograma de atividades ao longo dos semestres, descritas na Seção 5. A marcação • indica previsão feita no início do doutorado. A marcação [ ] se refere ao relato e previsão, agora no final do 1º semestre de 2015. As principais diferenças do previsto foram: as disciplinas terminaram no primeiro ano; a revisão da literatura, os acréscimos aos modelos atuais com o foco no participante da rede, e a implementação computacional, estas três atividades estão sendo realizadas constantemente e devem durar até a entrega e defesa da tese.

	2013		2014		2015	
Atividade	1º	2º	1º	2º	1º	2º
1	[•]	[•]	•	•		
2	[•]	[•]	[•]	[ ]	[ ]	[ ]
3	[ ]	[•]	[•]	[•]	[•]	[ ]
4	[ ]	[•]	[•]	[•]	[•]	[•]
5					[•]	[•]
6	[•]	[•]	[•]	[•]	[•]	[•]
7	[•]	[•]	[•]	[•]	[•]	[•]

Este projeto foi inicialmente dividido segundo as etapas a seguir e usadas como referência na Tabela 5.1:

1. Cumprimento dos créditos obrigatórios em disciplinas\*.
2. Revisão da literatura †.
3. Acréscimos aos modelos atuais com o foco no participante da rede.
4. Implementação computacional ‡.
5. Escrita da tese §.
6. Escrita e publicação dos resultados em artigos ¶.
7. Trocas com pessoas externas, estabelecimento de colaborações ||.

\*Introdução ao Processamento de Língua Natural (SCC5908, 12 créditos), Mineração de Dados não Estruturados (SCC5920, 12 créditos), Visualização Computacional (SCC5836, 12 créditos), e Introdução à Web Semântica (SCC5929, 8 créditos). No mestrado, fazia mais de 20 créditos na por semestre na graduação, 6 disciplinas na pós em um ano (66 créditos) e pesquisa, e fechei todas com A. Estranhamente, no doutorado fechei todas as disciplinas com B, fiz menos disciplinas na pós, não fiz graduação e despendi tempo.

†A literatura para o trabalho proposto é ampla e este aprofundamento tem sido constante.

‡Há implementações computacionais de provas de conceito, bibliotecas, rotinas básicas e rotinas para replicar resultados do grupo de pesquisa. Engenhocas para gerar arte audiovisual a partir de redes.

§A escrita da tese pode tomar vários rumos: pode consistir de um conjunto de artigos ou de uma monografia final. Acho mais provável que seja um conjunto de artigos focados nas direções dadas na Seção 4.4.3.

¶Conseguimos finalizar um artigo (13). Há ao menos mais dois em condições de publicação (37, 50). Além destes, há mais estes artigos no arXiv (28, 34, 38, 51), todos referentes ao trabalho do doutorado. Foram publicados em revista os artigos sobre o AA (40) e análise quantitativa de pintura (52), ambos sem a colaboração do orientador.

||Os experimentos de coleta e difusão de informação dispararam reuniões, visitas e colaborações. Este processo foi iniciado logo antes do doutorado e pode ser apreciado, por exemplo, pelas visitas a São Carlos de parceiros de pesquisa, pela integração do pesquisador ao grupo de pesquisa Nexus, vinculado ao CNPq, e ao aporte do PNUD/ONU à pesquisa, sobre o qual a Presidência da República se posicionou como beneficiária (28).

