

DCA

DEPARTAMENTO DE ENGENHARIA DE
COMPUTAÇÃO E AUTOMAÇÃO
INDUSTRIAL



FACULDADE DE ENGENHARIA
ELÉTRICA E DE COMPUTAÇÃO



UNICAMP

IA013 – Tópico 6 – Parte 3: Redes complexas e outros modelos em rede



www.lbic.fee.unicamp.br

Fernando José Von Zuben

Parte desse material foi produzido por Alan Godoy Souza Mello e George Barreto Pereira Bezerra (DCA/FEEC/Unicamp).

Introdução

Contextualização

- Em diversos ramos da ciência é comum a formulação de problemas valendo-se de estruturas em redes (grafos).
- Essa abordagem contrapõe-se às abordagens mais reducionistas, uma vez que ela considera não apenas as partes formadoras como também as interações dessas partes.
- Capaz de explicar o surgimento de efeitos advindos de comportamento emergente.

Introdução

Histórico

- Início das pesquisas no meio do século passado.
- A pesquisa em redes complexas se beneficia do aumento do poder de processamento, o que permite a análise de dados de redes reais.
- Pesquisa inicial de redes reais pautava-se pelos modelos aleatórios de ERDÖS & RÉNYI (1959).
- Modelos mais realistas aparecem com as redes *small world* (WATTS & STROGATZ, 1998).
- Proposta das redes *scale-free* por BARABÁSI & ALBERT (1999).

Introdução

Onde estão presentes?

- Ciências Sociais: redes de interação social, redes de relações entre palavras de uma língua, redes de colaboração científica, redes de contato sexual.
- Ciências Econômicas: sistema de interação de consumidores com produtores, rede de fornecedores para a síntese de produtos de alta tecnologia e alto valor agregado.
- Ciências Biológicas: na ecologia, em redes tróficas e redes sociais intra e inter-específicas, na genética, em redes de interações gênicas, e na medicina, na dispersão de doenças infecciosas, em redes metabólicas e em redes neurais reais.

Introdução

Onde estão presentes?

- Como exemplo de redes de interações sociais, temos o experimento de Stanley Milgram (6 níveis de distância).
- As redes de relações entre palavras de uma língua envolvem tanto as palavras com as quais outras palavras se relacionam (como elas se conectam em uma frase), como a rede de palavras em um dicionário;
- Redes de citações foram as primeiras a serem estudadas nesse contexto de redes complexas (NEWMAN, 2001a,b);
- A rede neural do verme *Caenorhabditis elegans* é livre de escala e forma uma rede mundo pequeno (*small world*).

Introdução

O experimento de Milgram

- O efeito de mundo pequeno surge a partir do trabalho de MILGRAM (1967), que realizou um experimento no qual um conjunto de pessoas deveria enviar cartas a seus conhecidos e estes a seus contatos, recursivamente, até que estas cartas chegassem a um conjunto de indivíduos alvos.
- Muitas das cartas se perderam, porém aquelas que chegaram a seu destino passaram, em média, pela mão de seis pessoas, uma distância pequena quando comparada à quantidade de pessoas existentes no planeta e à pequena quantidade de conexões sociais entre estas pessoas.

Introdução

O experimento de Milgram

- No experimento de MILGRAM (1967), os envolvidos possuíam apenas informações como primeiros nomes, cidades em que residem e emprego dos destinatários.
- Logo, cada pessoa podia repassar a carta a outro indivíduo que ela imaginava estar mais próximo do alvo, sem qualquer conhecimento mais profundo sobre a topologia da rede social.
- Mesmo assim, os participantes foram capazes de encontrar caminhos eficientes, conectando as pessoas de origem e de destino.

Introdução

Caenorhabditis elegans

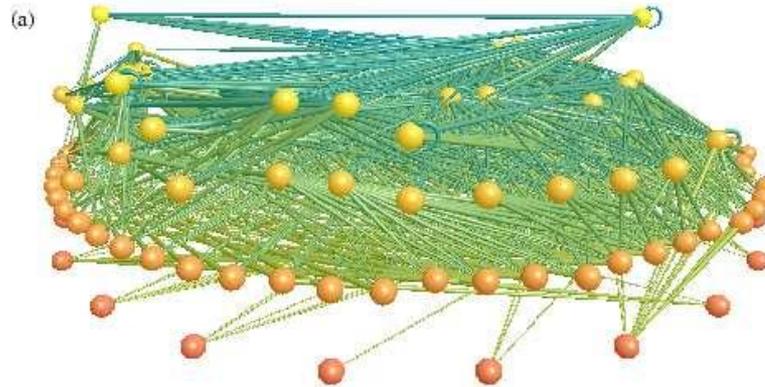
- É um nematelminto (verme cilíndrico).



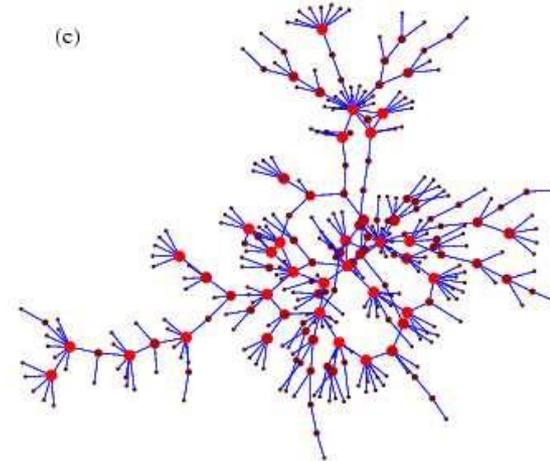
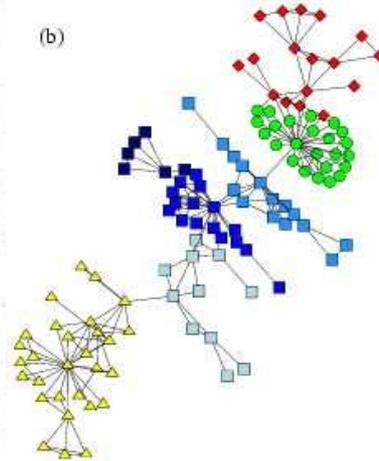
C. elegans

Introdução

Rede trófica inter-espécies em um lago (inclui relações presa-predador)



Colaboração entre cientistas em uma instituição de pesquisa



Rede de contato sexual

Introdução

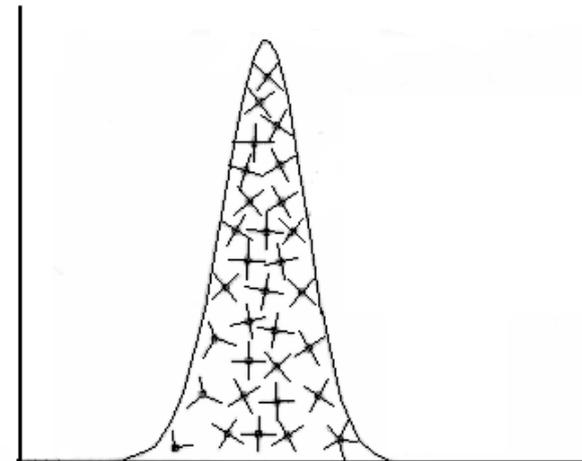
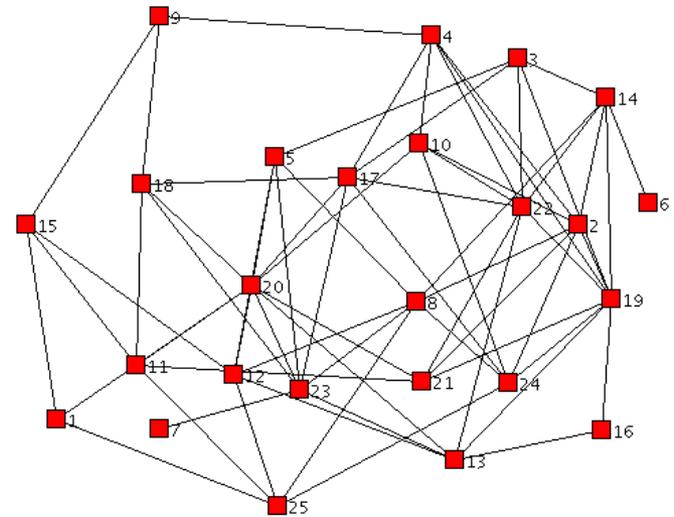
Onde estão presentes?

