

# TECNOLOGIAS DE REDES COMPLEXAS

Renato Fabbri

II - Simpósio Internacional de Teoria Crítica

INSTITUTO DE FÍSICA DE SÃO CARLOS

UNIVERSIDADE DE SÃO PAULO

03 de Novembro, 2016



**NEXOS**  
Teoria Crítica e Pesquisa  
Interdisciplinar

## Introdução

- redes complexas e de interação humana

## Medidas básicas

## Paradigmas de redes

## Métodos para caracterização de redes de interação

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

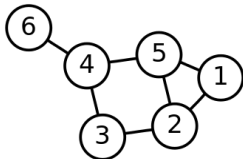
## Alguns resultados

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

## Conclusões

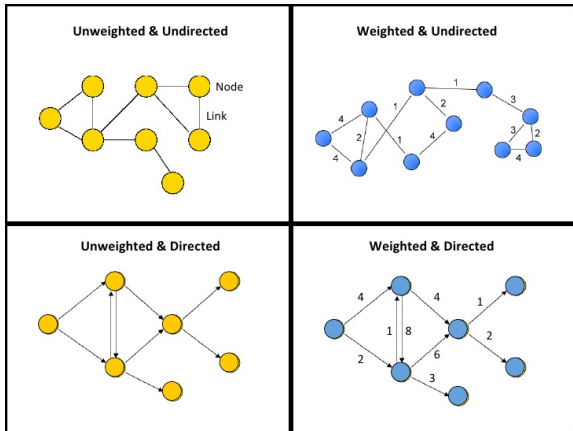
# Definições

- Grafo: estrutura matemática constituída por vértices (também nós ou pontos) e arestas (também ligações).
- Rede complexa: grafo com
  - grandes dimensões;
  - características não triviais;
  - representações de estruturas reais/naturais/empíricas.



# Definições

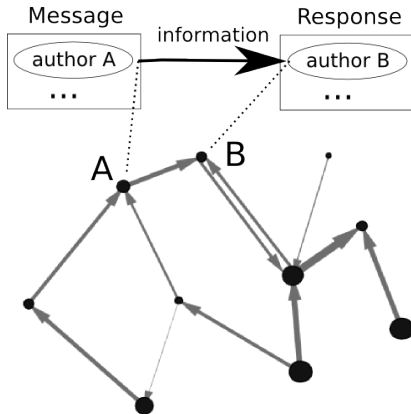
Um grafo (e uma rede) pode ser:



A nomenclatura dos elementos básicos das redes complexas varia bastante de acordo com a área (e.g. sociologia, economia, ciência da computação, física estatística, matemática).

- Vértice: nó, ponto, agente, participante, ator, etc.
- Aresta: ligação, link, bond, relação, vínculo, etc.

# Redes de interação e de amizade



# Problematização

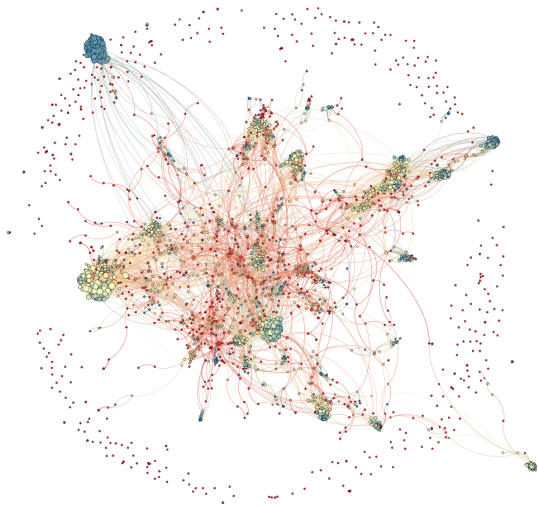
São  $10^{80}$  **átomos no universo observável**, uma referência de escala. Considere o número  $N$  de pessoas necessário para haver **mais redes possíveis do que átomos no universo**. Cada aresta é uma variável de Bernoulli fruto de cada par de vértices: a aresta pode estar presente ou não.

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

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.

