

ESTABILIDADE TOPOLÓGICA E DIFERENCIAÇÃO TEXTUAL EM REDES DE INTERAÇÃO HUMANA

orientador: Prof. Dr. Osvaldo N. de Oliveira jr.
candidato: Renato Fabbri

SIFSC 6

INSTITUTO DE FÍSICA DE SÃO CARLOS

UNIVERSIDADE DE SÃO PAULO

03 de Outubro, 2016



Introdução

- redes complexas e de interação humana

Materiais

- dados de email, Facebook, Twitter, IRC, Participa.br, AA

Métodos

- estatística circular - redes de interação - setorialização de Erdős
- PCA de medidas topológicas - Kolmogorov-Smirnov para textos
- audiovisualização de dados
- considerações tipológicas e humanísticas

Resultados

- estabilidade temporal e topológica - diferenciação textual
- dados sociais ligados - peças artísticas e mapeamentos sensoriais
- software

Conclusões

Introdução

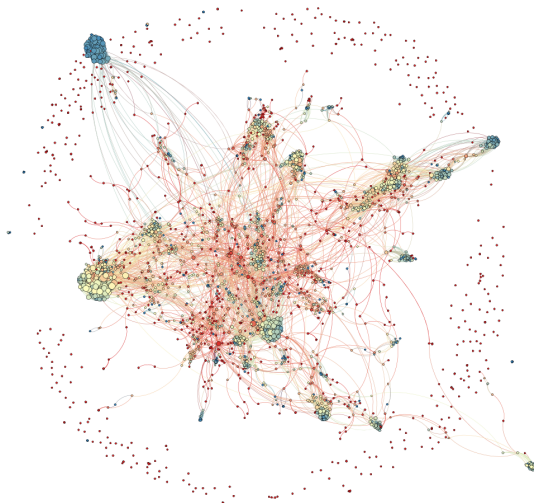
São 10^{80} **átomos no universo observável**, uma referência de escala. Considere o número N de pessoas necessário para haver **mais redes possíveis do que átomos no universo**. Cada aresta é uma variável de Bernoulli fruto de cada par de vértices: a aresta pode estar presente ou não.

$$\begin{aligned} 2^{\binom{N}{2}} > 10^{80} &\Rightarrow \log_2[2^{\binom{N}{2}}] > \log_2(10^{80}) \Rightarrow \binom{N}{2} > \frac{\log_{10}(10^{80})}{\log_{10}2} \Rightarrow \\ &\Rightarrow \frac{N \cdot (N - 1)}{2} > \frac{80}{\log_{10}2} \Rightarrow N > 23,5988 \end{aligned}$$

Isso justifica a utilidade de **paradigmas** para as redes, e das **medidas genéricas** para cada vértice e para a rede, instrumental para as **redes complexas**, incluindo as **redes de interação humana**.

Sistema complexo \Rightarrow Constituído de várias partes cuja interação exhibe comportamento emergente. É usual considerar que um sistema complexo: processa informação, exhibe mecanismos adaptativos, pode exhibir mecanismos de reprodução. Um sistema complexo é integrado a outros sistemas complexos e ao meio em que subsiste.

Introdução



- objetivos e justificativas

O objetivo geral é aprofundar a caracterização das redes virtuais de interação humana com a teoria das redes complexas. Objetivos específicos:

- melhor compreensão sobre nossas estruturas sociais
- entrega de um legado tecnológico para registrar e disponibilizar os desenvolvimentos
- entrega de um legado em dados sociais com fontes diversas
- delineio da prática de estudo das próprias estruturas sociais do pesquisador

A justificativa principal é que há um hiato entre a teoria atual de redes complexas e a caracterização das estruturas de interação humana representadas como redes complexas. Justificativas secundárias são:

- a área de redes complexas é recente e reconhecidamente útil
- a caracterização das redes complexas derivadas de estruturas sociais possui aspectos qualitativos passíveis de quantização
- os textos produzidos pelos setores dos hubs, intermediários e periféricos ainda não foram comparados entre si
- tenho um perfil adequado à transdisciplinaridade envolvida

