

## Redes Complexas: O Próximo Passo?

M. Cristina F. de Oliveira & Roberto Pinho

Workshop de Mapeamento Visual de Coleções de  
Documentos

ICMC-USP 2005

## Roteiro

- Introdução
- Modelos de redes
- Métricas
- Ferramentas
- Redes complexas e Texto
- Redes complexas e Coleções de Documentos
- Conexões

setembro 2005

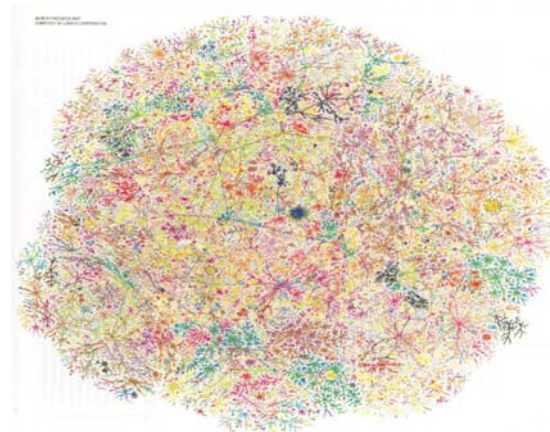
2

## Introdução às Redes Complexas

- Internet
- Cérebro
- Reações entre proteínas
- Relações sociais
- Sistemas rodoviários...

setembro 2005

3



## Modelos de Redes complexas

- Random
- Scale-free
- Hierarchical

setembro 2005

5

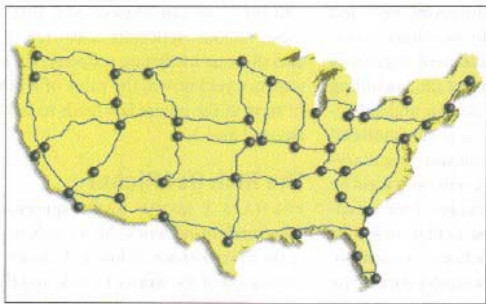
## Modelo aleatório (Random)

- Conexões: maior parte dos nós possui o mesmo grau
- Construção: igual probabilidade de conexão para novo nó
- Exemplo: Rodovias

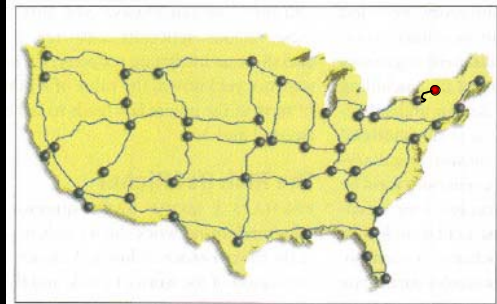
setembro 2005

6

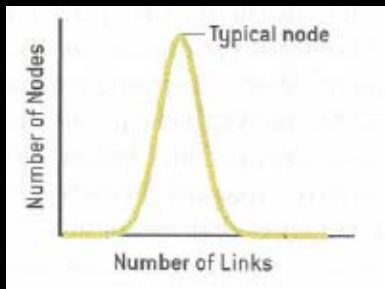
## Modelo aleatório - exemplo



## Modelo aleatório - exemplo



## Modelo aleatório - distribuição



setembro 2005

9

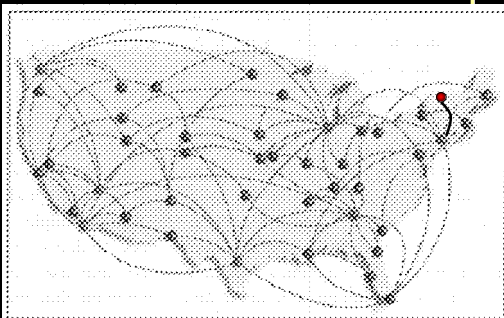
## Modelo livre de escala (scale-free)

- Conexões: alguns nós (*hubs*) possuem grande número de conexões
- Construção: Maior probabilidade de conexão a *hubs*. *preferential attachment*
- Exemplo: ligações entre aeroportos

