

Why Network Science can contribute to the advancement of Science *(and your research)*

Daniel R. Figueiredo
LAND – PESSC/COPPE/UFRJ

V Workshop Científico LogIA – PenCogLin
19-20/11 – UFC, Fortaleza

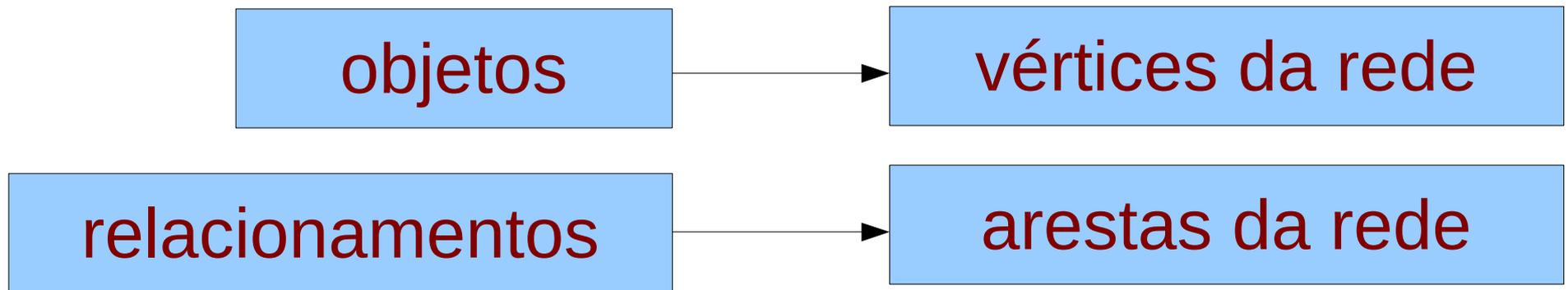


Roteiro

- I) Redes complexas**
- II) Características das redes**
- III) Vantagem cumulativa em redes**
- IV) Meus dois centavos**

Definindo Redes

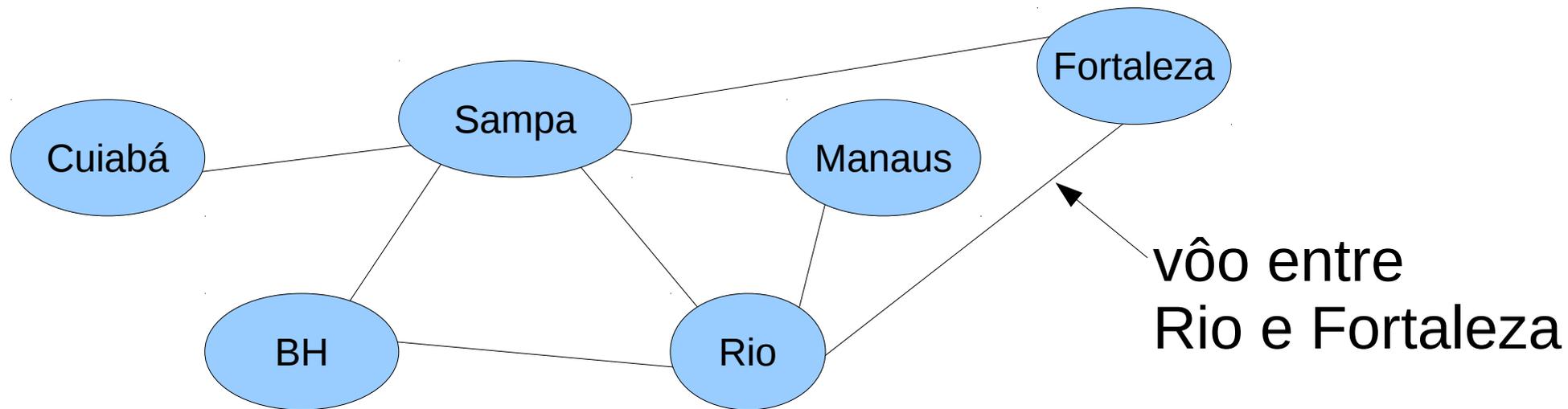
- Abstração que permite codificar relacionamento entre pares de objetos



- Redes como ferramenta para representar *estrutura*
 - objetos e relacionamentos físicos ou abstratos