Materiais

- mensagens de e-mail, com horário de envio, ID da mensagem, ID da mensagem anterior na thread se existente, ID do remetente, texto do título e corpo
- redes de Facebook: em formato GML ou GDF geralmente baixadas do Graphviz, mas também raspadas de minha própria conta. As únicas informações da rede são: nome e ID de cada amigo, aresta entre cada par de amigos que forem amigos entre si. Nas redes de interação constam arestas dirigidas. As redes eram de pessoas que me mandavam elas de suas contas ou minhas pessoais ou de grupos dos quais participava
- IRC: logs de canais de IRC podem auxiliar na generalização dos resultados
- Twitter: milhares (talvez milhões) de tweets permitiram observação de redes de interação (retweet), relacionamentos por vocabulário e hashtags, e padrões do vocabulário em si
- Participa.br: redes de amizade e de interação, texto de postagens, comentários, etc

- estatística circular
- obtenção das redes de interação
- setorialização de Erdős
- PCA de medidas topológicas
- testes de Kolmogorov-Smirnov dos textos
- audiovisualização de dados
- considerações tipológicas e humanísticas

- estatística circular

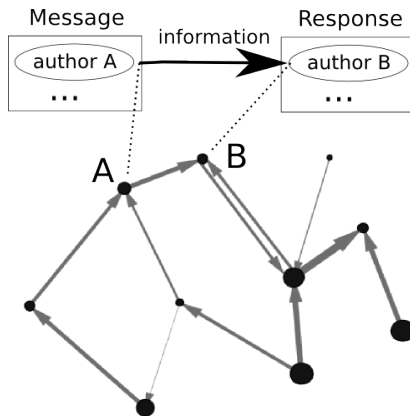
Considere $\theta = 2\pi \frac{\text{medida}}{\text{periodo}}$, $z_i = e^{i\theta}$ e $m_n = \frac{1}{N} \sum_{i=1}^N z_i^n$ o n-ésimo momento:

$$\begin{aligned} R_n &= |m_n| \\ \theta_\mu &= \text{Arg}(m_1) \\ \theta'_\mu &= \frac{\text{period}}{2\pi} \theta_\mu \end{aligned} \tag{1}$$

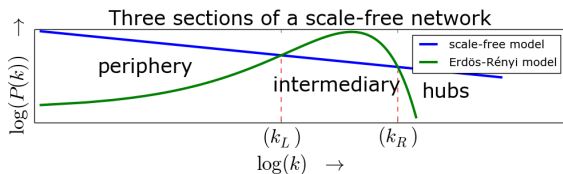
$$\begin{aligned} \text{Var}(z) &= 1 - R_1 \\ S(z) &= \sqrt{-2 \ln(R_1)} \\ \delta(z) &= \frac{1 - R_2}{2R_1^2} \end{aligned} \tag{2}$$

Usamos também $\frac{b_h}{b_l}$ entre a maior b_h e a menor b_l incidência nos histogramas.

- redes de interação



- setorialização de Erdős



$$\sum_{x=k_i}^{k_j} \tilde{P}(x) < \sum_{x=k_i}^{k_j} P(x) \Rightarrow i \text{ é intermediário} \quad (3)$$

$$P(k) = \binom{2(N-1)}{k} p_e^k (1-p_e)^{2(N-1)-k} \quad (4)$$

$$\text{onde } p_e = \frac{z}{N(N-1)}$$

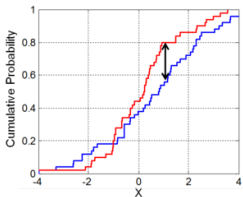
- PCA de medidas topológicas

Médias e desvios das medidas j nas componentes k fruto de L observações l :

$$\begin{aligned}\mu_{V'}[j, k] &= \frac{\sum_l^L V'[j, k, l]}{L} \\ \sigma_{V'}[j, k] &= \sqrt{\frac{(\mu_{V'} - V'[j, k, l])^2}{L}} \\ \mu_{D'}[k] &= \frac{\sum_l^L D'[k, l]}{L} \\ \sigma_{D'}[k] &= \sqrt{\frac{(\mu_{D'} - D'[k, l])^2}{L}}\end{aligned}\tag{5}$$

Foco nas medidas de centralidade e clusterização mais usuais. Inseridas medidas de simetria potencialmente novas.