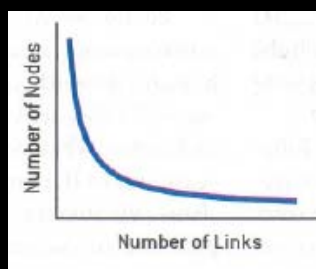
setembro 2005

10

## Modelo livre de escala - exemplo



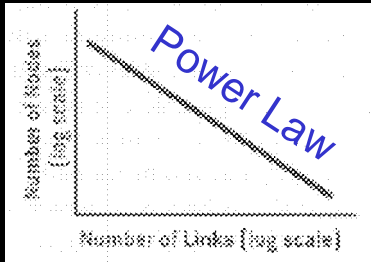
## Modelo livre de escala - distribuição



setembro 2005

12

## Modelo livre de escala - distribuição



setembro 2005

13

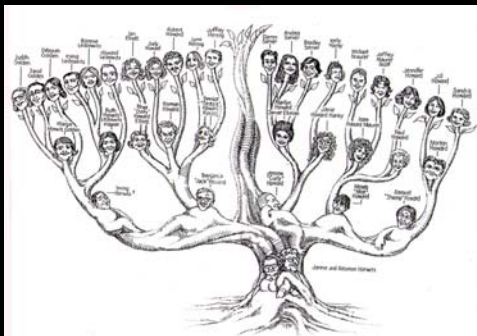
## Modelo hierárquico

- Conexões: nós com poucas conexões em *clusters*, ligados por *hubs*
- Construção: estrutura inerente
- Exemplo: relações familiares
- É também *scale-free*

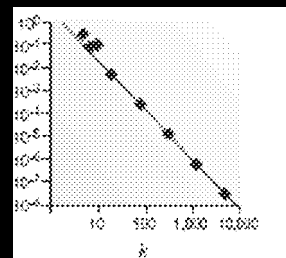
setembro 2005

14

## Modelo hierárquico - exemplo



## Modelo hierárquico - distribuição



setembro 2005

16

## Em números...

- $k$  - grau de um nó
- $P(k)$  - distribuição de probabilidade para valores de  $k$
- $\langle k \rangle$  - grau médio
- $C_I$  - coeficiente de agrupamento
- $\langle l \rangle$  - média menores caminhos

Μαθηματικά ...

setembro 2005

17

## Coeficiente de agrupamento

$$C_I = 2n_I / k(k-1)$$

- $n$  - número de conexões entre os  $k$  vizinhos de  $I$
- *Máximo de conexões:*  $k_1(k_1-1)/2$
- $\langle C \rangle$  - coeficiente médio
- $C(k)$  - coeficiente médio dos nós de grau  $k$

setembro 2005

18

## Menor caminho

$l_{AB}$  distância de A a B

$$l_{AB} \neq l_{BA}$$

$\langle l \rangle$  - média menor caminho

setembro 2005

19

## Comportamento dos modelos

- Aleatório
  - $P(k)$  é uma distribuição de Poisson
  - $C(k)$  é constante, agrupamento independe do número de conexões
  - $l \sim \log N$  "Mundo pequeno"

setembro 2005

20

## Comportamento dos modelos

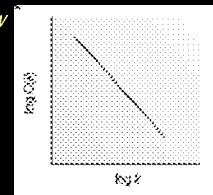
- Livre de escala
  - $P(k) \sim k^{-\gamma}$  *power law*
  - $2 < \gamma < 3$
  - Preferential attachment:  $\Pi_i = k_i / \sum k_j$
  - $C(k)$  é constante, agrupamento independe do número de conexões
  - $l \sim \log \log N$  "Mundo ultra pequeno"

setembro 2005

21

## Comportamento dos modelos

- Hierárquico
  - $P(k) \sim k^{-\gamma}$  *power law*
  - $2 < \gamma < 3$
  - $C(k) \sim K^{-1}$



setembro 2005

22

## Comportamento dos modelos

- Aleatório
  - $P(k)$  é uma distribuição de Poisson
  - $C(k)$  é constante, agrupamento independe do número de conexões
  - "Mundo pequeno"

setembro 2005

23

## Comportamento dos modelos

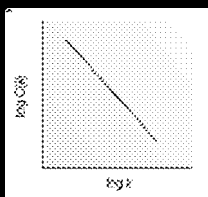
- Livre de escala
  - $P(k)$ : *power law*
  - *Preferential attachment*
  - $C(k)$  é constante, agrupamento independe do número de conexões
  - "Mundo ultra pequeno"

setembro 2005

24

## Comportamento dos modelos

- Hierárquico
  - $P(k)$  : *power law*
  - $C(k)$  : *power law*



setembro 2005

## Importante !

- $P(k)$  e  $C(k)$  independem do número de Nós
- Podem ser usados para caracterizar uma rede.