- Ciências Exatas: representação de redes elétricas, *backbone* da Internet, *links* entre páginas na Web, representação de circuitos integrados, modelagem de mecânica estatística, organização de redes *peer-to-peer*, redes de preferências em um sistema de recomendação, rede ferroviária e redes de dependência entre softwares.
- Uma área de destaque no estudo de redes complexas é a Física.

Redes Aleatórias

Modelo de Erdős e Rényi

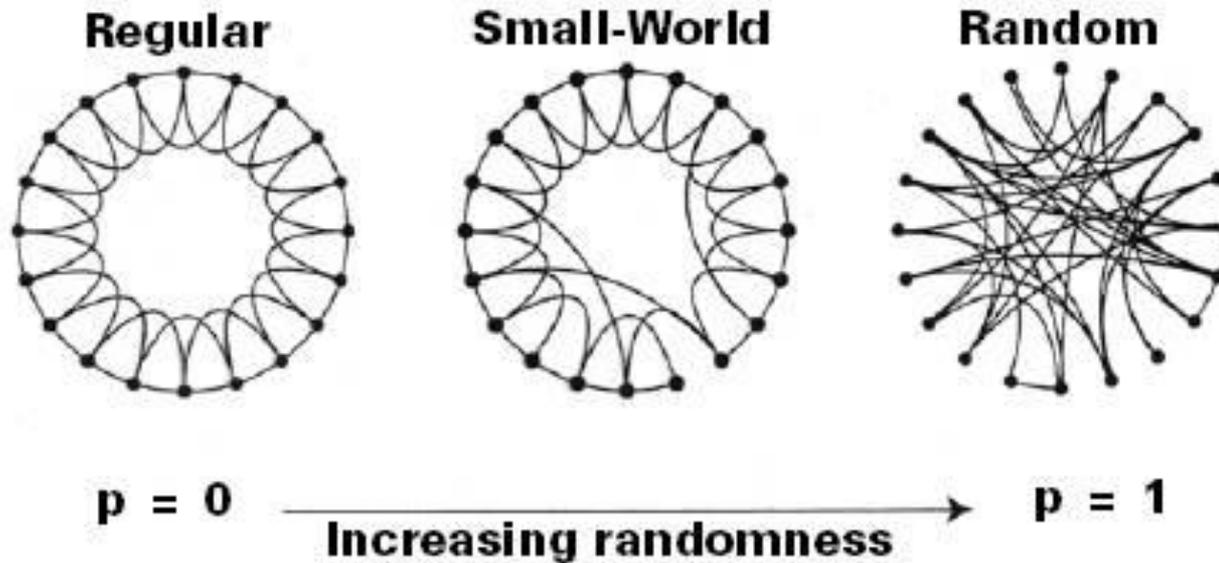
- **Links são adicionados com igual probabilidade entre os nós.**
- **Conectividade segue uma curva de distribuição normal.**
- **Para redes grandes, todos os nós terão aproximadamente o mesmo número de links.**



Redes Small World

Modelo de Watts-Strogatz

- Redes fortemente clusterizadas.
- Algumas conexões aleatórias.



Redes Small World

Modelo de Watts-Strogatz

- Início: a rede é iniciada como um anel com N vértices, cada qual conectado a seus K vizinhos mais próximos.
- Aleatoriedade: cada aresta do grafo original é reconectada com probabilidade p , sem a possibilidade da criação de laços ou multi-arestas. A reconexão de uma aresta é feita ao desligá-la de um de seus extremos, conectando-a a um vértice selecionado aleatoriamente da rede.
- Pode-se observar que, quando $p = 0$, a rede final mantém-se como um anel ordenado, enquanto que quando $p = 1$, a rede torna-se plenamente aleatória.

Redes Livres de Escala (Scale-Free)

Em busca de redes que reflitam a realidade

- Modelos anteriores, como o modelo de grafo aleatório de Erdős & Rényi e o modelo de redes small-world de Wattz & Strogatz, propunham formas de geração de grafos em que a quantidade de vértices no sistema era definida a priori, sendo permitida a criação apenas de novas ligações no decorrer do tempo.
- Apesar de suprir algumas das características comuns às redes complexas, esses modelos mostraram-se insuficientes para explicar a origem de tais sistemas. Em ambos os casos, por exemplo, a distribuição de graus segue uma distribuição de Poisson.

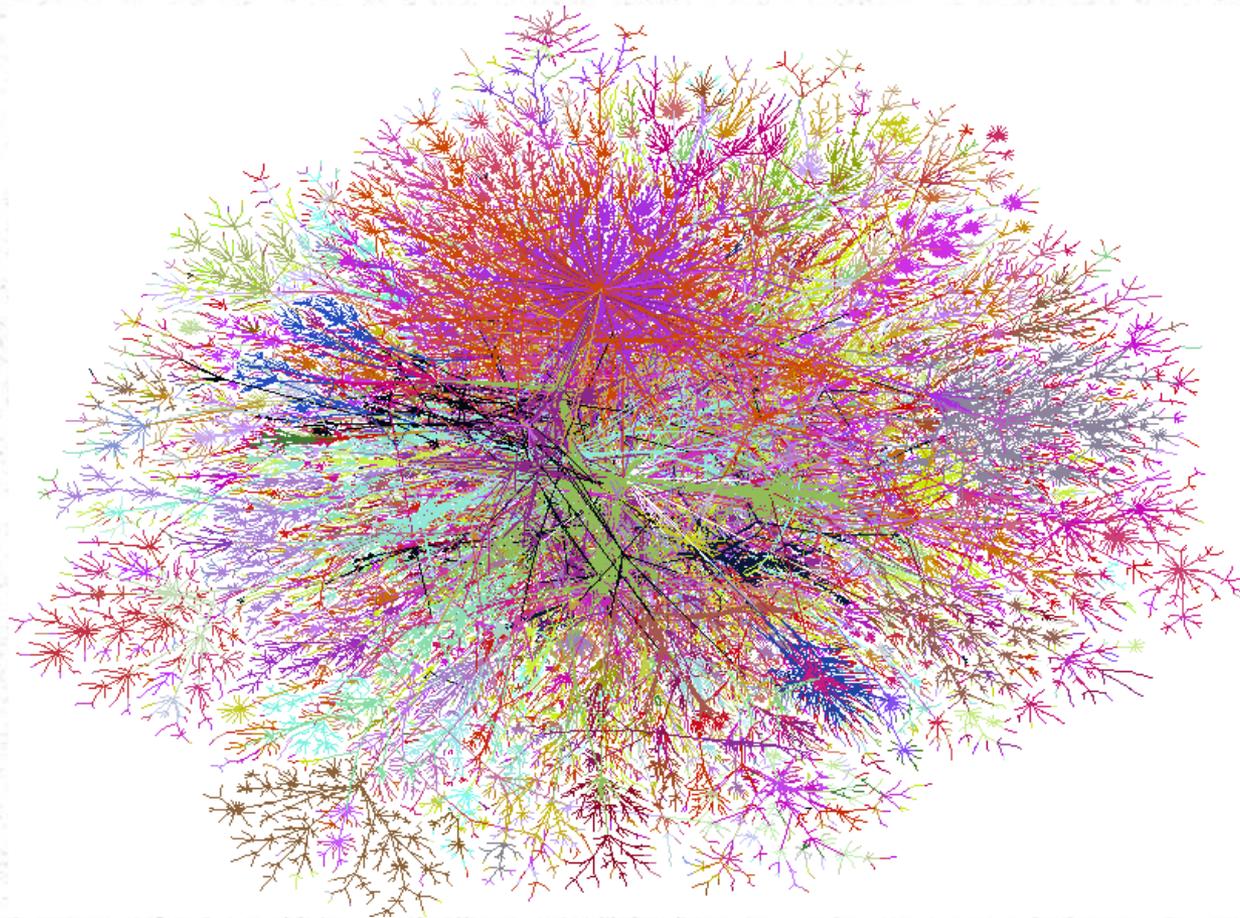
Redes Livres de Escala (Scale-Free)