- teste de Kolmogorov-Smirnov para incidências em textos

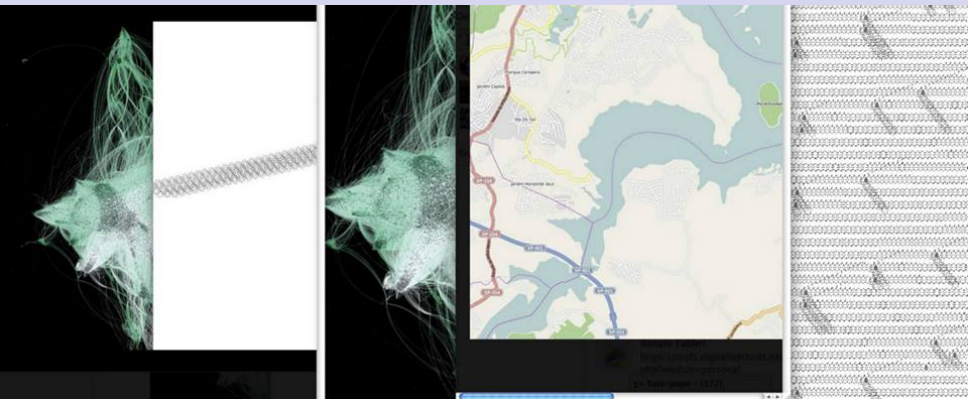


$$D_{n,n'} > c(\alpha) \sqrt{\frac{n+n'}{nn'}} \Rightarrow F_{1,n} \neq F_{2,n'} \quad (6)$$

$$c(\alpha) < \frac{D_{n,n'}}{\sqrt{\frac{n+n'}{nn'}}} = c'(\alpha) \quad (7)$$

α	0.1	0.05	0.025	0.01	0.005	0.001
$c(\alpha)$	1.22	1.36	1.48	1.63	1.73	1.95

- audiovisualização de dados



Permite maior contato com as estruturas de interesse, o que facilita a condução da pesquisa para questões mais fundamentais e observáveis. Algumas estratégias utilizadas:

- versinus, imagens, animação abstrata com música, sonificações
- roteiros automatizados de realização de arte social
- freakcoding (subgênero do livecoding)
- arte governamental

- considerações tipológicas e humanísticas

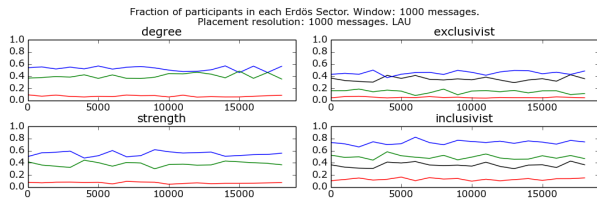
- redes de seres humanos
- consideração do fator estigmatizante
- apreciação do meio em que a rede é observada
- experimentos percolatórios
- física antropológica

- estabilidade temporal
- diferenciação textual
- iniciação da nuvem brasileira de dados ligados participativos
- aparato em software

- estabilidade temporal e topológica

- medidas circulares praticamente iguais para todas as listas e em todas as escalas de segundos a semestres
- constância dos tamanhos dos setores de Erdős, compatível com as expectativas da literatura. Ainda não achamos formalização para esta expectativa e talvez esta seja a primeira
- estabilidade das componentes principais.
- tipologia não estigmatizante de participante.