setembro 2005

26



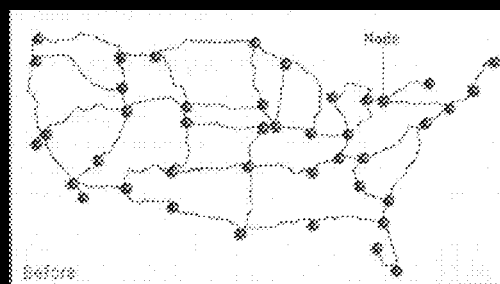
## A Bahia é um ovo (São Carlos também)

- *Small World Problem*
- Dois nós podem ser conectados em alguns (poucos) passos
- Redes sociais: pessoas bem conhecidas se conhecem
- Stanley Milgram, 1967 : seis graus de separação

setembro 2005

27

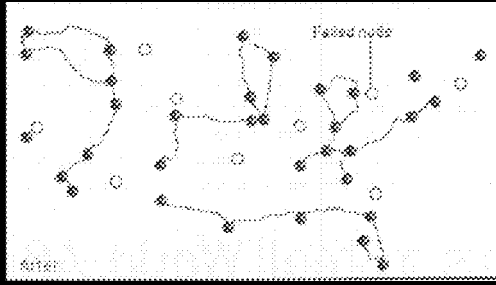
## Robustez - redes aleatórias



setembro 2005

28

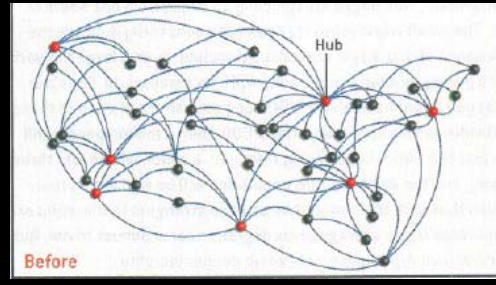
## Robustez - redes aleatórias



setembro 2005

29

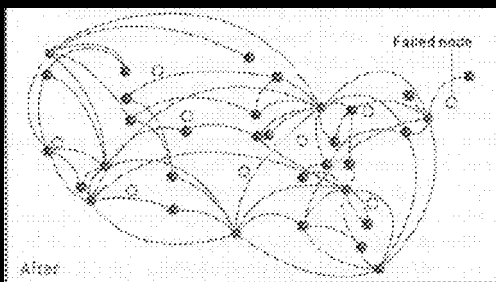
## Robustez - redes scale free



setembro 2005

30

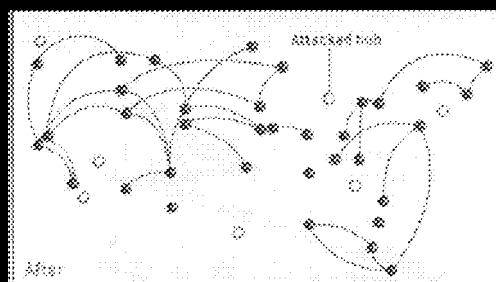
## Robustez - redes scale free



setembro 2005

31

## Robustez - redes scale free



setembro 2005

32

## Ferramentas

- Diversas para análise e visualização de grafos e redes em geral
  - JUNG (Java Universal Network/Graph Framework): linguagem para modelagem, análise e visualização de dados representados como grafos ou redes
  - Graphviz, TouchGraph: layout e desenho de grafos
  - Pajek, Otter: análise e visualização de redes

setembro 2005

33

## Ferramentas

- Muitas voltadas especificamente para análise e visualização de redes sociais
  - Aqna
  - infinder: softw. para detecção de '*motifs*' em redes
  - NetMiner
- Sites com *links* para ferramentas e programas
  - <http://cran.r-project.org/src/contrib/PACKAGES.html#sna>
  - [http://www.insna.org/INSNA/soft\\_inf.html](http://www.insna.org/INSNA/soft_inf.html)
  - <http://cyvision.if.sc.usp.br/~francisco/networks/software.html>

setembro 2005

34

## Aplicação em Modelagem de Textos (NILC)

- Trabalho no TIL 2005: Lucas Antiquiera, Graça Nunes, Osvaldo Novais e Luciano Costa
- Objetivo: avaliar qualidade de textos adotando uma abordagem não lingüística
- Extração de medidas estatísticas de textos modelados como redes complexas

setembro 2005

35

## Textos como Redes Complexas

- Pré-processamento
  - Remoção de *stopwords*
  - Lematização das palavras restantes

setembro 2005

36

↓ Remoção de stopwords

- Joana criou uma obra magnífica. **Concebeu** a idéia de sua escultura num domingo à tarde, ao passear pelo jardim. Nessa época, sua mente estava povoada por um único pensamento: criar a obra de sua vida.