Em busca de redes que reflitam a realidade

- Através da análise das redes reais, é possível observar que, em geral, elas não são fechadas a novos nós.
- Em redes de interação social é comum o aparecimento de novos membros – através do nascimento ou da mudança de pessoas, por exemplo –, assim como de tempos em tempos surgem novas espécies em redes biológicas, alterando sua estrutura de relações.
- Também é possível verificar que, em algumas redes, para cada nova relação criada há uma probabilidade maior de que o nó que a receba seja um nó altamente conectado. Tal característica pode ser vista na rede de citações científicas, onde artigos já bastante citados apresentam maior probabilidade de receber novas citações.

Redes Livres de Escala (Scale-Free)

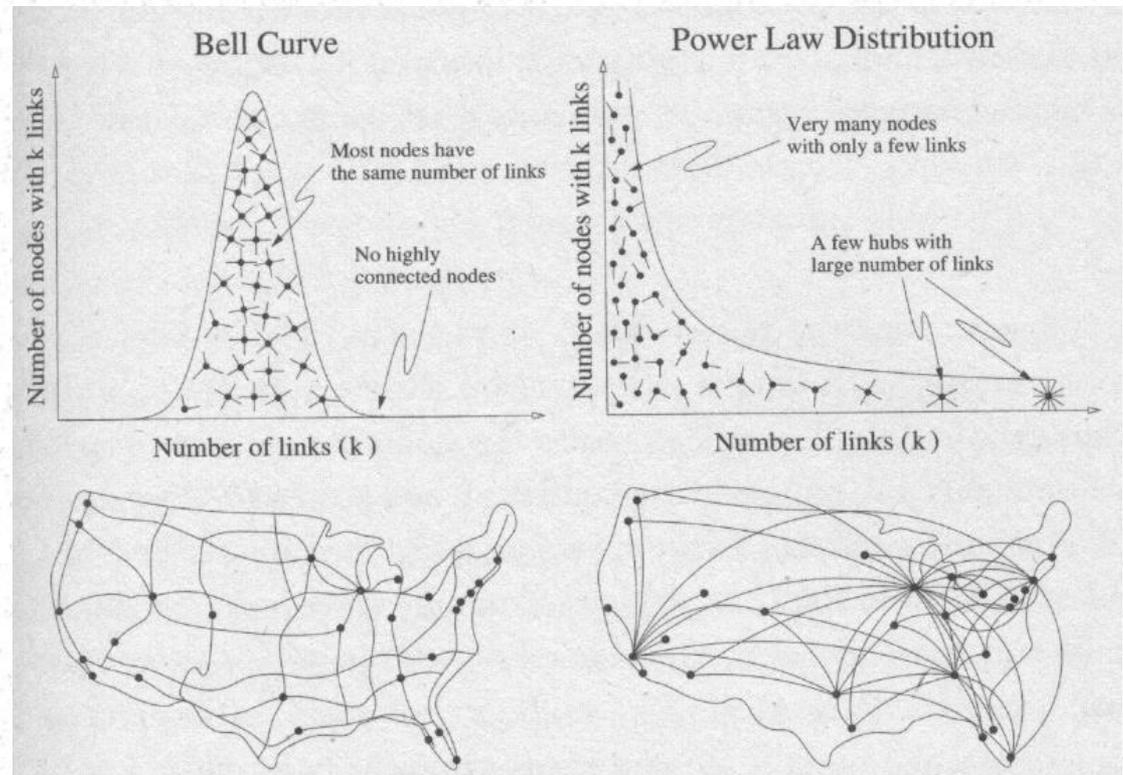
Mapeando a Internet



Redes Livres de Escala (Scale-Free)

Características

- Conectividade segue lei da potência $p(k)=k^{-\lambda}$
- Hubs
- Hierarquia
- Small world



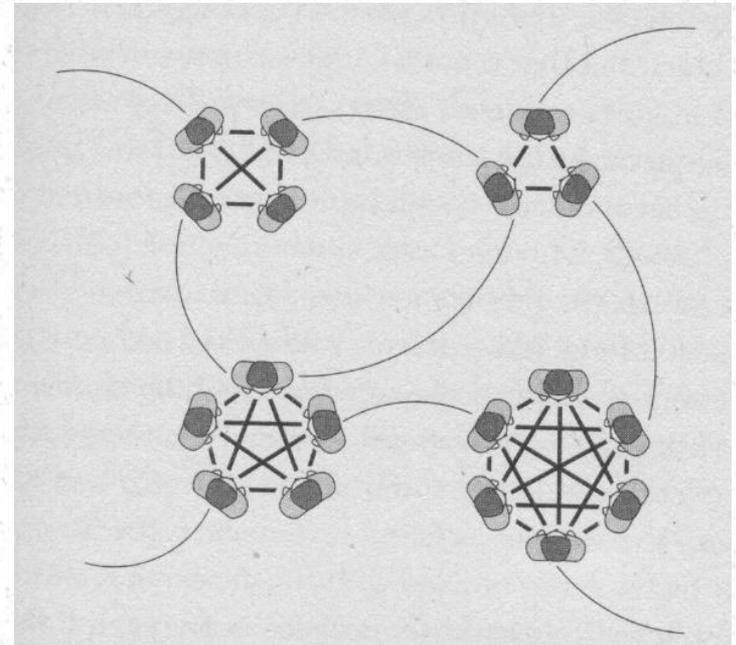
Exemplos de Redes Scale-Free

- Internet
- Redes de interações de espécies
- Redes neurais
- Redes gênicas e proteicas
- Sistemas de tráfego aéreo
- Sistemas de distribuição de energia elétrica
- Redes de interações sociais
- Redes de relações comerciais

