

# ESTABILIDADE TOPOLÓGICA E DIFERENCIAÇÃO TEXTUAL EM REDES DE INTERAÇÃO HUMANA: REDES COMPLEXAS PARA O PARTICIPANTE E A FÍSICA ANTROPOLÓGICA

Orientador: Prof. Dr. Osvaldo N. O. jr.  
candidato: Renato Fabbri

Seminário de Qualificação

INSTITUTO DE FÍSICA DE SÃO CARLOS

UNIVERSIDADE DE SÃO PAULO

24 de Julho, 2015



# Roteiro

## Prelúdio/Entrée/Appetizers

- complexidade (e pareidolia)
- três aspectos equânimes da distribuição livre de escala
- $f = \frac{v}{\lambda}$  está na origem da desigualdade?
- metassensores    - ideias ideais

## Introdução

- redes complexas e de interação humana
- ontologia do trabalho    - vocabulário do trabalho

Materiais: dados de email, Facebook, Twitter, Participa.br, AA, IRC

## Métodos

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

## Resultados

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

## Conclusões

# Prelúdio/Entrée/Appetizers

Digressão teórica sobre complexidade:

- contextualiza o seminário
- ajuda a inferir os níveis de rigor e abstração mais adequados
- exemplifica consequências e hipóteses
- incentiva a instrução, modelagem e especulação

Roteiro:

- complexidade (e pareidolia)
- três aspectos equânimes da distribuição livre de escala
- $f = \frac{R}{\lambda}$  na origem da desigualdade?
- metasensores
- ideias ideais

- complexidade (e pareidolia)



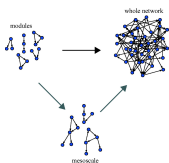
(Composition X, Kandinsky - 1939)

# - três aspectos equânimes da distribuição livre de escala

Em uma rede livre de escala, a probabilidade  $p(k_i)$  de escolher ao acaso um vértice de grau  $k_i$  é:  $p(k_i) = C.k_i^{-\alpha}$ , em geral  $1,5 < \alpha < 3$ .

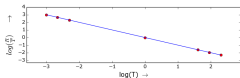
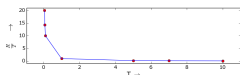


ao menos para as redes humanas, os participantes transitam continuamente entre os setores de hubs, intermediários e periféricos. Exceção para redes muito pequenas



o tempo disponível para cada participante é o mesmo, e a distribuição nas diferentes redes praticamente sempre implica no mesmo participante ser hub, intermediário e periférico em diferentes redes. Por exemplo: somos hubs em algumas redes de nossos trabalhos e de nossas famílias, periféricos em algumas redes de trabalhos e famílias de conhecidos

Relação básica da lei de potência



o meio aloca a mesma quantidade de recurso ao longo da conectividade em uma rede livre de escala. Ou seja, há uma distribuição equânime de recursos ao longo da conectividade:

$$f = \frac{v}{\lambda} \Rightarrow \log(f) = -1\log(\lambda) + \log(v) \Rightarrow \alpha = 1.$$

Ao mesmo tempo,  $f.\lambda = v = \text{constante}$ .

$f = \frac{v}{\lambda}$  está na origem da desigualdade?

$$f = \frac{v}{\lambda_1 \lambda_2}, \quad \lambda \approx \lambda_1 \approx \lambda_2 \Rightarrow \log(f) = -2\log(\lambda) + \log(v) \quad (1)$$

Em uma rede, temos essencialmente  $E$  arestas e  $N$  vértices. Assumindo linearidade:

$$\text{Recursos} = \alpha N + \beta E \approx \beta E = \beta N \frac{E}{N} = \beta \frac{N \bar{k}}{2} \quad (2)$$

onde  $\bar{k}$  é o grau médio. O termo  $\alpha N$  pode ser descartado se assumirmos interesse em rede de interação, onde  $N$  participantes que não interagem pode ser considerado o estado de mínimo/nulo emprego de recursos pelo sistema (para a interação de interesse e observável).

$$f = \frac{v}{\text{Recurso}_i} = \frac{v}{\beta E_i} = \frac{2}{\beta} \frac{v}{N_i \cdot \bar{k}_i} \equiv \frac{v}{\lambda_1 \lambda_2} \quad (3)$$

$N$  e  $\bar{k}$  são sempre diretamente proporcionais à quantidade de recursos alocados se fixado um deles ( $N$  ou  $\bar{k}$ ).