↓ Lematização

- Joana **criar** uma obra **magnífico**. **Conceber** a idéia de sua escultura num domingo à tarde, ao passear pelo jardim. Nessa época, sua mente **estar** **povoado** por um único pensamento: criar a obra de sua vida.

setembro 2005 37

## Textos como Redes Complexas

- Rede
  - Cada palavra restante representa um nó
  - Cada par de palavras consecutivas é associado a uma aresta direcionada
  - Cada aresta recebe um peso que indica quantas vezes a respectiva associação de palavras ocorre no texto

setembro 2005 38

## Textos como Redes Complexas

setembro 2005 39

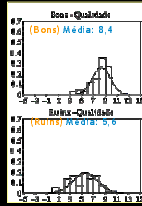
## Experimento

- Dois grupos de textos
  - 10 textos do gênero informativo, produzidos por universitários experientes (textos "bons")
  - 10 redações de vestibular (textos "ruins")
- Juízes: 6 alunos de curso de Letras
  - Notas de 0 a 10, para o critério "qualidade"
  - Não foi fornecida qualquer definição de qualidade aos juízes

setembro 2005 40

## Experimento

- Juízes diferenciaram os dois grupos
  - Diferença nas médias comprovada pelo teste *t-student*
- Textos modelados como redes, e métricas extraídas
  - Variabilidade permite analisar se existe relação entre as medidas provenientes das redes e as notas dadas pelos juízes



setembro 2005

41

## Métricas Extraídas

- Grau médio de entrada e grau médio de saída
- Caminho mínimo médio entre todos os pares de nós
- Coeficiente de aglomeração (*clustering*) médio
- Dinâmica do número de componentes
  - Variação do número de componentes na medida em que novas arestas vão sendo inseridas: componentes fracamente conexos

setembro 2005

42

## Resultados

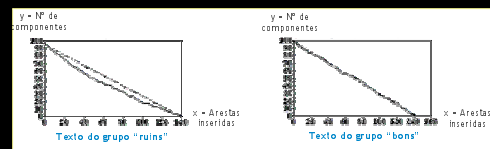
- Medidas permitem diferenciar entre textos 'bons' e 'ruins'!
  - Qualidade tende a cair na medida em que os valores do grau médio aumentam
  - Qualidade diminui com o aumento do coeficiente de aglomeração médio
  - Textos com notas maiores apresentam caminhos mínimos também maiores
- Várias implicações para processamento automático de língua, caracterização de textos de corpus, etc.

setembro 2005

43

## Resultados

### Dinâmica do número de componentes



setembro 2005

44

## Mapeamento de Coleções de Documentos

- Scientometrics (Derek Price, 1922-1983)
  - Estudo quantitativo de comunicações científicas (bibliometria aplicada à literatura científica)
  - Redes de citações e co-citações
- Relações entre artigos científicos
  - Mapas de coleções de documentos
    - Redes de Co-Autoria (*co-authorship networks*)
    - Redes de Citações (*citation networks, co-citation networks*)
    - Redes Temáticas: palavras-chave e termos em comum
    - Outras?

setembro 2005

45

## Semantic Association Networks

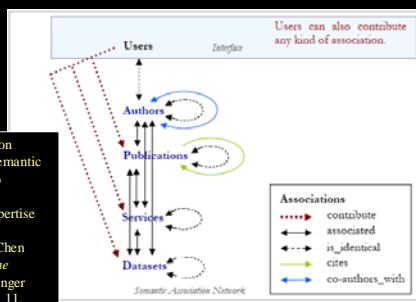
- *Semantic Association Networks* (Katy Börner)
  - Redes de publicações, ligadas via citações e/ou via autoria, de um mesmo provedor ou entre provedores

setembro 2005

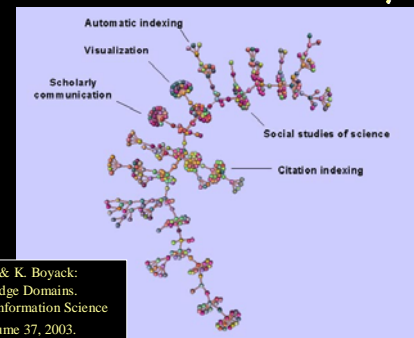
46

## Semantic Association Networks

Semantic Association Networks: Using Semantic Web Technology to Improve Scholarly Knowledge and Expertise Management; V. Geroinenko & C. Chen (eds.) *Visualizing the Semantic Web*, Springer Verlag, 2<sup>nd</sup> Ed., Ch. 11.