Tabela 5.2 – Relação de tarefas feitas e por fazer. Há literatura pronta e vários documentos escritos e em mãos para serem aprofundados. O mais urgente parece ser uma revisão e aprofundamento de estatística e física estatística, e confirmar os experimentos percolatórios contínuos (veja Seção 4.4.2).

	feito	por fazer
<b>escrita</b>	artigo de estabilidade em redes de interação humana (13); artigo sobre a Ontologia de Participação Social (37); ensaio descrevendo simbiose com PNUD/ONU e SG-PR (28); artigo com descrição psicofísica da música no áudio digital (50); produtos PNUD 3, 4 e 5, descrevendo sistemas de classificação, recomendação, ontologias e triplificações para participação social com métodos de redes complexas e processamento de linguagem natural (33, 35, 36); artigo sobre AA (40); versões iniciais e rascunhos dos artigos sobre física antropológica (29), sobre votação contínua por aprovação e participação (51), sobre diferenças da produção textual nos setores de Erdős (34), sobre visualização de redes de interação em evolução temporal (38), sobre audiovisualização de redes de interação em evolução temporal (38), sobre performance audiovisual via controle coletivo de código e projeção ao vivo (53)	publicar artigos no arXiv; repassar produtos PNUD um e dois; “Complex Networks Gradus ad Parnassum”, um compêndio de redes complexas que utiliza a existência em rede do leitor para instrumentalizá-lo; artigo sobre tipologia de agentes humanos em redes de interação; versão desenvolvida do escrito sobre física antropológica; documentação do pacote Python oficial “percolation” (43); artigo com o método de levantamento de ontologias orientado aos dados; artigo sobre os dados participativos ligados brasileiros; artigo com os experimentos de coleta e difusão de informação; versão final do ensaio do AA (41)
<b>leitura</b>	documentação de redes complexas; documentação de web semântica; amadurecimentos coletivos frutos das difusões de informação; numerosos artigos da Wikipédia, protocolos e manuais de software; cursos do Coursera, alguns completos; literatura de PLN; literatura de visualização de dados e mineração de dados; artigos, exemplos especiais são (4, 17, 54, 55)	estatística e física estatística, talvez manuais de R também; terminar livros referência de redes complexas; absorver uma literatura mínima sobre antropologia; visita à topologia tradicional e teoria de grafos na computação
<b>experimentos</b>	experimentos contínuos/cíclicos e outros efêmeros	confirmar experimentos contínuos/cíclicos
<b>comunidade</b>	repassados resultados para comunidades estudadas; confirmada permissão dos desenvolvedores do Gmane para utilizar os dados das listas para pesquisa	repassar às comunidades estudadas um resumo dos resultados, em linguagem mais acessível que os artigos
<b>disciplinas</b>	cursadas disciplinas Introdução ao Processamento de Linguagem Natural, Mineração de dados; Visualização de dados; Introdução à web semântica	-/-
<b>considerar banca</b>	-/-	preparar apresentação; apresentar e anotar contribuição da banca; conduzir com orientador
<b>software</b>	telões de streaming de estruturas sociais; funcionalidades escolhidas da MMISSA (Monitoramento Massivo e Interativo da Sociedade pela Sociedade para Aproveitamento); engenhoras no AARS (A Análise de Redes Sociais) e MyNSA ( <i>Monitoring yields Natural Streaming and Analysis</i> ); rotinas de triplificação de dados; rotinas de construção de ontologias; rotinas para, dada a rede social, sintetizar música e animação visual sincronizados; rotinas com fundamentos e provas de conceitos para genérica classificação e recomendação de recursos	finalizar pacotes oficiais da linguagem Python; estação de monitoramento massivo; sistema de navegação semântica enriquecido com recomendação de recursos
<b>dados</b>	dados triplificados do Participa.br, do Cidade Democrática, do AA	revisar dados triplificados; triplificar dados do Facebook, Twitter e listas de email
<b>ontologias e vocabulários</b>	OPS, OPa, OPP, OCD, Ontologiaa, OBS e VBS iniciais	ontologias e vocabulários revisados
<b>audiovisualização</b>	versinus; prelúdio social; four hubs dance	músicas focando em algum dos participantes da rede; mais músicas sobre as redes do Facebook; mais músicas sobre as redes de email; rotinas para fazer animação abstrata sobre rede de interação e mixar com clipe do youtube; sonificação de dados semânticos e renderização de imagens sincronizadas