## - metassensores

Cada vértice subsiste no tempo, é um observador, um sensor, uma unidade de processamento de informação e matéria. As arestas correspondem ao relacionamento entre os dois sensores, criando um sensor composto, dual. A associação sucessiva de sensores implica em um meta-sensor. Considerações:

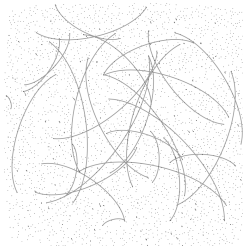
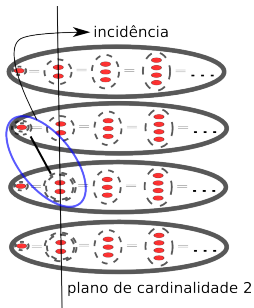
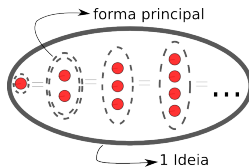
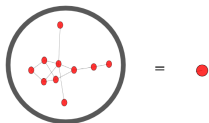
- a teoria básica de complexidade estabelece: estes sistemas são adaptativos para eficiência e sobrevivência, processam informação, interagem entre si, por vezes se replicam (e.g. idosos em uma praça)
- as arestas são criadas através da alocação de tempo pelos participantes. O recurso básico é tempo. A rede é um sensor de sinal temporal (quadrático?)
- a percepção também lineariza a lei de potência:  $\Delta_8 = 2^{\frac{f_1 - f_0}{f_0}}$
- o universo é o sistema complexo total, i.e. sensor máximo?
- densidade média no caso tridimensional:  $f = \frac{v}{\lambda_1 \lambda_2 \lambda_3} \equiv \frac{v}{r^3}$

Exemplo do sensor sobre a Unicamp (geração de conhecimento, ganha amigos, concentração segundo lei de potência, motor dialético, sensor colorido).

## - ideias ideais

Teoria física das ideias.

Uma ideia ideal é um objeto físico idealizado, como uma superfície ideal.





# Introdução

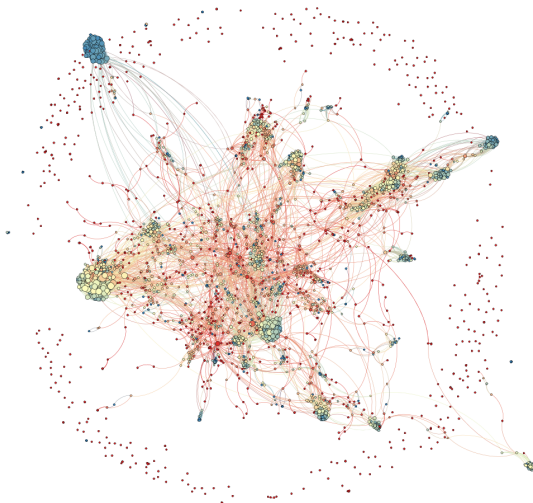
São  $10^{80}$  **átomos no universo observável**, uma referência de escala. Considere o número  $N$  de pessoas necessário para haver **mais redes possíveis do que átomos no universo**. Cada aresta é uma variável de Bernoulli fruto de cada par de vértices: a aresta pode estar presente ou não.

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

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**.

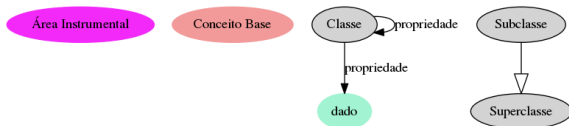
**Sistema complexo**  $\Rightarrow$  Constituído de várias partes cuja interação exibe comportamento emergente. É usual considerar que um sistema complexo: processa informação, exibe mecanismos adaptativos, pode exibir mecanismos de reprodução. Um sistema complexo é integrado a outros sistemas complexos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

# Introdução



## - ontologia do trabalho

Formalização dada ontologia OWL com a formalização das áreas envolvidas.



Legenda da OT

Situar alguns indivíduos na ontologia, que são alguns dos feitos.  
Contemplar de redes de interação humana em evolução temporal até  
Redes complexas, estatística e ignorância?