## Document Co-citation Map



K. Börner, C. Chen & K. Boyack: Visualizing Knowledge Domains. *Annual Review of Information Science & Technology*, Volume 37, 2003.

## Mapeamento de Coleções de Documentos

- Teoria de redes complexas
  - Modelo para extrair métricas e aprender sobre a dinâmica do processo de colaboração e da evolução e surgimento de áreas?
  - Classes de redes apresentam propriedades topológicas que caracterizam a sua conectividade e a dinâmica dos processos que ela modela
  - Medidas associadas às propriedades topológicas descrevem o comportamento da rede
    - Em muitas aplicações, ênfase na dinâmica e na simulação de comportamentos (para análise e previsão)

setembro 2005

49

## Mapeamento de Coleções de Documentos

- email K. Börner, 29/08 (Ph.D. ou Pos-Doc na Universidade de Indiana)
  - "Our major research agenda is to build computational models that simulate individual scientists and their production and consumption of scholarly works, and to validate these models to large-scale databases of scholarly output. We would like to analyze and compare these empirical corpora and simulated data in terms of general statistics, local and global network properties, and general laws that describe the mechanisms of science..."

setembro 2005

50

## Mapeamento de Coleções de Documentos

- Representações visuais
  - Se relações são modeladas como redes complexas, quais as implicações para a interação com visualizações geradas para apoiar análise e busca?
    - Mecanismos específicos de filtragem, identificação de agrupamentos, consulta?

setembro 2005

51

## Conexões

- Algoritmo "Clique Averaging" (Serge Belongie, UCSD)
  - Aborda o problema do *clustering* (particionamento) em domínios em que as relações de afinidade envolvem mais de 2 elementos
  - Nem sempre a medida de similaridade existe para pares de elementos de dados; para alguns problemas é necessário considerar relações entre 3 ou mais elementos
  - Instância do problema de particionamento em hipergrafos

setembro 2005

52

## Conexões

- Hipergrafo ponderado, não direcionado
  - par  $H = (V, h)$ ;  $V$  é o conjunto de vértices, e sub-conjuntos de  $V$  são os hiper-arcos
  - Função  $h$  associa pesos (não negativos) a cada hiper-arco
- "Clique Averaging" é uma solução para particionamento em hipergrafos

setembro 2005

53

## Conexões

- Autores do "Clique Averaging" o aplicam para tratar problemas difíceis em Visão Computacional, p.ex.:
  - Agrupamento de imagens invariante a condições de iluminação: imagens de um mesmo objeto, coletadas sob diferentes condições de iluminação
  - *k-lines clustering*: dado um conjunto de pontos em  $\mathbb{R}^d$ , particioná-los em um número de linhas  $d$ -dimensionais

setembro 2005

54

## Conexões

- *Co-authorship and co-citation networks* podem ser tratados como hipergrafos
- Interessante modelar redes complexas como hipergrafos?
- Seria o Clique Averaging um algoritmo interessante para detectar partições representativas?

setembro 2005

55

## Referências

- Albert-Laszlo and Eric Bonabeau Barabasi, "Scale-Free Networks", *Scientific American* (New York, NY), May 2003, 68. Page 9
- [http://en.wikipedia.org/wiki/Six\\_degrees\\_of\\_separation](http://en.wikipedia.org/wiki/Six_degrees_of_separation)
- <http://cyvision.if.sc.usp.br/~francisco/networks/>

setembro 2005

56

## Referências

- Luciano da Fontoura Costa, "Redes 'Complexas': modelagem simples da natureza", *Ciência Hoje*, Vol. 36, n. 213, março, 2005.
- Albert-László Barabási & Zoltán N. Oltvai, "Network Biology: understanding the cell's functional organization". *Nature Reviews Genetics* 5, 101-113 (2004)

setembro 2005

57