Redes Sociais

Grande relevância em seu estudo

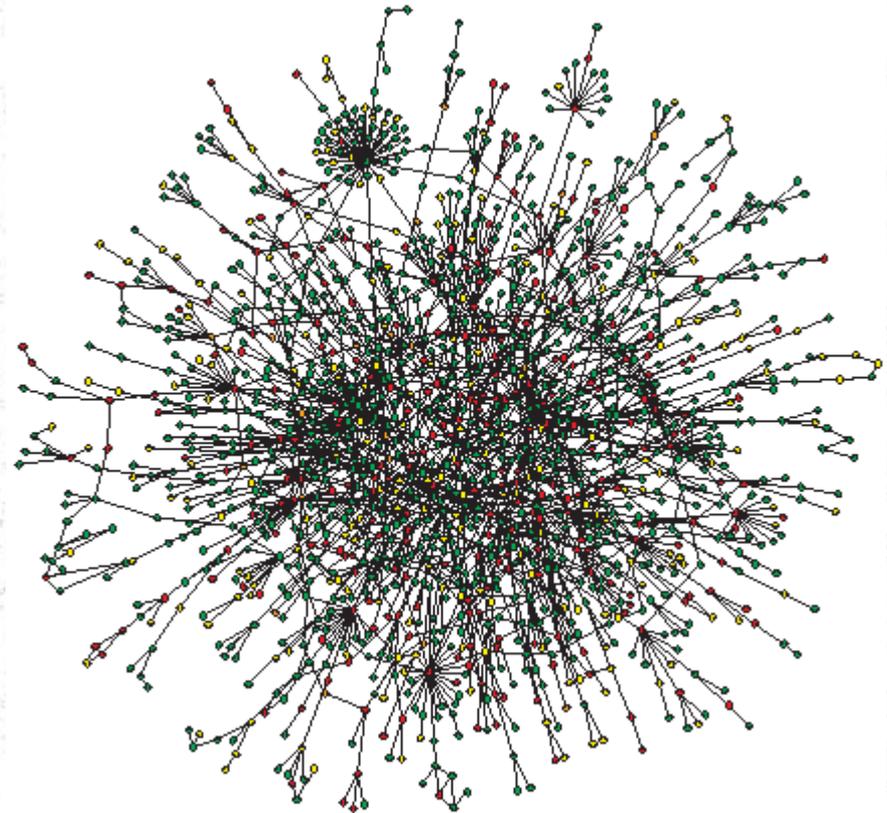
- Orkut, Facebook, LinkedIn
- Redes sociais tendem a ser clusterizadas.
- Importância dos links fracos.



Redes gênicas e proteicas

Características:

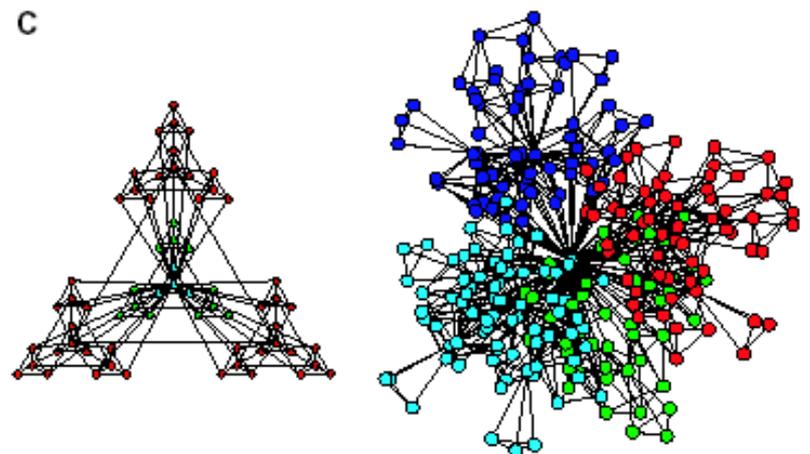
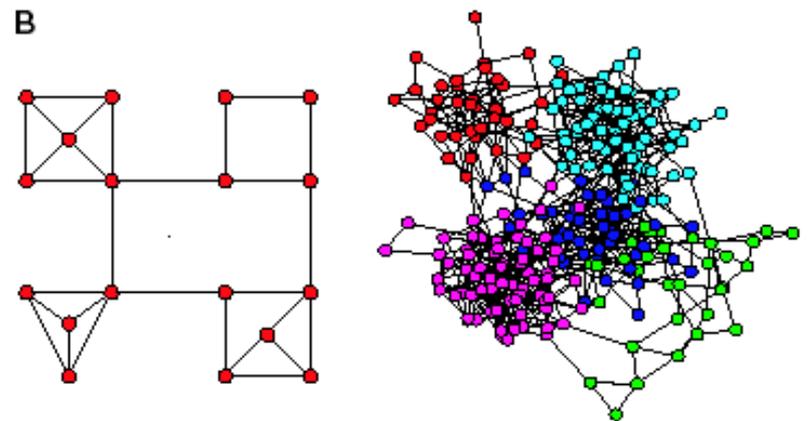
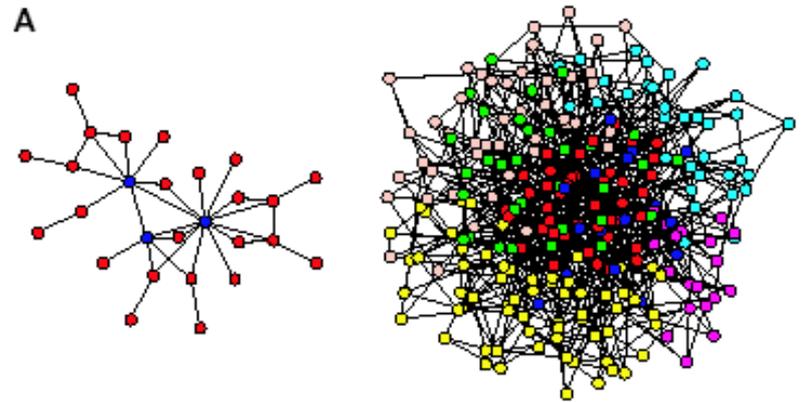
- Power law distribution $P(k) = k^{-\lambda}$
- Formação de módulos
- Small world



A) Scale-free network

B) Rede modular

C) Modularização hierárquica



Distribuição em lei de potência

- Ao contrário do que ocorre em um grafo aleatório, onde espera-se uma distribuição de graus binomial ou de Poisson, nas redes complexas a distribuição de graus se dá segundo a seguinte equação:

$$p(k) \propto k^{-\gamma}$$

sendo k o grau do nó e γ uma constante positiva.

- Essa distribuição é chamada de lei de potência.