## - vocabulário do trabalho

Vocabulário SKOS, compreendendo principais definições, assim como os casos de ambiguidades, sinônimos e formas preferidas. Complementa as definições e conceitualização da OT. Termos especiais:

- física antropológica
- estabilidade temporal
- setorialização de Erdős, setores de Erdős
- rede complexa
- contribuição tecnológica, científica, para o ensino, para a extensão universitária

## - objetivos e justificativas

O objetivo geral é comprovar e aprofundar a utilização das redes complexas pelo participante. 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, de fontes heterodoxas e apropriadas para a análise conjunta
- delineio da prática de estudo das próprias estruturas sociais do pesquisador

A justificativa principal é que há um hiato proeminente entre o potencial desta frente científica e tecnológica e a utilização que os participantes destas redes fazem dela. Justificativas secundárias são:

- a área é recente e reconhecidamente útil e potente
- as instituições já fazem uso destes conhecimentos há séculos (talvez milênios). A frente civil está começando a surgir
- 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
- Participa.br: redes de amizade e de interação, texto de postagens, comentários, etc
- Twitter: milhões de tweets permitiram observação contínua de redes de interação (retweet), relacionamentos por vocabulário e hashtags, e padrões do vocabulário em si
- materiais coletados com entrevistas e oficinas com especialistas
- estruturas semânticas e dados etiquetados
- estruturas sociais do qual faço parte

- estatística circular
- obtenção das redes de interação
- setorialização de Erdős
- PCA de medidas topológicas
- testes de Kolmogorov-Smirnov dos textos
- web semântica de dados ligados
- 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:

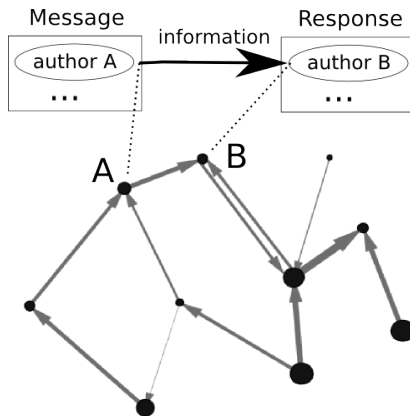
$$\begin{aligned} R_n &= |m_n| \\ \theta_\mu &= \text{Arg}(m_1) \\ \theta'_\mu &= \frac{\text{period}}{2\pi} \theta_\mu \end{aligned} \tag{4}$$

$$\begin{aligned} \text{Var}(z) &= 1 - R_1 \\ S(z) &= \sqrt{-2 \ln(R_1)} \\ \delta(z) &= \frac{1 - R_2}{2R_1^2} \end{aligned} \tag{5}$$

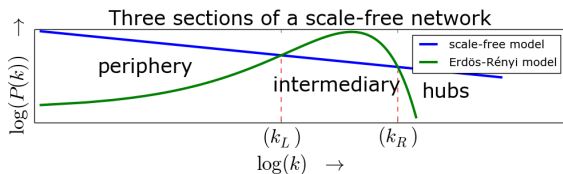
Usamos também  $\frac{b_h}{b_l}$  entre a maior  $b_h$  e a menor  $b_l$  incidência nos histogramas.



- redes de interação



## - setorialização de Erdős



$$\sum_{x=k_i}^{k_j} \tilde{P}(x) < \sum_{x=k_i}^{k_j} P(x) \Rightarrow i \text{ é intermediário} \quad (6)$$

$$P(k) = \binom{2(N-1)}{k} p_e^k (1-p_e)^{2(N-1)-k} \quad (7)$$

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

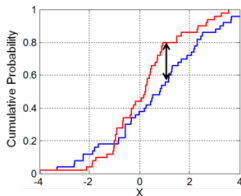
## - PCA de medidas topológicas

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

$$\begin{aligned}\mu_{V'}[j, k] &= \frac{\sum_l^L V'[j, k, l]}{L} \\ \sigma_{V'}[j, k] &= \sqrt{\frac{(\mu_{V'} - V'[j, k, l])^2}{L}} \\ \mu_{D'}[k] &= \frac{\sum_l^L D'[k, l]}{L} \\ \sigma_{D'}[k] &= \sqrt{\frac{(\mu_{D'} - D'[k, l])^2}{L}}\end{aligned}\tag{8}$$