## 6 - Conclusões e previsão

---

Há, a princípio, uma confirmação de que os conhecimentos de redes complexas possuem aplicações diversas e potencialmente benéficas para o participante. Por exemplo, os experimentos apresentaram modificações da estrutura social para comportar a pesquisa, e podem ser usados para comportar outros empreendimentos. Os estudos de estabilidade e diferenciação em redes de interação humana apontam na direção de tipologias de redes e de participantes, com base nos setores de Erdős e com componentes principais típicas e estáveis. Um legado conceitualizações formalizadas associado a dados ligados e abertos é conveniente para *benchmarks* e para apresentar estes resultados às comunidades acadêmicas e interessadas nas aplicações, para as quais foram adiantadas ontologias, vocabulários, rotinas de conversão de dados relacionais em RDF e os dados em si.

Há, em alguns casos extremos, considerações na base da área, com implicações sobre a própria constituição das redes complexas (como na Seção 4.4.3). Ao mesmo tempo, os métodos utilizados são potencialmente novos (como na Seção 3.3). Há diversos trabalhos na bibliografia e, caso haja disponibilidade para visitar itens da literatura produzida, recomendamos, nesta mesma ordem, (10, 13, 24, 28, 33). Dois exemplos de trabalhos de terceiros com influência direta do material nesta monografia, são a tese de doutorado defendida ano passado (56); e o relatório de consultoria prestada por professores da UnB ao PNUD junto à SG-PR (57).

Uma direção simples para concluir a pesquisa consiste em focar no documento *Complex Networks Gradus ad Parnassum*, que está planejado como uma apresentação das redes complexas através da entrega, para o leitor, de formas de observar e interagir com suas redes, beneficiando-se (9). Uma direção menos pedagógica, porém mais usual e simples, é explorar as estabilidades encontradas: até que número de agentes a distribuição dos participantes nos setores e a formação das componentes principais se mantém? Para quais redes? Como caracterizar a intermitência dos agentes enquanto a distribuição de grau é estável? Se o texto produzido pelos setores é diferente, em quais aspectos é igual e em quais se diferencia? Os resultados se mantêm em ambas as línguas português e inglês?

O orientador solicitou ênfase do conteúdo do artigo de estabilidade temporal (13) nesta monografia, o que me esforcei por fazer. Os complementos principais ao conteúdo aqui apresentado são: a seção de resultados do corpo do artigo, e o documento de informações de suporte (58).

## **Agradecimentos**

Autores agradecem o apoio financeiro concedido pelo CNPq (140860/2013-4, projeto 870336/1997-5), Programa das Nações Unidas para o Desenvolvimento (contrato 2013/000566, projeto BRA/12/018) e FAPESP; a prestatividade do corpo do IFSC, da CPG e dos funcionários; o espaço concedido pelo IEASC/USP para algumas reuniões e visitas de parceiros de pesquisa; o suporte intelectual e o interesse dos membros do Nexos-Sudeste (Grupo de pesquisa/CNPq (26)); ao suporte institucional e intelectual da Secretaria Geral da Presidência da República; ao suporte intelectual e tecnológico do labMacambira.sf.net e todas as comunidades de software e cultura livre direta e indiretamente relacionadas a este trabalho.