Rede de Transporte Aéreo

- ❑ Objeto: cidades
- ❑ Relacionamento: voo comercial entre cidades



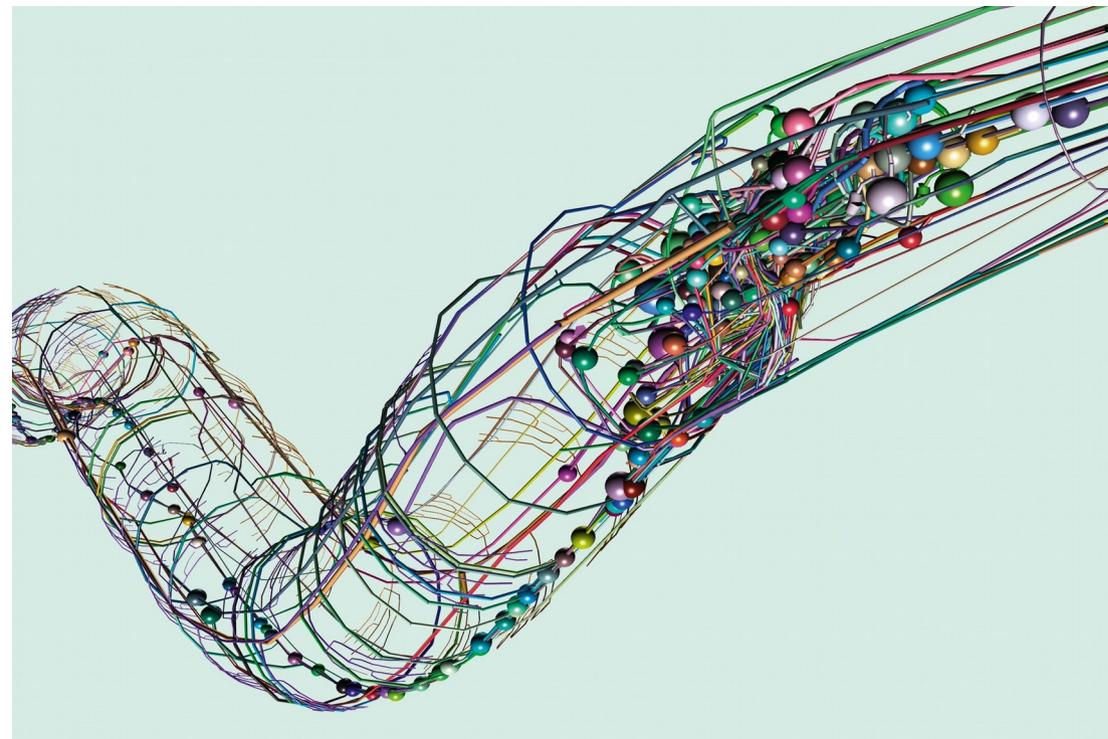
**Rede codifica a estrutura
induzida pelo relacionamento**

Rede Neural

- ❑ *C. elegans* (roundworm)
- ❑ Sinapses entre neurônios



- ❑ 302 neurônios
- ❑ completamente mapeada anos 70
 - prêmio Nobel 2002



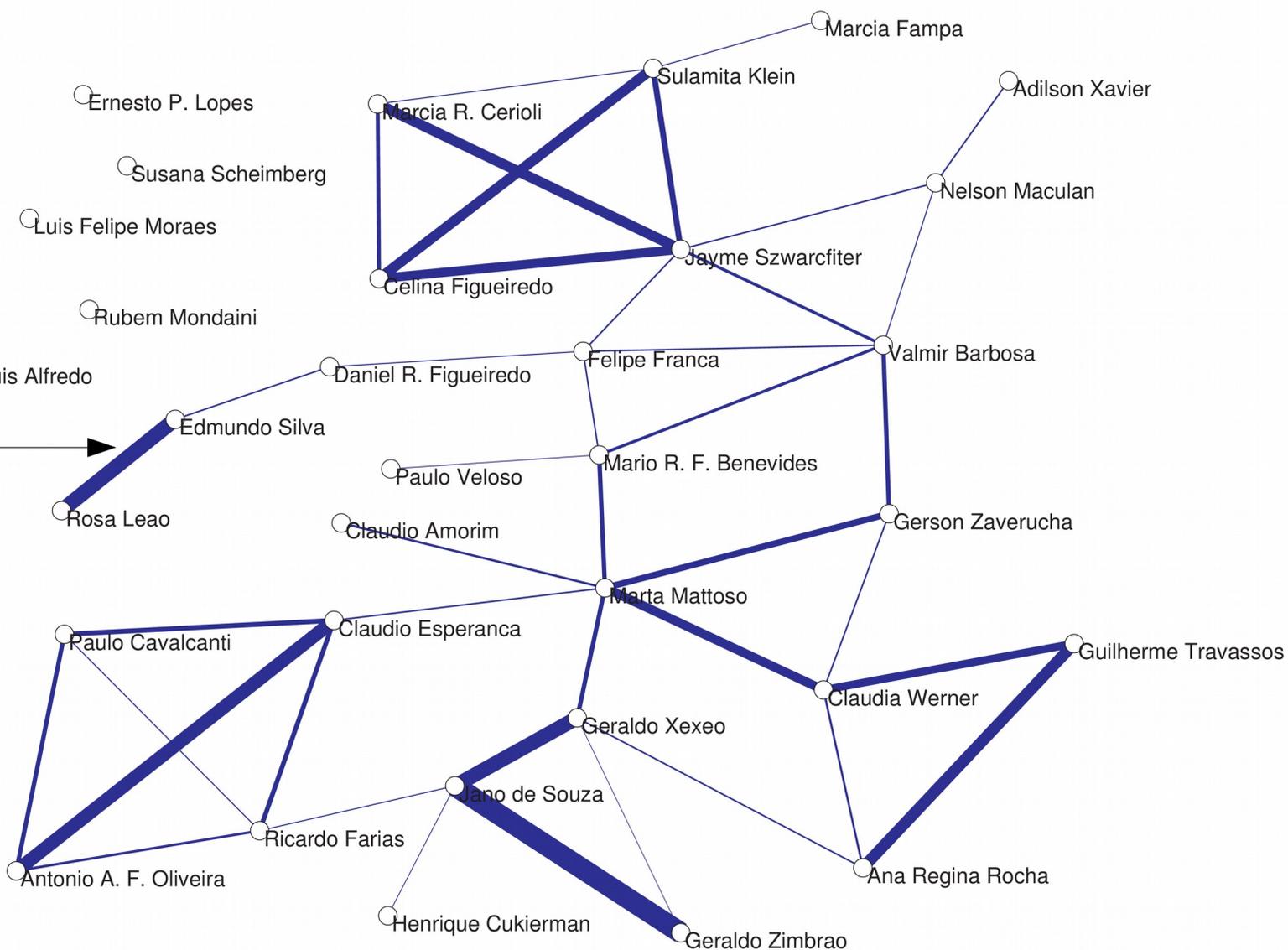
Importância da
estrutura?