Distribuição em lei de potência

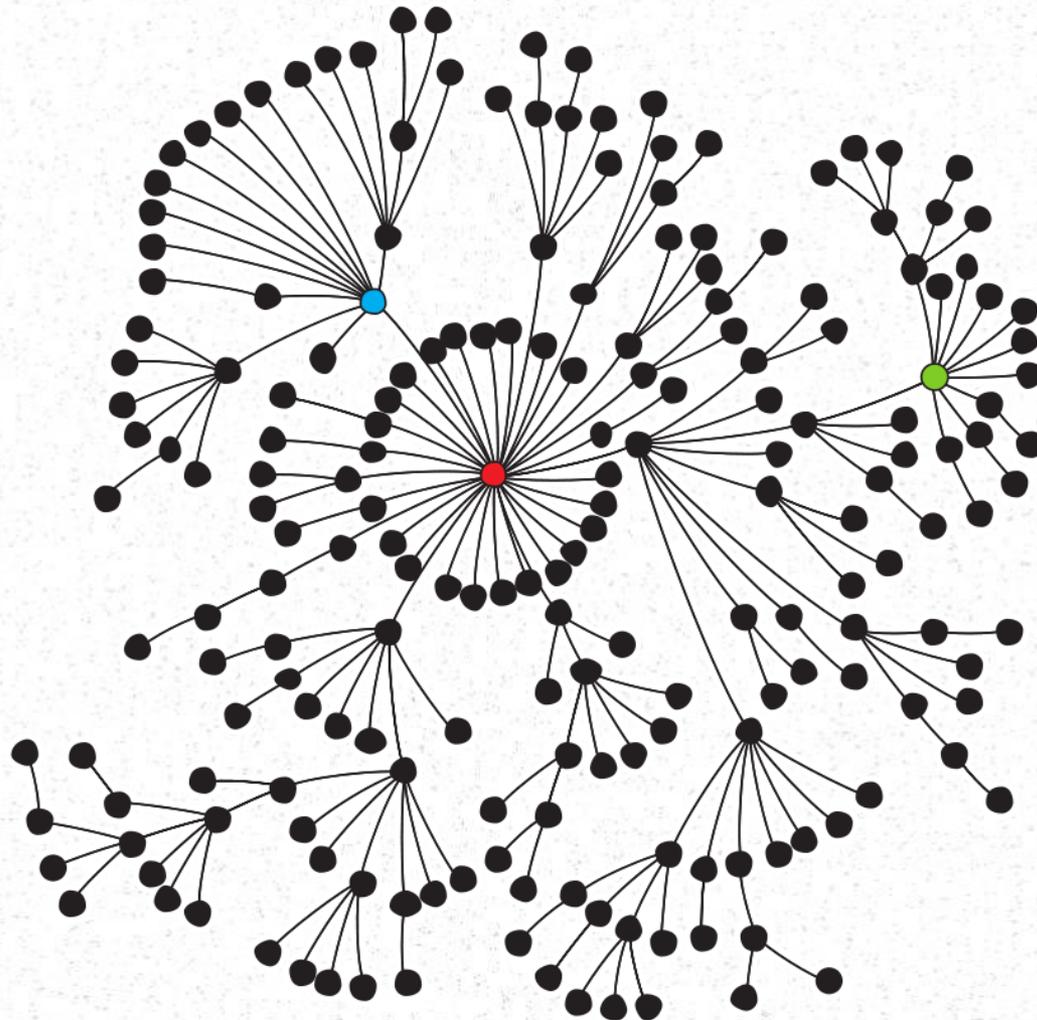
- Apesar de serem esparsas, redes complexas apresentam baixa distância média entre seus nós.
- Essa distância é proporcional a $\log(N)$, onde N é o número de nós na rede, em oposição à relação linear esperada entre o diâmetro da rede e N .
- Essa propriedade é chamada de propriedade de mundo pequeno (ou *small world*).
- Essas redes também apresentam um alto agrupamento (clusterização) e seus nós formam comunidades.

Métodos de Crescimento

Modelo Livre de Escala (Modelo BA)

- Mecanismo proposto por Barabási & Albert, em seu artigo "*Emergence of scaling in random networks*", de 1999.
- Principais modelos anteriores (o grafo aleatório de Erdős & Rényi e as redes *small world* de Wattz & Strogatz) explicavam parcialmente as propriedades de redes complexas.
- Modelos anteriores apresentavam número fixo de nós, alterando apenas a forma como esses estabelecem ligações.

Métodos de Crescimento



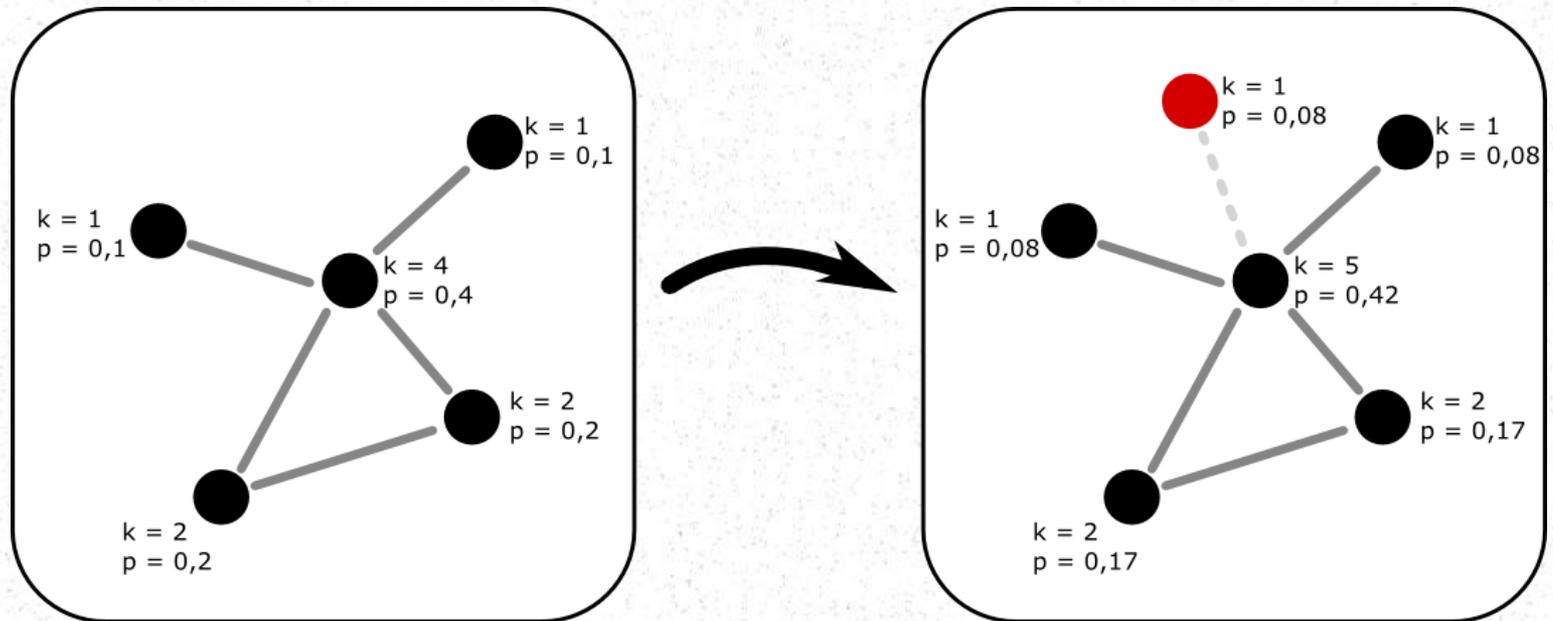
Métodos de Crescimento

Modelo Livre de Escala (Modelo BA)

- O modelo pode ser dividido em duas etapas:
 - Crescimento: a rede deve ser iniciada com uma quantidade m_0 de nós, sendo um novo nó adicionado a cada passo. Junto do nó são adicionadas m ($m \leq m_0$) arestas
 - Conexão preferencial (*preferential attachment*): cada nova aresta deve ligar o novo nó a um nó i existente na rede, escolhido com probabilidade

$$p(i) = \frac{k_i}{\sum k_j}$$

Métodos de Crescimento



Métodos de Crescimento

Modelo Livre de Escala (Modelo BA)

- De acordo com Barabási & Albert, essas duas características são necessárias e suficientes para garantir as características livres de escala.
- De fato, o modelo BA oferece um modelo mais realista, uma vez que, ao contrário das antigas propostas, grande parte das redes reais são abertas ao aparecimento de novos nós.
- Após alguns passos, as propriedades estatísticas da rede tornam-se independentes de tempo e tamanho da rede.