# Referências

---

- 1 Mark Newman. *Networks: an introduction*. Oxford University Press, 2010.
- 2 Bruno Latour. Reassembling the social. an introduction to actor-network-theory. *Journal of Economic Sociology*, 14(2):73–87, 2013.
- 3 Christian Bird, Alex Gourley, Prem Devanbu, Michael Gertz, and Anand Swaminathan. Mining email social networks. In *Proceedings of the 2006 international workshop on Mining software repositories*, pages 137–143. ACM, 2006.
- 4 Alexei Vázquez, João Gama Oliveira, Zoltán Dezső, Kwang-II Goh, Imre Kondor, and Albert-László Barabási. Modeling bursts and heavy tails in human dynamics. *Physical Review E*, 73(3):036127, 2006.
- 5 Brian Ball and Mark EJ Newman. Friendship networks and social status. *arXiv preprint arXiv:1205.6822*, 2012.
- 6 A Petland. Reinventing society in the wake of big data. edge. org, 2012. <http://edge.org/conversation/reinventing-society-in-the-wake-of-big-data>.
- 7 Alex Pentland. *Social Physics: How Good Ideas Spread-The Lessons from a New Science*. Penguin, 2014.
- 8 NetworkX Developers. Networkx, 2010.
- 9 R. Fabbri. Complex networks gradus ad parnassum. 2015. <https://github.com/ttm/gradus/raw/master/article.pdf>.
- 10 Renato Fabbri. Python package to observe time stability in the gmane database, 2015. <https://pypi.python.org/pypi/gmane>.
- 11 Wikipedia. Gmane — Wikipedia, the free encyclopedia, 2013. Online; accessed 27-October-2013.
- 12 Kyle Marek-Spartz, Paula Chesley, and Hannah Sande. Construction of the gmane corpus for examining the diffusion of lexical innovations. 2012.
- 13 R. Fabbri. Time stability in human interaction networks. *arXiv preprint arXiv:1310.7769*, 2015. <http://arxiv.org/abs/1310.7769>.
- 14 Matthew O. Jackson. Social and economic networks: Models and analysis, 2013. <https://class.coursera.org/networksonline-001>.
- 15 Wikipedia. Directional statistics, 2015. [Online; accessed 16-June-2015].
- 16 Elizabeth A Leicht and Mark EJ Newman. Community structure in directed networks. *Physical review letters*, 100(11):118703, 2008.
- 17 MEJ Newman. Community detection and graph partitioning. *arXiv preprint arXiv:1305.4974*, 2013.
- 18 Ian Jolliffe. *Principal component analysis*. Wiley Online Library, 2005.
- 19 Wikipedia. Kolmogorov–smirnov test, 2015. [Online; accessed 16-June-2015].