Rede de Colaboração

- Objeto: pesquisadores
- Relacionamento: publicação em conjunto

PESC/COPPE (DBLP 09)

Intensidade do relacionamento



Definindo Complexidade

- ❑ Difícil definição, mas com seguintes ingredientes

objetos

interação

emergência

- ❑ Do trivial ao não-trivial
- ❑ Sistemas complexos

Redes Complexas

objetos

relacionamentos

interação

emergência



cultura

memória

apagão

**crise
financeira**

Fenômenos que ainda não explicamos!

Network Science

- ❑ Entender como e porque as coisas se conectam e consequências desta conectividade



“Coisas que se conectam” —————▶ **Redes**

“Como, por que, e consequências” —————▶ **Complexo**

- ❑ **Estrutura** assume papel central
 - pedaço que faltava?

Traga a estrutura para sua pesquisa!

Roteiro

- I) Redes complexas
- II) Características das redes**
- III) Vantagem cumulativa em redes**
- IV) Meus dois centavos**

Quatro Importantes Características

- ❑ Observada em diversas redes reais
 - a partir do final da década de 90



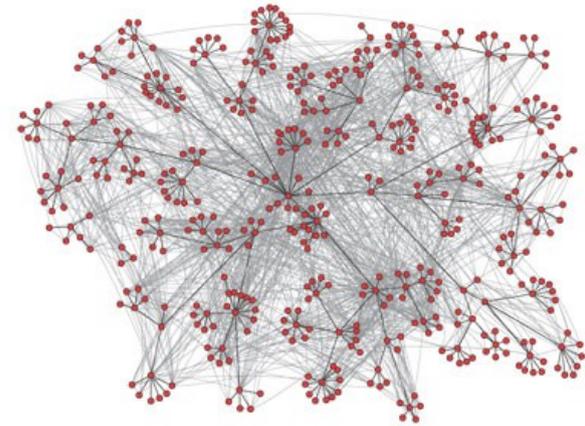
■ Mundo pequeno



■ Meus amigos são amigos



■ Normalidade ausente