- estabilidade temporal e topológica



	1h	2h	3h	4h	6h	12h
0h	3.66	6.42	8.20	9.30	10.67	33.76
1h	2.76					
2h	1.79	2.88	2.47	3.44	23.09	33.76
3h	1.10					
4h	0.68	1.37	4.35	21.03	37.63	66.24
5h	0.69					
6h	0.83	2.07	18.75	25.05	28.61	66.24
7h	1.24					
8h	2.28	6.80	18.95	23.60	37.63	66.24
9h	4.52					
10h	6.62	14.23	18.95	25.05	37.63	66.24
11h	7.61					
12h	6.44	12.48	18.95	25.05	37.63	66.24
13h	6.04					
14h	6.47	12.57	18.68	23.60	37.63	66.24
15h	6.10					
16h	6.22	12.58	15.88	17.59	28.61	66.24
17h	6.36					
18h	6.01	11.02	12.73	17.59	28.61	66.24
19h	5.02					
20h	4.85	9.23	12.73	17.59	28.61	66.24
21h	4.38					
22h	4.06	8.36	12.73	17.59	28.61	66.24
23h	4.30					

	PC1		PC2		PC3	
	μ	σ	μ	σ	μ	σ
cc	0.89	0.59	1.93	1.33	21.22	2.97
s	11.71	0.57	2.97	0.82	2.45	0.72
s ⁱⁿ	11.68	0.58	2.37	0.91	3.08	0.78
s ^{out}	11.49	0.61	3.63	0.79	1.61	0.88
k	11.93	0.54	2.58	0.70	0.52	0.44
k ⁱⁿ	11.93	0.52	1.19	0.88	1.41	0.71
k ^{out}	11.57	0.61	4.34	0.70	0.98	0.66
bt	11.37	0.55	2.44	0.84	1.37	0.77
asy	3.14	0.98	18.52	1.97	2.46	1.69
μ asy	3.32	0.99	18.23	2.01	2.80	1.82
σ asy	4.91	0.59	2.44	1.47	26.84	3.06
dis	2.94	0.88	18.50	1.92	3.06	1.98
μ dis	2.55	0.89	18.12	1.85	1.57	1.32
σ dis	0.57	0.33	2.74	1.63	30.61	2.66
λ	49.56	1.16	27.14	0.54	13.25	0.95

- diferenciação textual

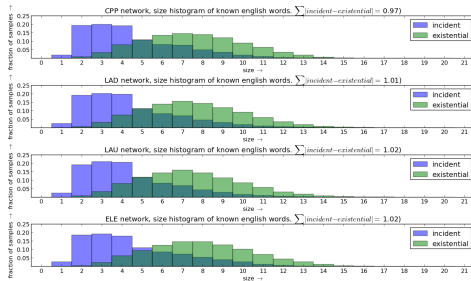
- o texto produzido por cada setor de Erdös é extremamente diferente um do outro, maiores do que a diferença entre texto produzido por redes diferentes ou mesmo por setores iguais de redes diferentes
- hubs produzem mais adjetivos. Periféricos mais substantivos, etc
- correlações não triviais
- combinação moderada de medidas topológicas e textuais; prevalência (não extrema) de componentes de texto ou topologia
- constância da existência - incidência nos textos observados

- diferenciação textual

α	0.1	0.05	0.025	0.01	0.005	0.001
$c(\alpha)$	1.22	1.36	1.48	1.63	1.73	1.95

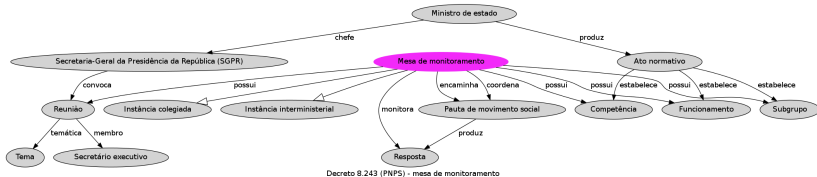
list \ measure	H-P	H-I	I-P
CPP	5.58	2.54	7.82
LAD	7.67	2.07	8.35
LAU	6.23	1.63	5.98
ELE	3.42	0.77	2.81