- 20 Renato Fabbri. Video visualizations of email interaction network evolution, 2013-5. [https://www.youtube.com/playlist?list=PLf\\_EtaMqu3jVodaqDjN7yaSgsQx2Xna3d](https://www.youtube.com/playlist?list=PLf_EtaMqu3jVodaqDjN7yaSgsQx2Xna3d), [https://www.youtube.com/playlist?list=PLf\\_EtaMqu3jWYQiJZYhVlJVngb7vsf6na](https://www.youtube.com/playlist?list=PLf_EtaMqu3jWYQiJZYhVlJVngb7vsf6na), [https://www.youtube.com/playlist?list=PLf\\_EtaMqu3jVb7CTt59t3ZnrmXuGON3c0](https://www.youtube.com/playlist?list=PLf_EtaMqu3jVb7CTt59t3ZnrmXuGON3c0), [https://www.youtube.com/playlist?list=PLf\\_EtaMqu3jVFS\\_AJZm\\_Hu09pywnSWaNF](https://www.youtube.com/playlist?list=PLf_EtaMqu3jVFS_AJZm_Hu09pywnSWaNF), [https://www.youtube.com/playlist?list=PLf\\_EtaMqu3jU-1j4jiIUiyMqyVSzIYeh6](https://www.youtube.com/playlist?list=PLf_EtaMqu3jU-1j4jiIUiyMqyVSzIYeh6), [https://www.youtube.com/playlist?list=PLf\\_EtaMqu3jUZpAX3cKPC5J0t3q836CLy](https://www.youtube.com/playlist?list=PLf_EtaMqu3jUZpAX3cKPC5J0t3q836CLy), [https://www.youtube.com/playlist?list=PLf\\_EtaMqu3jUY0\\_XfJdqQELdbFnpqYEfb](https://www.youtube.com/playlist?list=PLf_EtaMqu3jUY0_XfJdqQELdbFnpqYEfb).
- 21 Renato Fabbri. Image gallery of email interaction networks., 2013. [http://hera.ethymos.com.br:1080/redes/python/autoRede/gmane.linux.audio.devel\\_3000-4200-280/](http://hera.ethymos.com.br:1080/redes/python/autoRede/gmane.linux.audio.devel_3000-4200-280/).
- 22 Renato Fabbri. Online gadget for making email interaction network images, gml files and measurements, 2013. <http://hera.ethymos.com.br:1080/redes/python/autoRede/escolheRedes.php>.
- 23 R. Fabbri. Prelúdio social (audiovisualização de rede social), 2015. <https://www.youtube.com/watch?v=9uIiGQBWYo>.
- 24 R. Fabbri. Four hubs dance (audiovisualização de rede social), 2015. <https://www.youtube.com/watch?v=1EOnEVmqmmc>.
- 25 Renato Fabbri. Python package for social data scraping, analysis and art., 2015. <https://pypi.python.org/pypi/social>.
- 26 Grupo nexos: Teoria crítica e pesquisa interdisciplinar. <http://dgp.cnpq.br/dgp/espelhogrupo/5624425913774111>. Acessado: 2015-Jun.
- 27 M. Canevacci. Antropologia, psicanálise, comunicação. 2012. <https://goo.gl/rZoR9D>.
- 28 Renato Fabbri. Ensaio sobre o auto-aproveitamento: um relato de investidas naturais na participação social. *arXiv preprint arXiv:1412.6868*, 2014.
- 29 R. Fabbri. What are you and i? [anthropological physics fundamentals], 2015. [https://www.academia.edu/10356773/What\\_are\\_you\\_and\\_I\\_anthropological\\_physics\\_fundamentals\\_](https://www.academia.edu/10356773/What_are_you_and_I_anthropological_physics_fundamentals_).
- 30 World Wide Web Consortium et al. Rdf 1.1 concepts and abstract syntax. 2014.
- 31 World Wide Web Consortium et al. Owl 2 web ontology language document overview. 2012.
- 32 Alistair Miles and Sean Bechhofer. Skos simple knowledge organization system reference. *W3C recommendation*, 18:W3C, 2009.
- 33 R. Fabbri. Proposta de regras de extração de conteúdos da api do portal e suas ferramentas para alimentação de eventual/hipotética base/nuvem de conhecimento de participação social. 2014. <https://github.com/ttm/pnud5/blob/master/latex/produto.pdf?raw=true>.
- 34 R. Fabbri. A connective differentiation of textual production in interaction networks. 2013. <http://arxiv.org/abs/1412.7309>.
- 35 R. Fabbri. *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. <https://github.com/ttm/pnud3/blob/master/latex/produto.pdf?raw=true>.