■ Tudo Conectado

It's a small world after all

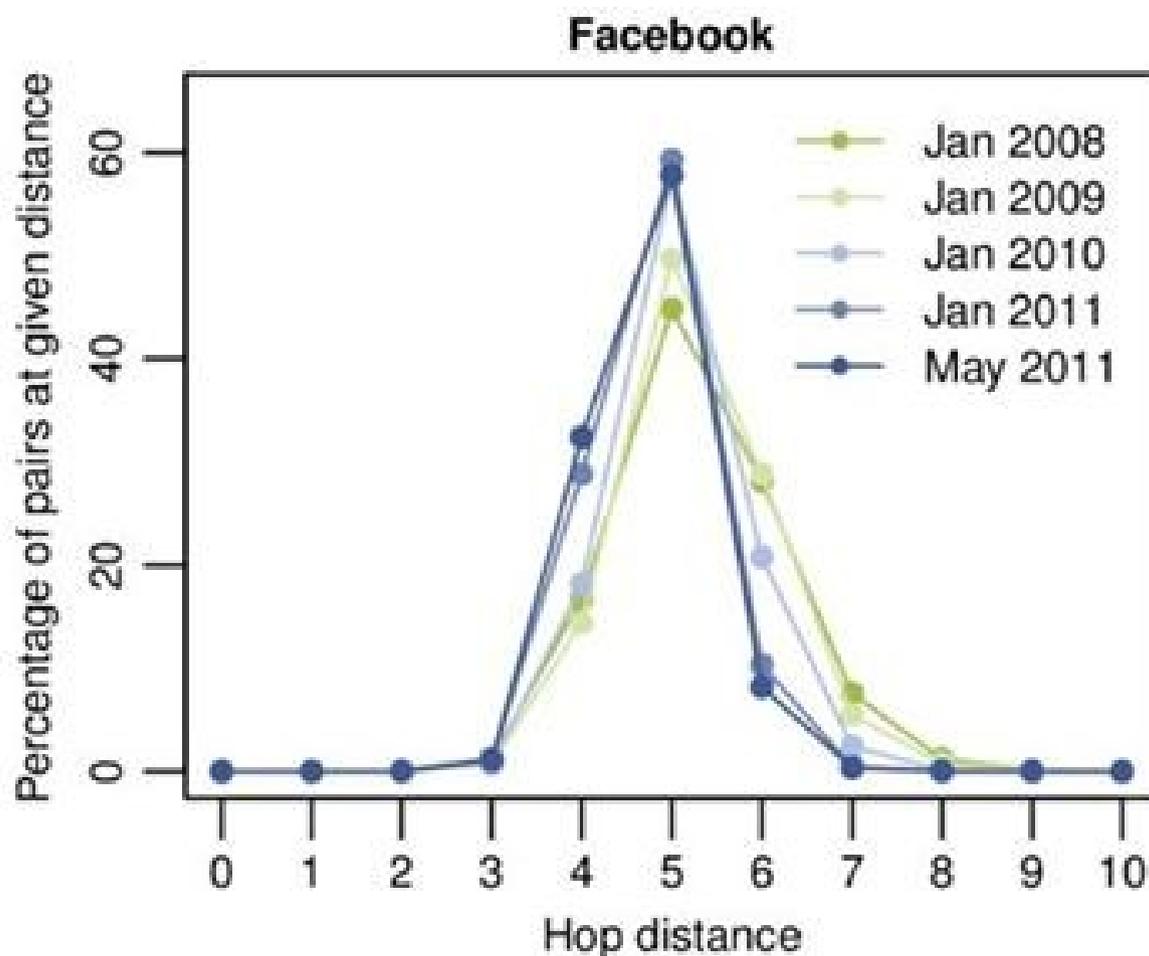


- Distância média **muito** pequena, diâmetro também
 - mesmo para redes muito grandes
- Web (parte) - 10^8 vértices ← 7.5
- Rede de colaboração - 10^6 nodes ← 4.9
- Facebook - 10^9 nodes ← 4.5
- e muitas outras!

**Ordens de grandeza menor!
Aparentemente da ordem de $\log n$**

Exemplo do Facebook

- Distribuição da distância ao longo do tempo (truncada em 10)



vértices (arestas)

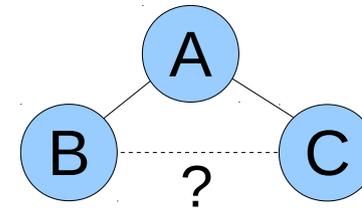
	fb
2007	13.0 M (644.6 M)
2008	56.0 M (2.1 G)
2009	139.1 M (6.2 G)
2010	332.3 M (18.8 G)
2011	562.4 M (47.5 G)
current	721.1 M (68.7 G)

- Rede cresce (50 x), mas distância média diminui!

Meus amigos também são amigos



- A relacionado com B e C faz com que B e C se relacionem mais provavelmente
- clusterização: chance de dois vizinhos serem vizinhos
- densidade: chance de dois vértices ao acaso serem vizinhos



	clusterização	densidade
■ AS graph - 10^4 nodes	← 0.39	0.00056
■ Facebook - 10^9 nodes	← 0.14	0.00000026
■ Biology coauthorship	← 0.67	0.00001

Ordens de magnitude maior!