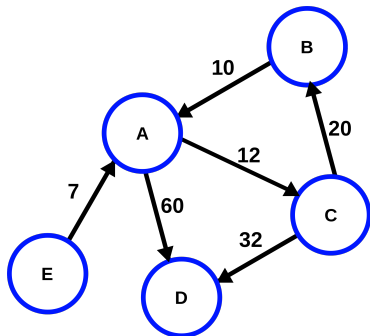
# Problematização





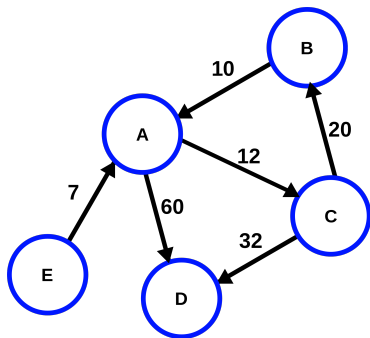
# Medidas básicas (1)

- Degree  $k_i$ : number of edges linked to vertex  $i$ .
- In-degree  $k_i^{in}$ : number of edges ending at vertex  $i$ .
- Out-degree  $k_i^{out}$ : number of edges departing from vertex  $i$ .



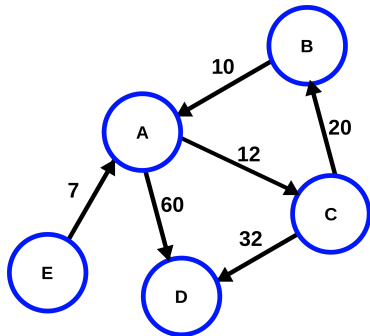
## Medidas básicas (2)

- Strength  $s_i$ : sum of weights of all edges linked to vertex  $i$ .
- In-strength  $s_i^{in}$ : sum of weights of all edges ending at vertex  $i$ .
- Out-strength  $s_i^{out}$ : sum of weights of all edges departing from vertex  $i$ .



## Medidas básicas (3)

- Betweenness centrality  $bt_i$ : fraction of geodesics that contain vertex  $i$ .
- Clustering coefficient  $cc_i$ : fraction of pairs of neighbors of  $i$  that are linked, i.e. the standard clustering coefficient metric for undirected graphs.



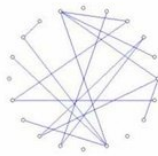
# Rede randômica de Erdős-Rényi

Clusterização baixa, grau característico ( $n \cdot p$ ), distância entre vértices é pequena.



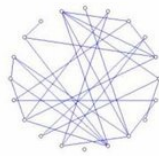
$p = 0$

(a)



$p = 0.1$

(b)

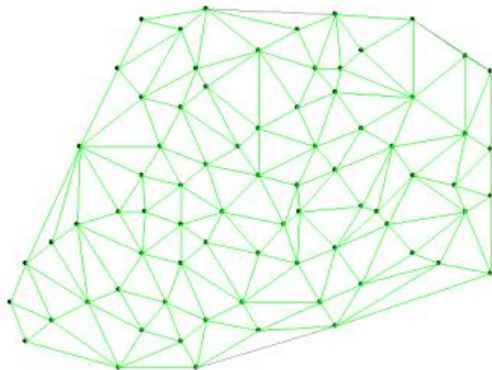


$p = 0.2$

(c)

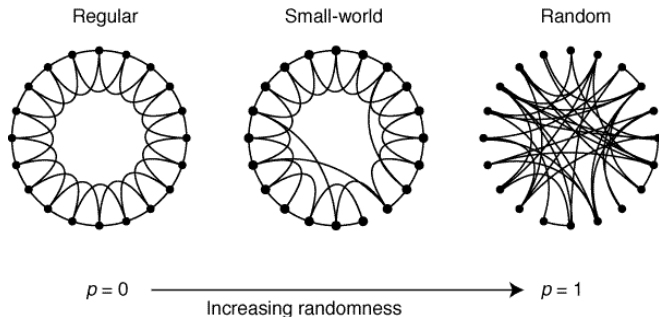
# Rede geográfica

Clusterização alta, grau característico, distância entre vértices é grande.



