

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

Exame de Qualificação
INSTITUTO DE FÍSICA DE SÃO CARLOS
UNIVERSIDADE DE SÃO PAULO

24 de Julho, 2015

Roteiro

Prelúdio/Entrée/Appetizers

- complexidade, pareidolia, perversão e magia do caos
- 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-Smirnoff 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

- próximos passos - ideais ideias - agradecimentos e referências

Prelúdio/Entrée/Appetizers

Digressão teórica:

- contextualiza o seminário
- sintoniza rigor e abstração
- exemplifica consequências e hipóteses
- incentiva a instrução, modelagem e especulação

Roteiro:

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

- complexidade, pareidolia, perversão e magia do caos

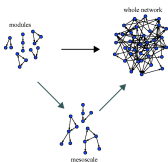


(Composition X, Kandinsky - 1939)

- três aspectos equânimes da distribuição livre de escala

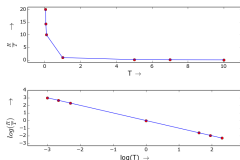


ao menos para as redes humanas, os participantes transitam continuamente entre os setores de hubs, intermediários e periféricos. Exceção para redes muito pequenas



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, $f \cdot \lambda = v = \text{constante}$.

$f = \frac{v}{\lambda}$ 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 αN 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).

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

N e \bar{k} são sempre diretamente proporcionais à quantidade de recursos alocados se fixados 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:

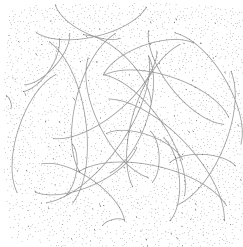
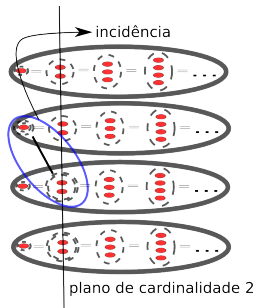
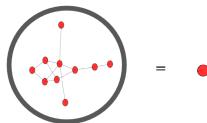
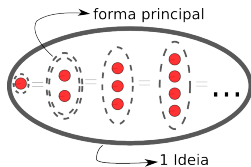
- teoria básica de complexidade sugere isso: sistemas adaptativos para eficiência e sobrevivência, processam informação, interagem entre si, por vezes se replicam (e.g. idosos em uma praça).
- densidade média de esferas no caso tridimensional:
$$f = \frac{v}{\lambda_1 \lambda_2 \lambda_3} \equiv \frac{3}{4\pi} \frac{v}{r^3}$$
- extensão da percepção também segue a lei de potência: $\Delta_8 = 2$
- 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?
- o universo é o sensor máximo?

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.

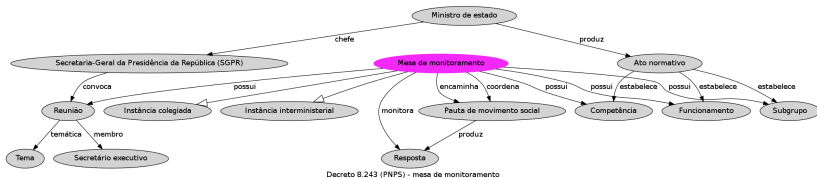
$$\begin{aligned}2^{\binom{N}{2}} > 10^{80} &\Rightarrow \log_2[2^{\binom{N}{2}}] > \log_2(10^{80}) \Rightarrow \binom{N}{2} > \frac{\log_{10}(10^{80})}{\log_{10}2} \Rightarrow \\&\Rightarrow \frac{N.(N-2)}{2} > \frac{80}{\log_{10}2} \Rightarrow N > 23,5988\end{aligned}$$

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

Sistemas complexos \Rightarrow sistemas que processam informação, exibem mecanismos adaptativos e de subsistência. Constituído de várias partes cuja interação implica em comportamento emergente. Sistema intrínseco ao meio em que subsiste e integrado a outros sistemas dinâmicos.

- ontologia do trabalho

diagrama da ontologia OWL com a formalização das áreas envolvidas. 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 polissemias e poligrafias.

Materiais

- Mensagens de e-mail, com horário de envio, ID da mensagem, ID da mensagem anterior na thread se existente, texto do título e corpo.
- Redes de Facebook: redes 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.