Normalidade Ausente



- Grau dos vértices é muito desigual
- Muitos com grau pequeno, poucos com grau muito grande

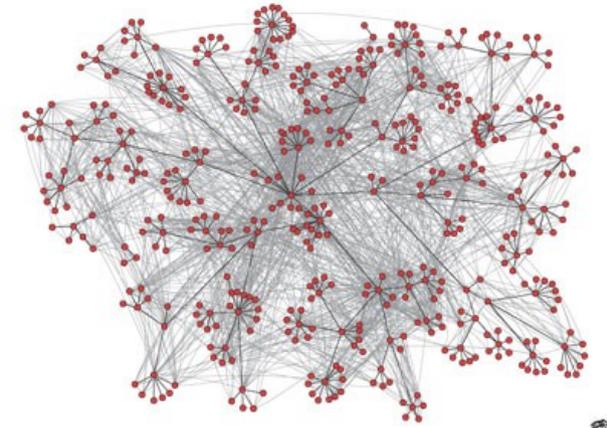
Parecido com distribuição da renda no Brasil!

- | | |
|---------------------------|---------------------------|
| ■ AS Graph – 10^4 nodes | ■ Citações – 10^6 nodes |
| ■ Grau médio: 5.9 | ■ Grau médio: 8.6 |
| ■ Grau máx: ~2100 | ■ Grau máx: ~9000 |

**Distribuição de grau possui cauda pesada
Abrange diversas orders de grandeza**

- Muito diferente da distribuição normal ou binomial

Tudo Conectado



- Redes não são conexas
- Maior componente conexa possui **quase todos** os vértices - CC gigante
- Outras componentes muito pequenas
 - muitas outras componentes
- Rede de palavras
 - *thesaurus* ~1K vértices, 97% CCG
 - Co-ocorrência ~ 500K vértices, 100% CCG
- Rede social, rede neural, etc

**Quase sempre quase
completamente conectada!**

Curiosidade

- **Fato I:** Muitas redes possuem propriedades estruturais peculiares (não esperadas)
 - grau, distância, clusterização, conectividade, etc
- **Fato II:** Muitas redes diferentes possuem propriedades estruturais semelhantes
 - Web, Facebook, AS Graph, Neural network, etc



- *“Million dollar question”*

Por que?



- Algumas respostas na literatura, mas nada muito definitivo

Roteiro

- I) Redes complexas
- II) Características das redes
- III) Vantagem cumulativa em redes**
- IV) Meus dois centavos**

Vantagem Cumulativa

50 pessoas



- Dois restaurantes, pouco conhecimento
- Em qual você entra?



6 pessoas



- VC:** Recursos acumulados promovem o acúmulo de mais recursos
 - preferential attachment, rich-gets-richer, Mathew effect, network effect*

Vantagem Cumulativa

- ❑ Presente na dinâmica de muitos processos
- ❑ Evolução das espécies (Yule'25)
 - espécies mais diversas atraem mais espécies
- ❑ Uso de palavras (Zipf'35)
 - palavras mais usadas atraem mais uso
- ❑ Formação de cidades (Simon'55)
 - cidades maiores atraem mais pessoas

Observada empiricamente em diferentes contextos



VC em Redes

❑ Como redes se formam ou evoluem?

❑ **Ideia:** VC guia processo de formação

o vértices muito relacionados atraem mais relacionamentos



atração pelo grau

❑ Observada empiricamente em muitas redes

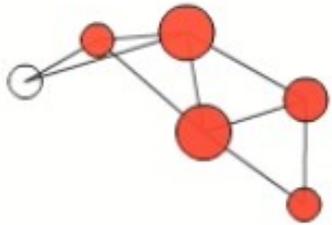
o web, rede de citação, rede de AS (Internet), Twitter, etc

Modelo BA

- ❑ Modelo simples para capturar VC em redes (*preferential attachment*)
 - proposto por Barabási-Albert, 99 [+23K citações]
- ❑ Modelo iterativo
 - iniciar com pequeno clique
 - a cada passo, adicionar 1 vértice na rede com grau m
 - escolher vizinhos (m pontas) com prob. proporcional ao graus