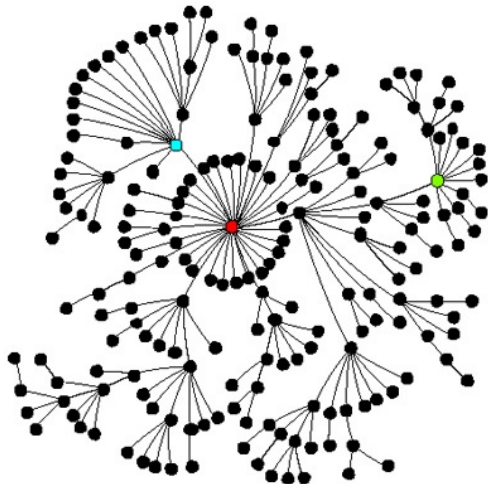
# Rede de pequeno mundo

Clusterização alta, grau característico ou não, distância entre vértices é pequena.



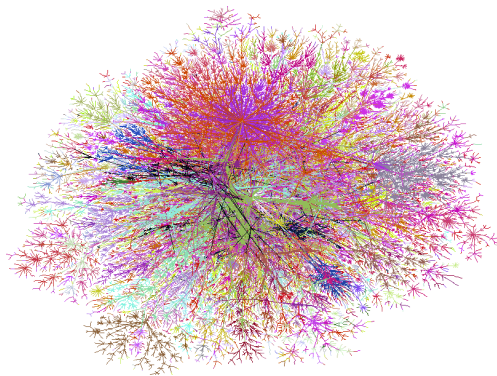
# Rede livre de escala

Clusterização alta ou não, não tem grau característico, distância entre vértices é pequena.



# Redes reais (1)

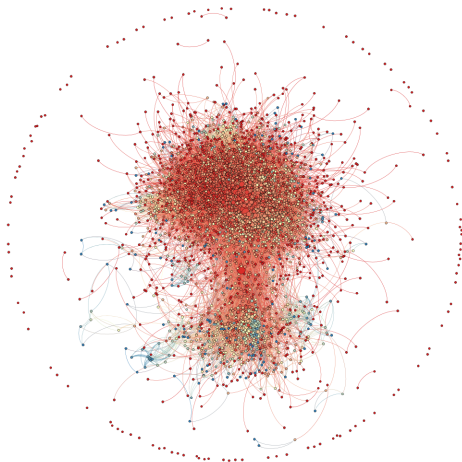
Redes reais, incluindo muitas redes sociais, econômicas e tecnológicas, tendem a ter clusterização alta, ser livre de escala e de pequeno mundo (sem conectividade/grau característico e com pequena distância entre vértices).





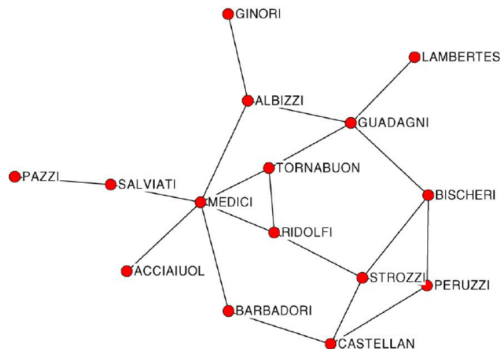
## Redes reais (2)

Redes reais, incluindo muitas redes sociais, econômicas e tecnológicas, tendem a ter clusterização alta, ser livre de escala e de pequeno mundo (sem conectividade/grau característico e com pequena distância entre vértices).



# Redes reais (3)

Padgett and Ansell, 1993



# Métodos para caracterização de redes de interação

- 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

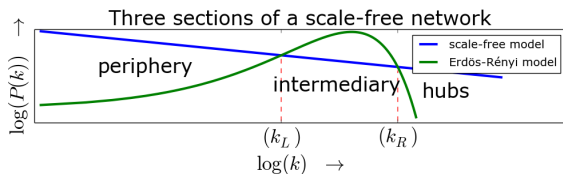
Considere  $\theta = 2\pi \frac{\text{medida}}{\text{periodo}}$ ,  $z_i = e^{i\theta}$  e  $m_n = \frac{1}{N} \sum_{i=1}^N z_i^n$  o n-ésimo momento:

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

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

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

## - setorialização de Erdős



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

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

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

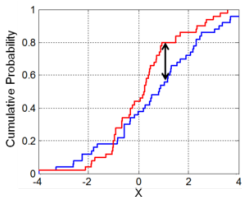
## - PCA de medidas topológicas

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

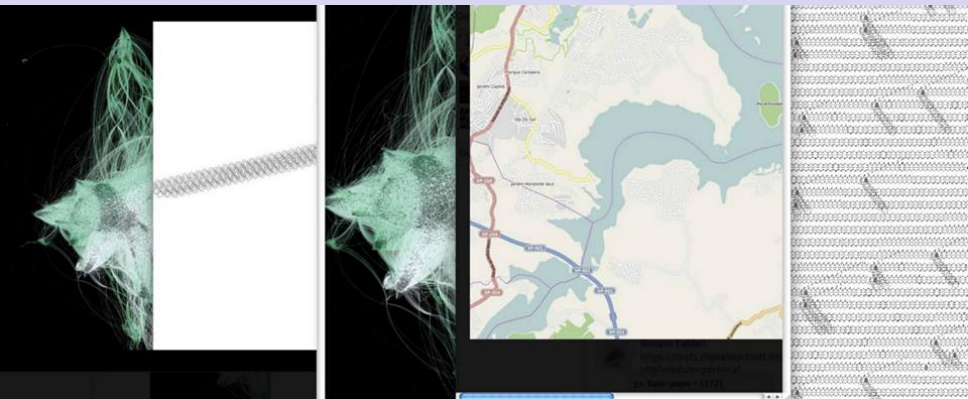


$$D_{n,n'} > c(\alpha) \sqrt{\frac{n+n'}{nn'}} \Rightarrow F_{1,n} \neq F_{2,n'} \quad (6)$$

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

$\alpha$	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

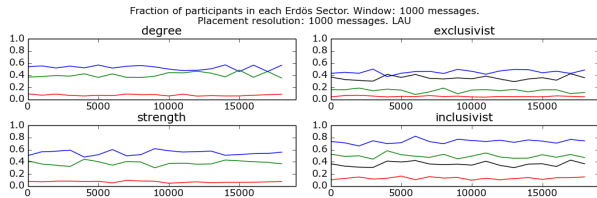
# Alguns resultados

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

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

## - diferenciação textual

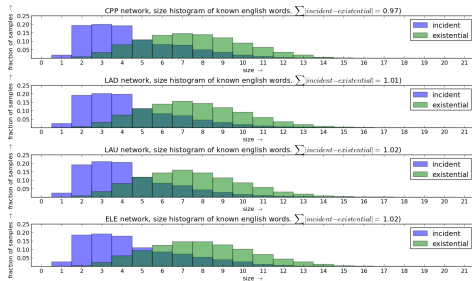
- o texto produzido por cada setor de Erdős é extremamente diferente um do outro, maiores do que a diferença entre texto produzido por redes diferentes ou mesmo por setores iguais de redes diferentes
- hubs produzem mais adjetivos. Periféricos mais substantivos, etc
- correlações não triviais
- combinação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

# - diferenciação textual

$\alpha$	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

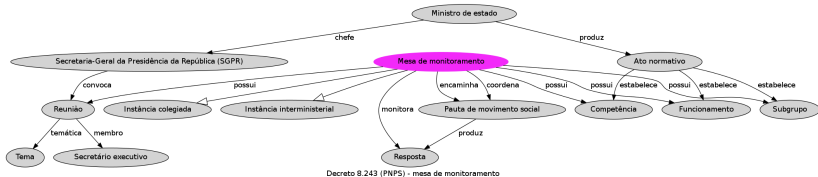
list \ measure	H-P	H-I	I-P
CPP	5.58	2.54	7.82
LAD	7.67	2.07	8.35
LAU	6.23	1.63	5.98
ELE	3.42	0.77	2.81

	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



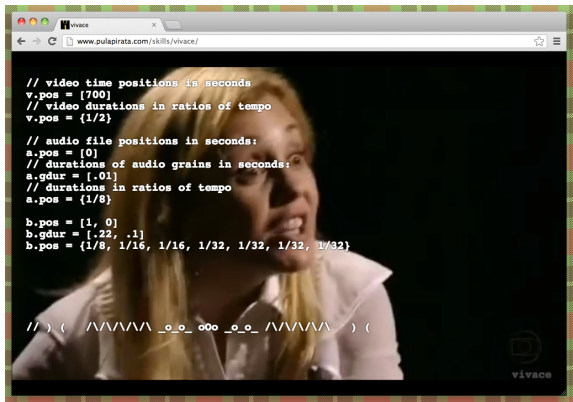
## - dados sociais ligados

- 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)

- 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

- [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 . . .