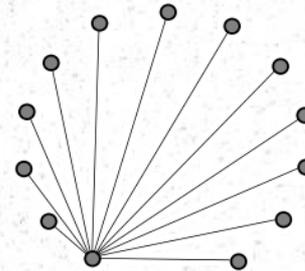
Métodos de Crescimento

Variantes do Modelo Livre de Escala

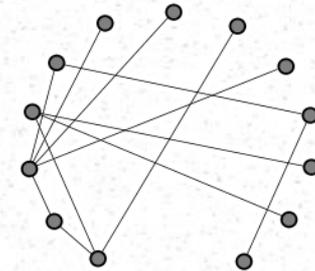
- Adição e remoção de arestas: possibilidade de adição ou remoção de uma ligação não necessariamente relacionada com o nó adicionado.
- *Fitness*: sugere que, em alguns casos, a distribuição de graus é devida à qualidade intrínseca dos nós:
 - Influência multiplicativa;
 - Influência aditiva – importante em modelos que permitam que vértices sejam criados sem arestas.
- Envelhecimento de vértices: atratividade de um nó depende também de sua idade.

Métodos de Crescimento

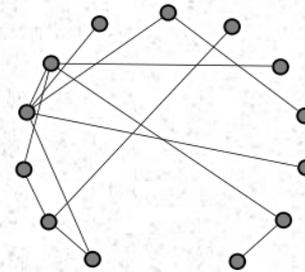
Exemplos de redes onde há o envelhecimento de nós. No primeiro caso o nó ganha atratividade, no segundo caso é o caso usual e nos outros o nó vai perdendo atratividade.



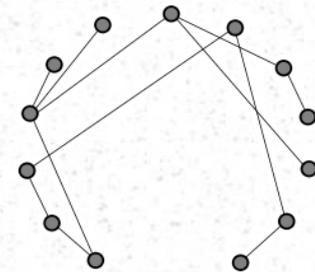
$\alpha = -10.0$



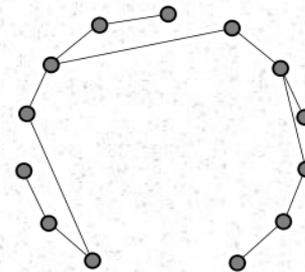
$\alpha = 0.0$



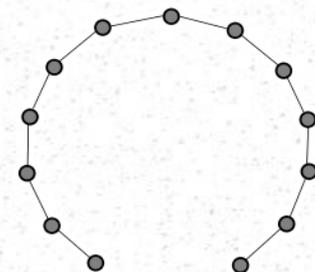
$\alpha = 0.5$



$\alpha = 1.0$



$\alpha = 2.0$



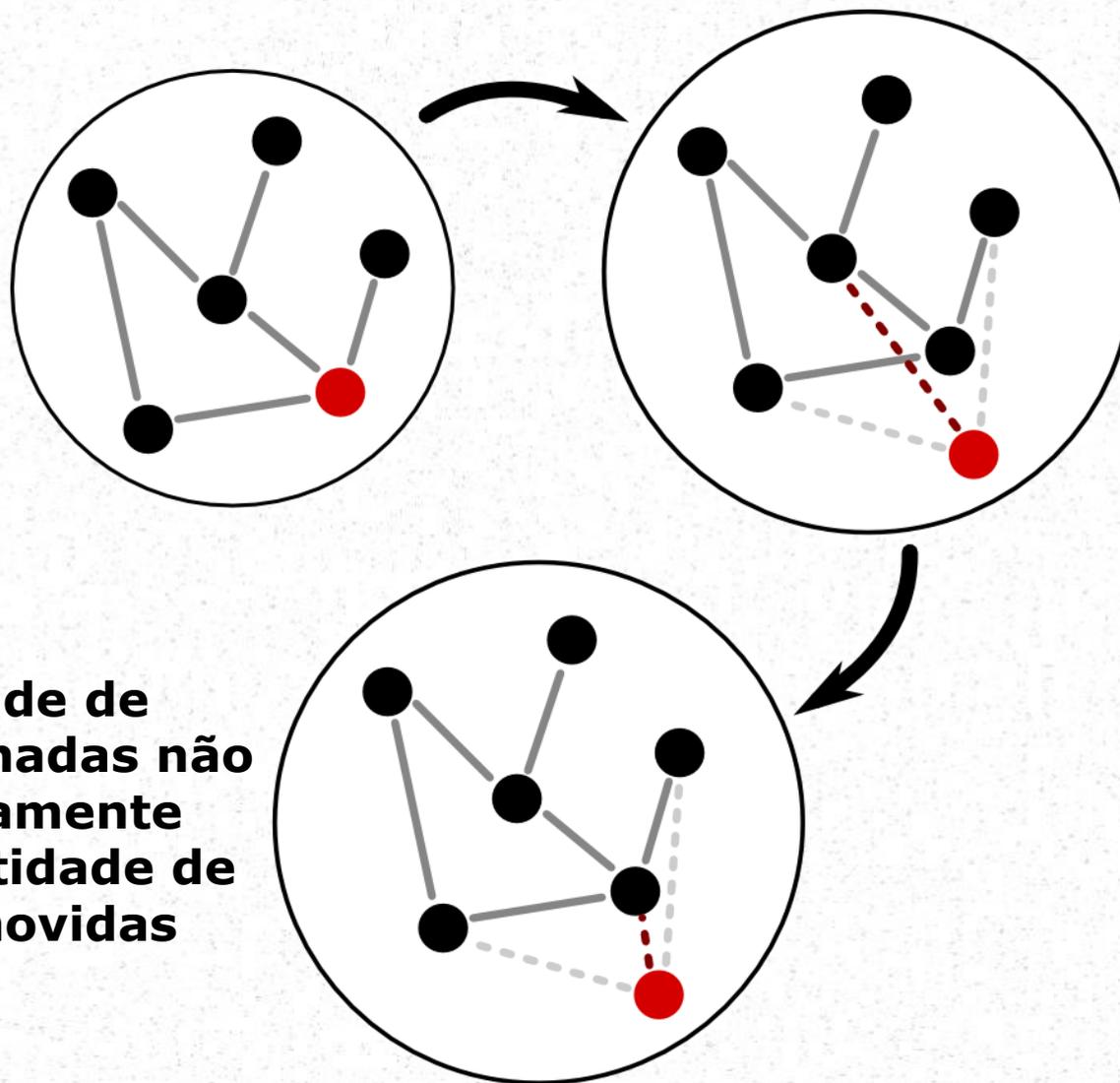
$\alpha = 10.0$

Métodos de Crescimento

Modelo de Crescimento Proteômico

- Pode ser dividido nas seguintes etapas:
 - Início: rede é iniciada com um conjunto pré-existente de vértices e arestas;
 - Crescimento: a cada passo um nó é duplicado;
 - Remoção: as arestas do novo nó são removidas com probabilidade δ ;
 - Adição: novas arestas são criadas a partir do novo nó com probabilidade α .
- Utiliza a ideia de *preferential attachment* de forma indireta: quanto maior o grau de um nó, maior a chance de ter uma aresta duplicada.