- 36 R. Fabbri. *adaptações e incrementos para a interface do portal federal de participação social e suas ferramentas*, 2014. <https://github.com/ttm/pnud4/blob/master/latex/produto.pdf?raw=true>.
- 37 Renato Fabbri, Rodrigo Bandeira de Luna, Ricardo Augusto Poppi Martins, et al. Social participation ontology: community documentation, enhancements and use examples. *arXiv preprint arXiv:1501.02662*, 2015. <http://arxiv.org/abs/1501.02662>.
- 38 R. Fabbri. Versinus: a visualization method for graphs in evolution. *arXiv preprint arXiv:1412.7311*, 2013. <http://arxiv.org/abs/1412.7311>.
- 39 R. Fabbri. Brazilian social participation data from Participa.br, Cidade Democrática and AA, in XML/RDF and Turtle/RDF, 2014. <http://datahub.io/organization/socialparticipation>.
- 40 Renato Fabbri, Ricardo Fabbri, Vilson Vieira, Daniel Penalva, Danilo Shiga, Marcos Mendonça, Alexandre Negrão, Lucas Zambianchi, and Gabriela Salvador Thumé. The algorithmic autoregulation software development methodology. *Revista Eletrônica de Sistemas de Informação ISSN 1677-3071 doi: 10.5329/RESI*, 13(2), 2014.
- 41 R. Fabbri. *The Algorithmic-Autoregulation essay: a collective and natural focus on self-transparency*. <https://github.com/ttm/ensaio/raw/master/ensaio.pdf>.
- 42 R. Fabbri. Python package with routines for analysis and synthesis of rdf social participation data, 2015. <https://pypi.python.org/pypi/participation>.
- 43 R. Fabbri. Python package for anthropological physics and social harnessing, 2015. <https://pypi.python.org/pypi/percolation>.
- 44 MMISSA: Monitoramento massivo e interativo da sociedade pela sociedade para aproveitamento. <http://mmissa.meteor.com>. Acessado: 2014-Dez-16.
- 45 Telões de streaming de estruturas sociais para o #ocupagov. <http://ocupagov.meteor.com>. Acessado: 2014-Dez-16.
- 46 Música social e govern art. <http://mm.meteor.com>. Acessado: 2014-Dez-16.
- 47 Receitas para percolação da própria rede. <https://dl.dropboxusercontent.com/u/22209842/doc/mit/difusao.pdf>. Acessado: 2015-Jun.
- 48 R. Fabbri. Nuvens cognitivas e a unificação da espécie humana. *CyberiuN Revista eletrônica*, 2013. <http://wiki.nosdigitais.teia.org.br/CyberiuN>, <http://cyberiuN.tumblr.com/post/64607669758/nuvens-cognitivas>.
- 49 Serge Galam. Sociophysics: A review of galam models. *International Journal of Modern Physics C*, 19(03):409–440, 2008.
- 50 Renato Fabbri, Vilson Vieira da Silva Junior, Antônio Carlos Silvano Pessotti, and Débora Cristina Corrêa. Psychophysics of musical elements in the discrete-time representation of sound. *arXiv preprint arXiv:1412.6853*, 2014.
- 51 Renato Fabbri and Ricardo Poppi. Continuous voting by approval and participation. *arXiv preprint arXiv:1505.06640*, 2015.
- 52 Vilson Vieira, Renato Fabbri, David Sbrissa, Luciano da Fontoura Costa, and Gonzalo Travieso. A quantitative approach to painting styles. *Physica A: Statistical Mechanics and its Applications*, 417:110–129, 2015.

- 53 Vilson Vieira, Guilherme Lunhany, Geraldo Magela de Castro Rocha Junior, Caleb Mascarenhas Luporini, Daniel Penalva, Ricardo Fabbri, and Renato Fabbri. Vivace: A collaborative live coding language. *arXiv preprint arXiv:1502.01312*, 2015.
- 54 Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.
- 55 Gergely Palla, Albert-László Barabási, and Tamás Vicsek. Quantifying social group evolution. *Nature*, 446(7136):664–667, 2007.
- 56 Chandra Wood Viegas. *Línguas em rede: para o fortalecimento da língua e da cultura Kokama*, 2015. <https://github.com/ttm/tese/blob/master/bib/chandra/VIEGASchandrawood.pdf?raw=true>.
- 57 P. Cruz, F. e Meireles. *Documento com proposta de extensão de ontologia para o Portal de Participação Social, dados de Conferências Nacionais e Conselhos Nacionais, incluindo propostas de articulação entre os dados desses instrumentos*. [https://github.com/ttm/tese/blob/master/bib/sgpr/pnud\\_fwcruez\\_pmeirelles\\_produto06\\_vfinal.docx?raw=true](https://github.com/ttm/tese/blob/master/bib/sgpr/pnud_fwcruez_pmeirelles_produto06_vfinal.docx?raw=true).
- 58 R. Fabbri. [upporting information] time stability in human interaction networks. *arXiv preprint arXiv:1310.7769*, 2015. <http://arxiv.org/src/1310.7769v6/anc/supportingInformation.pdf>.