	CPP-LAD	CPP-LAU	CPP-ELE	LAD-LAU	LAD-ELE	LAU-ELE
P	1.35	4.05	5.80	3.00	5.41	4.94
I	1.27	0.78	4.01	0.84	3.84	3.94
H	0.98	1.94	3.17	1.32	3.82	4.47



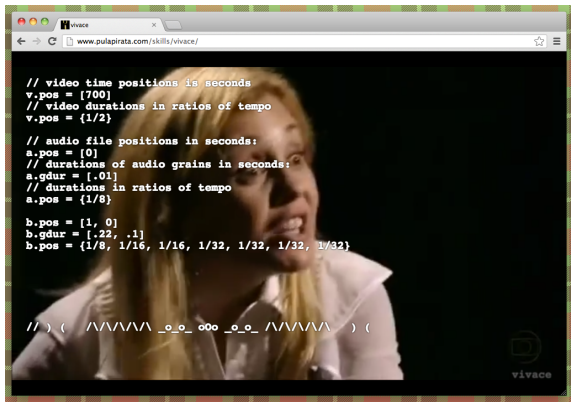
- dados sociais ligados

- síntese de ontologias (OWL) e vocabulários (SKOS) de estruturas sociais. OPS, OPa, OPP, Ontologiaa, OCD, OBS, VBS
- formalização de dados ligados a partir de dados relacionais participativos
- método de construção de ontologias orientado aos dados



- peças artísticas e mapeamentos sensoriais

- four hubs dance. Prelúdio social
- versinus (linha+senóide)
- outros casos: app online (PHP+python) para imagens do GMANE, sonificações
- apresentações artísticas: Crânio de Rilke (grupo de teoria crítica e fechamento do congresso internacional), Freakcoding



Pacotes oficiais da linguagem Python (PyPI) para compartilhamento preciso e eficiente dos desenvolvimentos. Mais especificamente:

- observação das estabilidades topológicas e diferenciações textuais. (Percolation)
- roteiros de triplificação de dados participativos brasileiros do Participabr, Cidade Democrática e AA (Participation)
- roteiros de triplificação de dados provenientes do Facebook, Twitter, e IRC (Social)
- roteiros de triplificação de dados provenientes de listas de email (Gmane)

Conclusões

- estabilidade temporal das redes de interação humana estão melhor quantizadas e são fortes: estatística da atividade ao longo do tempo, componentes principais e setores dos hubs, intermediários e periféricos
- o texto produzido por cada um dos setores é bastante diferente, com os intermediários se destacando dos setores dos hubs e periféricos
- legado em software, ontologias e dados
- possibilidades de tipologias com o aprofundamento das análises de estabilidade temporal e diferenciação textual
- 2 artigos publicados (não relacionados ao doutorado)
- 1 artigo submetido ao Physica A
- finalizar artigo de diferenciação textual
- finalizar o artigo explicitando de abertos e o método de construção de ontologias orientado aos dados

Conclusões

- tempo decorrido no projeto: 3 anos e 8 meses
- prazo máximo para defesa: Mar/2017

- [1] Renato Fabbri, *Estabilidade topológica e diferenciação textual em redes de interação humana: redes complexas para o participante e a física antropológica* (monografia de qualificação). Online em <https://github.com/ttm/tese/raw/master/ApresentacoesArtigos/quali/qualiFinal.pdf>
- [2] Renato Fabbri, *Slides da apresentação sobre estabilidade topológica e diferenciação textual em redes de interação humana: redes complexas para o participante e a física antropológica* (seminário de qualificação). Online em <https://github.com/ttm/tese/raw/master/ApresentacoesArtigos/quali/apresentacao/apresentacao.pdf>

OBRIGADO . . .