Métodos de Crescimento



A quantidade de arestas adicionadas não é necessariamente ligada à quantidade de arestas removidas

Métodos de Crescimento

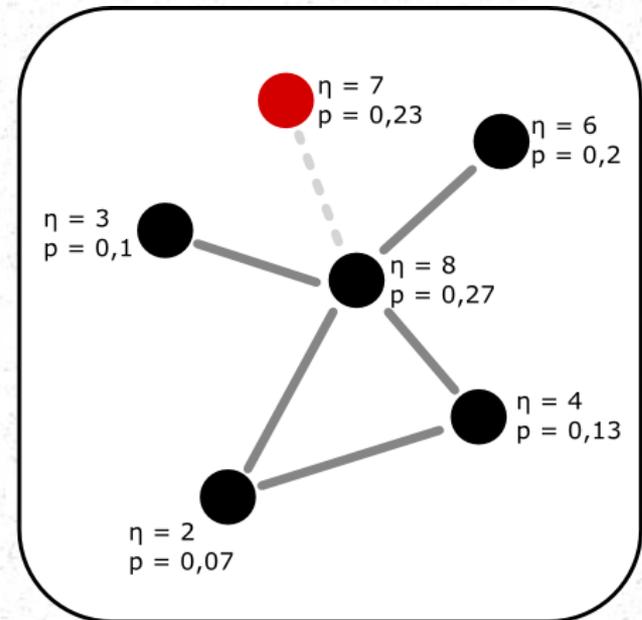
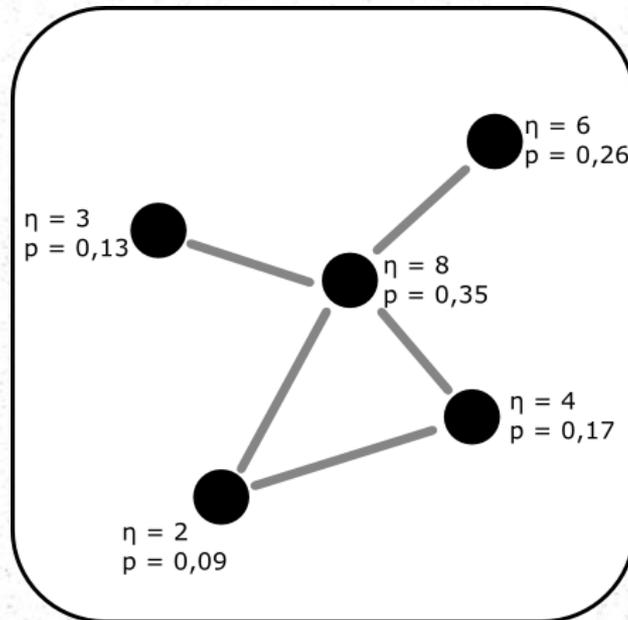
Modelo "good gets richer"

- Cada nó possui uma *fitness* η_i , sendo que a cada aresta criada os nós relacionados são escolhidos com probabilidade proporcional a esse *fitness*, segundo a seguinte equação:

$$p(i) = \frac{\eta_i}{\sum \eta_j}$$

- Apenas algumas distribuições de *fitness* levarão a uma rede livre de escala: uma distribuição de *fitness* em lei da potência é uma forma trivial de obter redes com características livres de escala

Métodos de Crescimento



Métodos de Crescimento

Modelo *Connecting Nearest Neighbors* (CNN)

- Modelo baseado em redes sociais: os nós têm maior probabilidade de conectar-se aos nós que se relacionam com seus vizinhos.
- O modelo pode ser descrito da seguinte forma:
 - As arestas podem ser conectadas por uma aresta real ou por uma aresta potencial;
 - Dois nós são conectados por arestas potenciais se não há arestas reais entre os nós, mas há um vizinho em comum com o qual os dois nós se conectem por uma aresta real.

Métodos de Crescimento

Modelo *Connecting Nearest Neighbors* (CNN)

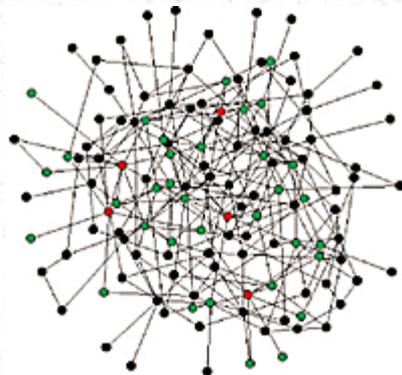
- Passo a passo, o modelo pode ser explicado da seguinte forma:
 - A rede começa com um único vértice;
 - A cada passo, deve-se criar um novo vértice com probabilidade $1 - u$, sendo esse vértice ligado a um outro vértice escolhido aleatoriamente;
 - Com probabilidade u uma aresta potencial é convertida em uma aresta real.

Métodos de Crescimento

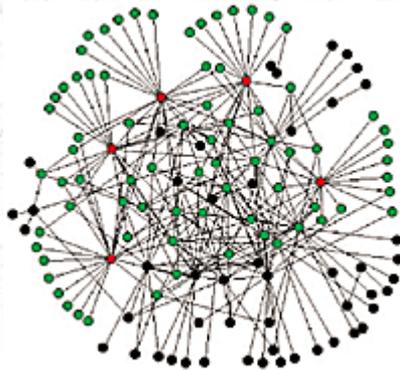
Modelo *Connecting Nearest Neighbors* (CNN)

- Redes geradas por esse modelo apresentam alto grau de clusterização.
- É possível verificar a ocorrência de *preferential attachment* nesse modelo: nós com mais vizinhos diretos (a uma aresta de distância) costumam possuir mais vizinhos a duas arestas de distância.
- O valor do expoente γ da lei de potência é determinado pelo valor de u .

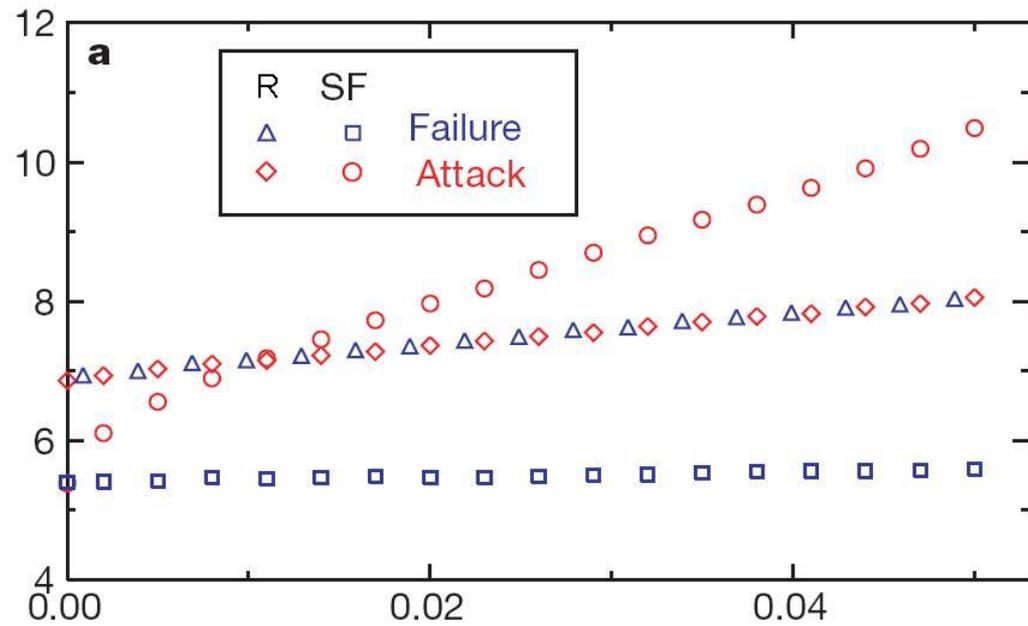
Tolerância a Falhas



Random



Scale-Free



Estrutura/função e Metadinâmica

- Em termos práticos, a representação do conhecimento empregando modelos de redes complexas permite que propriedades estruturais das redes determinem funcionalidades do sistema.
- Quando as redes apresentam metadinâmica, ou seja, quando nós e conexões são criados e eliminados ao longo do tempo, a emergência de certos tipos de estrutura (topologia da rede) é um resultado de fenômenos de auto-organização.