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

- teste de Kolmogorov-Smirnov para incidencias em textos



$$D_{n,n'} > c(\alpha) \sqrt{\frac{n+n'}{nn'}} \Rightarrow F_{1,n} \neq F_{2,n'} \quad (9)$$

$$c(\alpha) < \frac{D_{n,n'}}{\sqrt{\frac{n+n'}{nn'}}} = c'(\alpha) \quad (10)$$

$\alpha$	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

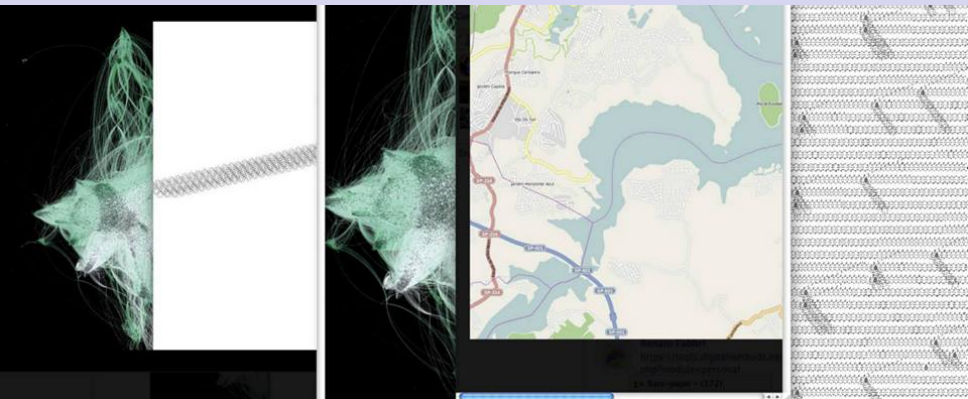
## - web semântica / dados ligados

Recomendação da W3C para a formalização de conceitualizações, o relacionamento dos dados a estas conceitualizações, o armazenamento de dados semanticamente enriquecidos, o relacionamento de dados de fontes diferentes, a navegação semântica e a inferência por máquina.

Observações especiais:

- permite a formalização de redes relativamente estáveis em nosso tecido social (em certas escalas temporais e de população).
- permite a análise conjunta de dados de diferentes fontes; desenvolvimento conceitual compartilhado
- padrão acadêmico para dados semânticos etiquetados; melhor formato para entregar os dados para a sociedade como um legado para análise e experimentos
- expressão em triplas "sujeito objeto predicado"
- pesado e um pouco complicado. Uso de ferramentas como Fuseki/Jena para facilitar os usos
- procurar uma notação mais poderosa para os dados ligados?

## - 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

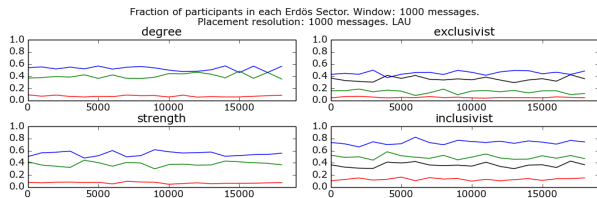
- estabilidade temporal
- diferenciação textual
- iniciação da nuvem brasileira de dados ligados participativos
- aparato em software
- beneficiamento
- ideias ideais (teoria física das ideias)



## - 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 achei formalização para esta expectativa e talvez esta seja a primeira
- estabilidade das componentes principais. Prevalência da centralidade, seguida da simetria e então clusterização dos participantes
- tipologia não estigmatizante de participante. Tipologia de rede

# - 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	18.68	23.60	37.63	66.24
17h	6.36					
18h	6.01	11.02	15.88	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	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
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 <sup>in</sup>	11.68	0.58	2.37	0.91	3.08	0.78
s <sup>out</sup>	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 <sup>in</sup>	11.93	0.52	1.19	0.88	1.41	0.71
k <sup>out</sup>	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
$\mu$ 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
dis	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

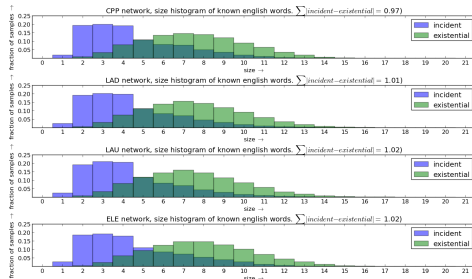
- o texto produzido por cada setor de Erdös é extremamente diferente um do outro, maiores do que a diferença entre texto produzido por redes diferentes ou mesmo por setores iguais de redes diferentes
- hubs produzem mais adjetivos. Periféricos mais substantivos, etc
- correlações não triviais
- combinação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