Modelo BA em Ação

- ❑ Clique inicial com 3 vértices, $m = 2$



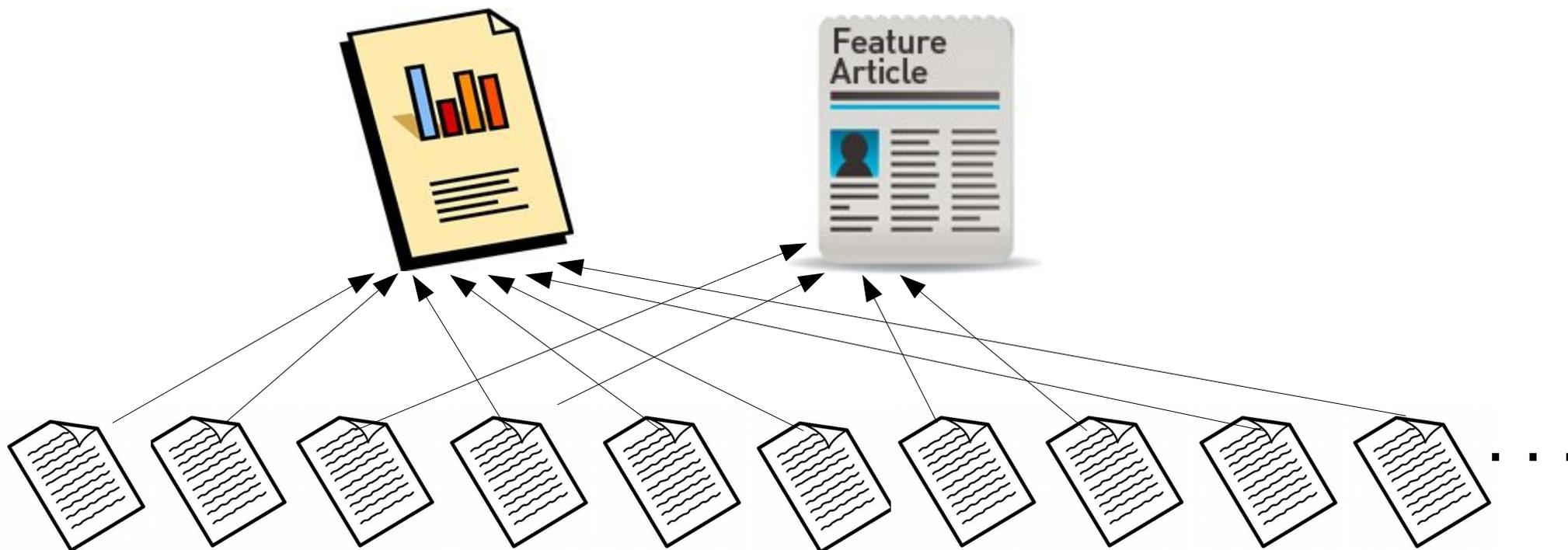
- ❑ Distribuição de grau segue *lei de potência*
- ❑ Explicação para desigualdade de grau nas redes

Roteiro

- I) Redes complexas
- II) Características das redes
- III) Vantagem cumulativa em redes
- IV) Meus dois centavos**

VC em Competições

- ❑ Agentes competindo por recursos
 - ex: artigos competindo por citações



Qual o impacto de VC na dinâmica?

Impacto de VC

- ❑ Modelo simples misturando habilidade e sorte
 - com VC: Polya urn
 - sem VC: Random walk
- ❑ Caracterização da intensidade e duração da competição
 - duração: instante do último empate
- ❑ Resultado: “*Struggle of the fittest*”
 - duração passa de exponencial para lei de potência
 - mais hábil pode levar muito mais tempo para ganhar quando temos VC

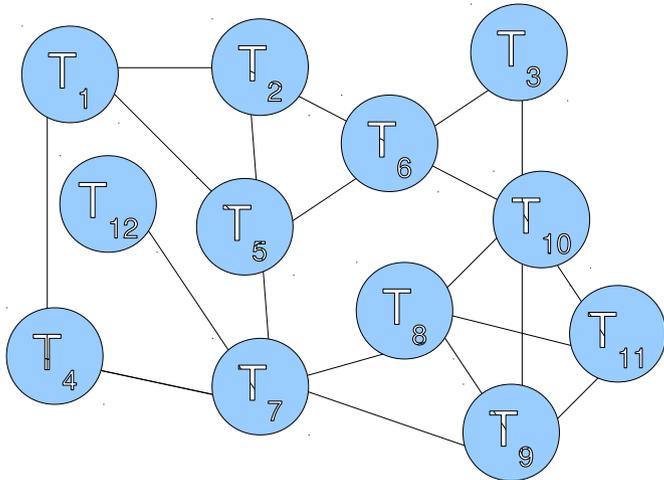
• Colaboração com: Bo Jiang (UMass), Bruno Ribeiro (Purdue) e Don Towsley (UMass)

Emparelhamento em Redes

- Mapear vértices de dois grafos relacionados
 - vértices remetem aos mesmos objetos