- Estatística circular
- (obtenção das redes de interação)
- Setorialização de Erdős
- PCA de medidas topológicas
- Testes de Kolmogorov-Smirnoff dos textos
- Web semântica
- Audiovisualização de dados
- Considerações tipológicas e humanísticas

- estatística circular

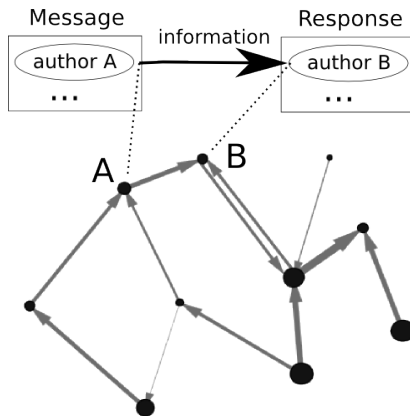
Com $m_n = \frac{1}{N} \sum_{i=1}^N z_i^n$ o n-ésimo momento:

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

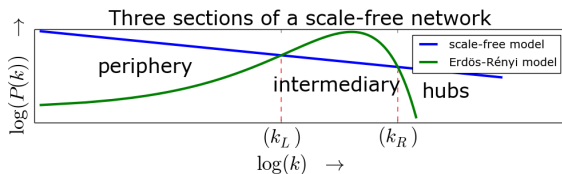
$$\begin{aligned} \text{Var}(z) &= 1 - R_1 \\ S(z) &= \sqrt{-2 \ln(R_1)} \\ \delta(z) &= \frac{1 - R_2}{2R_1^2} \end{aligned} \tag{5}$$

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.

- Kolmogorov-Smirnoff para textos

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

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

- web semântica / dados ligados

[!Procurar uma notação mais poderosa para os dados ligados.]

- Formalização de conceitualizações.
- Redes estáveis em nosso tecido social em certas escalas temporais e de pessoas. Passível de transições de fase, modificações abruptas em outras.
- Permite: análise conjunta de dados de diferentes fontes; inferência por máquina; desenvolvimento conceitual compartilhado.
- Recomendação da W3C; padrão acadêmico para dados semânticos etiquetados; melhor formato para entregar os dados para a sociedade como um legado para análise e experimentos.
- Pesado e um pouco complicado. Uso de ferramentas como Fuseki/Jena para facilitar os usos.

- audiovisualização de dados

- Versinus, imagens, animação abstrata com música, sonificações.
- Roteiros automatizados de realização de arte social.
- 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 Erdös, 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.

- diferenciação textual

- O texto produzido por cada setor de Erdös é extremamente diferente um do outro, maiores do que texto produzido por redes diferentes ou mesmo por setores iguais de redes diferentes.
- Hubs produzem mais adjetivos. Periféricos mais substantivos, etc.
- Formalização de dados ligados a partir de dados relacionais participativos.
- 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.

- 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.
- Outros casos: app online (PHP+python) para imagens do GMANE, sonificações.
- Apresentações artísticas: Crânio de Rilke, Freakcoding.

Pacote oficial da linguagem Python (PyPI) para:

- observação das estabilidades topológicas e diferenciações textuais. (Gmane)
- Acesso à nuvem de dados participativos brasileiros, junto aos scripts de triplificação. para 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 da melhor qualidade 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)

- beneficiamento

- Recomendação de recursos para enriquecimento da navegação semântica.
- Experimentos percolatórios.
- Fundamentação da origem da lei de potência nestes contextos.
- Enriquecimento das tipologias humanas e sociais através do cânone acadêmico para o assunto, que é humanístico.

- 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.
- Livre de escala na forma principal, possui diversos harmônicos.
- 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.
- Exemplos de 2002.

Conclusões

- Gradus 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. Apêndice com listagens úteis, como medidas, trabalhos de referência, etc.
 - Questões de física antropológica (ou como esta conceituação se resolver).
 - Aparato em software, protocolos e dados.
- Tipologias com o maior desenvolvimento das análises de estabilidade temporal e diferenciação textual
- Legado em software, ontologias e dados
- Estou aquém de equipe e qualidade de vida do que acredito adequado.
- Ignorância. Froteira da ilha sempre se ampliando.
- Gostarei de terminar a graduação da física aos poucos se eu tiver a oportunidade. Creio que causará bastante estímulo e isso pode ser conveniente daqui alguns anos. Talvez proficiências para titulação.

- próximos passos

Meu norte seria já poder estudar diretamente a cadeia de sistemas dinâmicos, do social ao universo. Com sua inércia, deve dar para tirar umas boas propriedades relacionando as frequências de sistemas complexos em escalas diferentes. As referências de escala individual ao universo observável devem implicar em propriedades.

Ou seja:

- muita coisa boa para fazer, seria bom delimitar o que vai para pós-doc ou próximo vínculo.
- Isso parece depender de alguma previsão de vínculo apropriado e de onde.
- Prioridade é permanecer em S. Carlos, de preferência relacionado ao IFSC. Possibilidades na UFSCar, IPRJ, UnB, UFABC.
- Acho que é na academia que conseguirei me dedicar para a pesquisa podendo imergir como gosto.

- [1] Anis Das Sharma, Alpa Jain, Kong Yu, "Dynamic Relationship and Event Discovery".
- [2] Nguyen Bach and Sameer Badaskar, Presentation on "Survey on Relation Extraction".

OBRIGADO . . .