# - diferenciação textual

$\alpha$	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

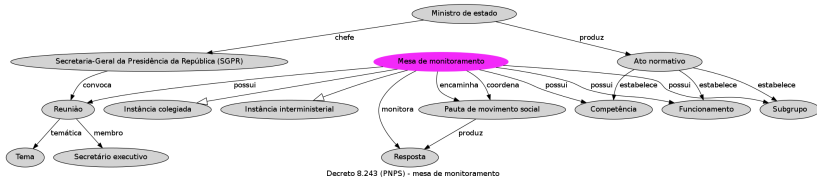
list \ measure	H-P	H-I	I-P
CPP	5.58	2.54	7.82
LAD	7.67	2.07	8.35
LAU	6.23	1.63	5.98
ELE	3.42	0.77	2.81

	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



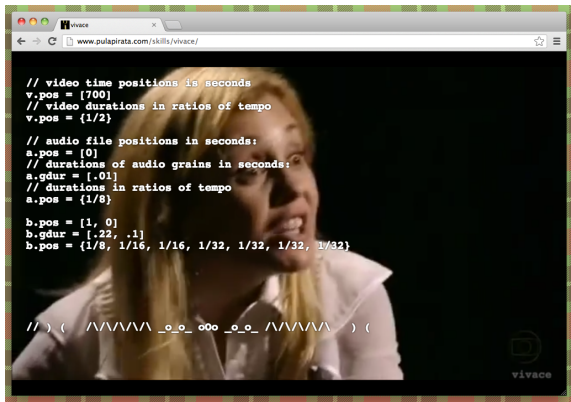
## - dados sociais ligados

- 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



## - software

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

- observação das estabilidades topológicas e diferenciações textuais. (Gmane)
- acesso à nuvem de dados participativos brasileiros, junto aos scripts de triplificação e análise (Participation)
- anotação automatizada e semântica de seus próprios dados virtuais provenientes do Facebook, Twitter, Diáspora, IRC, etc (Social)
- mapeamentos precisos de estruturas sociais em sonoras através de síntese (MASS)
- integração destes dados todos para análise conjunta e navegação semântica enriquecida. Realização de músicas e animações abstratas explorando estas propriedades-chave (Percolation)
- **streaming de estruturas sociais**
- **sistema de monitoramento massivo e aberto**

## - beneficiamento

- recomendação de recursos para enriquecimento da navegação semântica.
- experimentos percolatórios
- geração de arte a partir das estruturas sociais
- formalização de estruturas de governança (gerenciais) da sociedade
- sistema de construção de ontologias OWL orientado aos dados
- enriquecimento semântico de dados relacionais e sua tradução para dados ligados
- compreensão sobre as estruturas sociais: estabilidades, diferenciações
- fundamentação da origem da lei de potência nestes contextos
- contribuição ao cânone de tipologias humanas e sociais, que é humanístico, através da física



## - empréstimos antropológicos

- histórico do termo física antropológica, de Boaz a este trabalho
- aspectos reflexivos, biográficos. Estudo e exposição de si. Diário. Leitura da curtíssima autobiografia
- comparação entre física e antropologia. Considerações sobre uma ciência sólida
- *Social Physics* do Pentland (co-fundador e diretor do MIT Media Lab)

## - ideias ideais

- o que é
- como é formada a rede de uma ideia
- pelo mesmo raciocínio da origem da propriedade livre de escala, podem ser considerados sensores abstratos
- **ideias ideais como unificação das redes sociais formadas por indivíduos e por conceitos.** Sensores imateriais de informação?
- **ontologias são Rich Clubs?**
- exemplos de 2002. Discussão de algum caso partindo do sistema axiomático

- gradus ad parnassum unificando:
  - apresentação breve e instrumental da área para o indivíduo. Conceitos fundamentais. Paradigmas de redes. Sinonímias, e ambiguidades. Caracterização de redes humanas e beneficiamento para o indivíduo através de experimentos antropológicos, análise e navegação
  - consideração da física antropológica (ou como esta conceituação se resolver)
  - Apêndice com listagens úteis, como medidas, software, trabalhos de referência, protocolos, dados, etc.
- tipologias com o aprofundamento das análises de estabilidade temporal e diferenciação textual
- legado em software, ontologias e dados
- sugestão atual: finalizar artigo de diferenciação textual, escrever o gradus e um artigo explicitando o método de construção de ontologias orientado aos dados

- [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 . . .