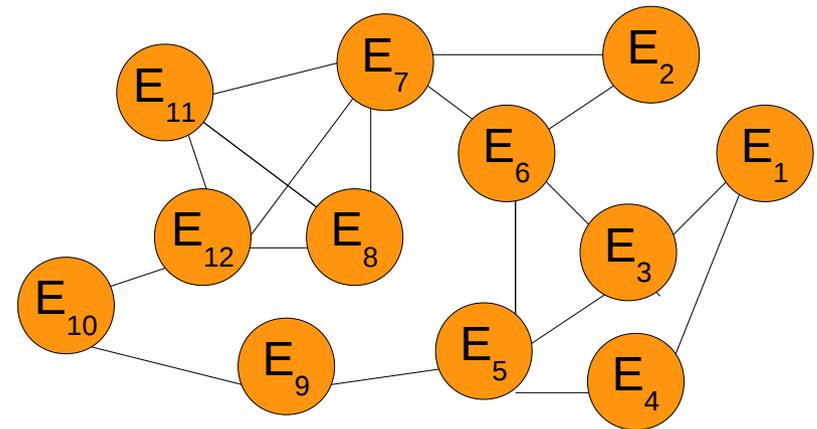
- Rede de telefonemas

- nó = número de telefone
- aresta = chamada telefônica

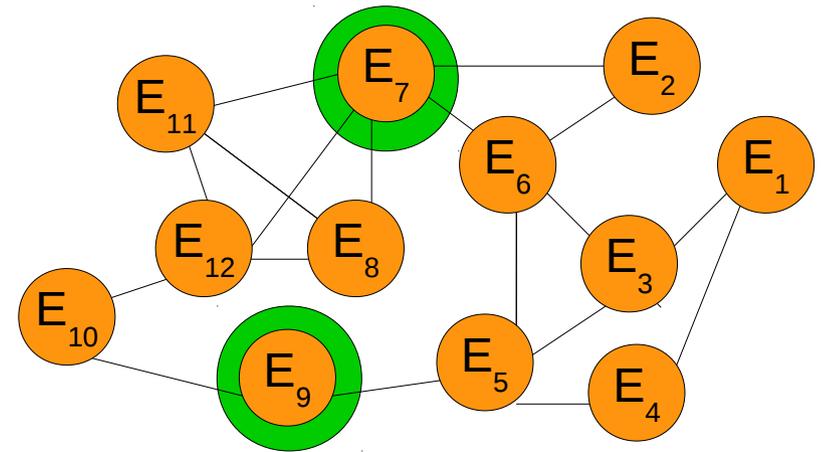
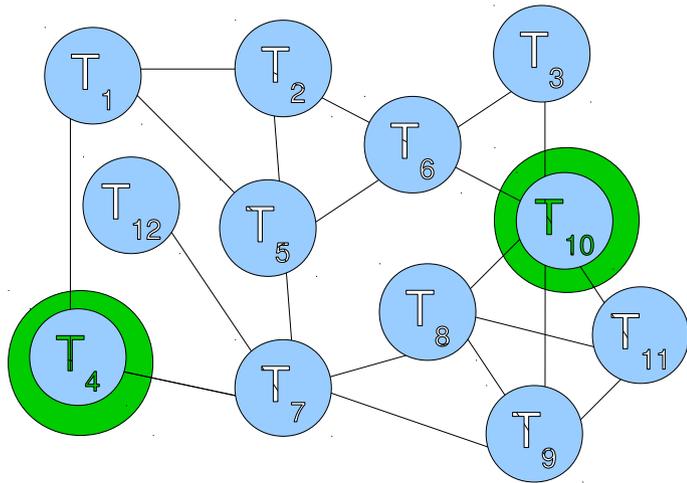


- Rede de emails

- nó = endereço de email
- aresta = troca de msgs



Emparelhamento em Redes



- Supor telefones e emails correpondam ao mesmo conjunto de pessoas
 - ex. Carlos E. F. Brito: $T_4 \leftrightarrow E_9$



Podemos mapear os vértices?

- Qual número de telefone corresponde a qual endereço de email?

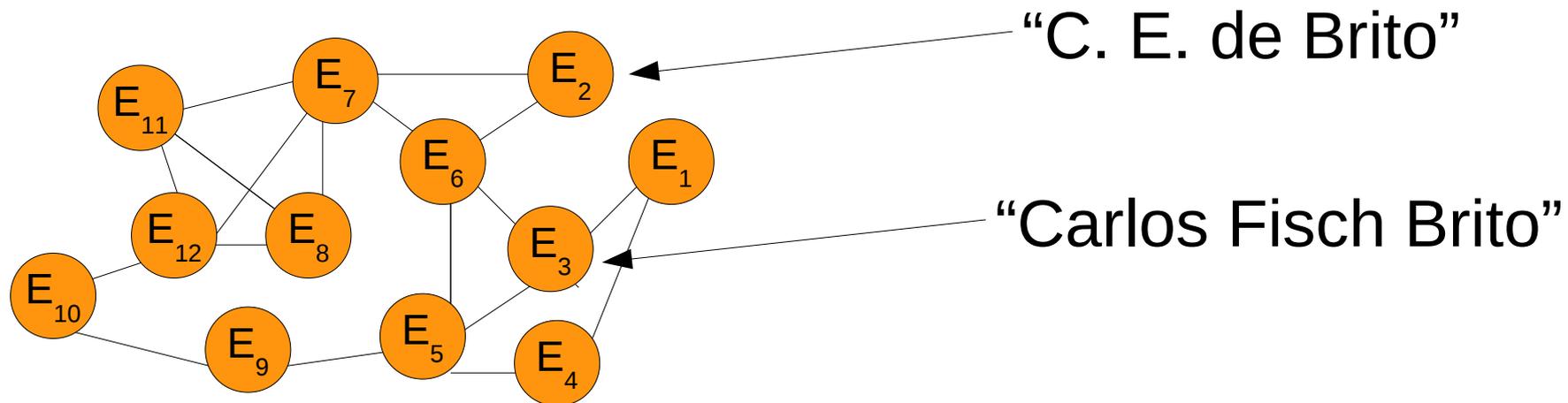
Emparelhamento em Redes

- ❑ Generalização do problema de isomorfismo
 - grafos são diferentes
- ❑ Abordagem algorítmica
 - modelagem probabilística Bayesiana utilizando estrutura como evidência
 - algoritmo eficiente iterativo - *max weight match*
- ❑ Abordagem teórica
 - papel da estrutura: condição estrutural para que solução correta seja possível
 - papel de múltiplas redes (e não duas)

♦ Colaboração com: M. Grossglauser (EPFL), J. Simões (UFRJ) e P. Pedarsani (ex-EPFL)

Ambiguidade em Redes

- ❑ Vértices remetem ao mesmo objeto
 - ex. redes de colaboração, vértices identificados pelo nome na publicação

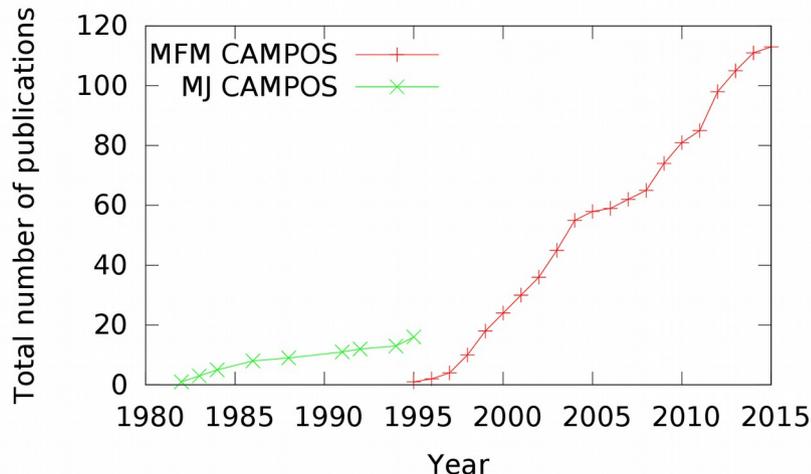


Como remover ambiguidades?

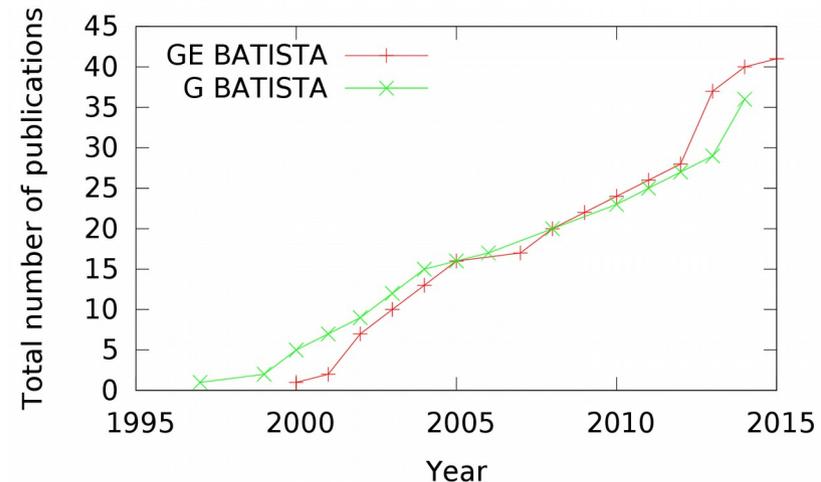
- ❑ Rede de pessoas e não de rótulos

Ambiguidade em Redes

- ❑ Classificação dos padrões de uso de nome
- ❑ Estudo empírico de casos positivos
 - DBLP e Google Scholar



Mudou de nome!



Convive com dois nomes!

- ❑ Reflexo das classes na estrutura da rede
 - auxiliar remoção de ambiguidade

Resumo da Ópera

- ❑ Redes Complexas → ❑ Como, por que e consequências das redes que nos cercam
- ❑ Caracterizando as redes que os cercam → ❑ Mesmas peculiaridades em redes diferentes
- ❑ Vantagem cumulativa em redes → ❑ Princípio unificador (e antigo)
- ❑ Meus dois centavos → ❑ Diferentes problemas focando na estrutura

Obrigado!



Perguntas ou comentários?

daniel@land.ufrj.br