Estrutura/função e Metadinâmica

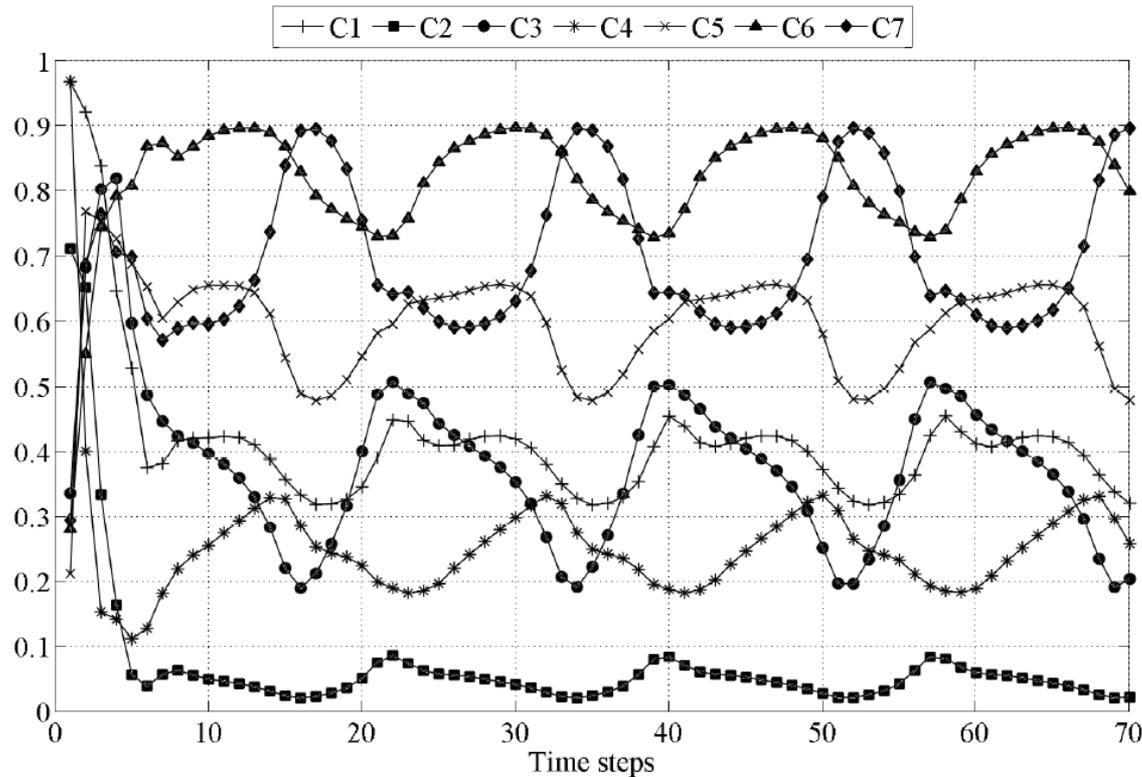
- Propriedades estruturais de redes complexas podem ser obtidas a partir do cômputo de índices como (Costa et al. 2007):
 - Distribuição de grau dos nós;
 - Grau de agrupamento ou formação de comunidades;
 - Grau de intermediação;
 - Grau de modularidade;
 - Grau de conectividade.

Outros modelos em rede

- Modelos em rede também podem ser sintetizados visando reproduzir simultaneamente o comportamento de múltiplas variáveis vinculadas a processos complexos, que podem ser descritos como sistemas dinâmicos acoplados, onde os padrões de acoplamento não são conhecidos a priori.
- Dados históricos (ou séries temporais) obtidos a partir do monitoramento das variáveis deste sistema são tomados como a informação disponível para se inferir um modelo em rede que explica o acoplamento entre as variáveis e permite prever a evolução temporal simultânea dessas variáveis.

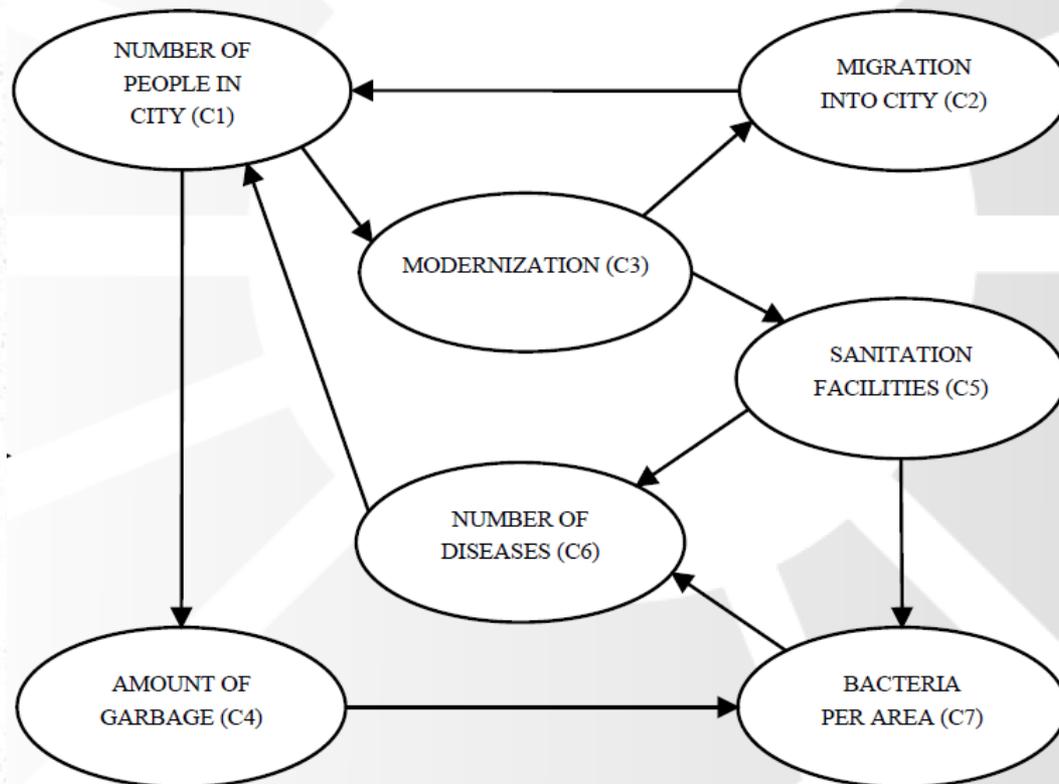
Outros modelos em rede

- Dados históricos de 7 variáveis, obtidos ao se monitorar um sistema complexo.



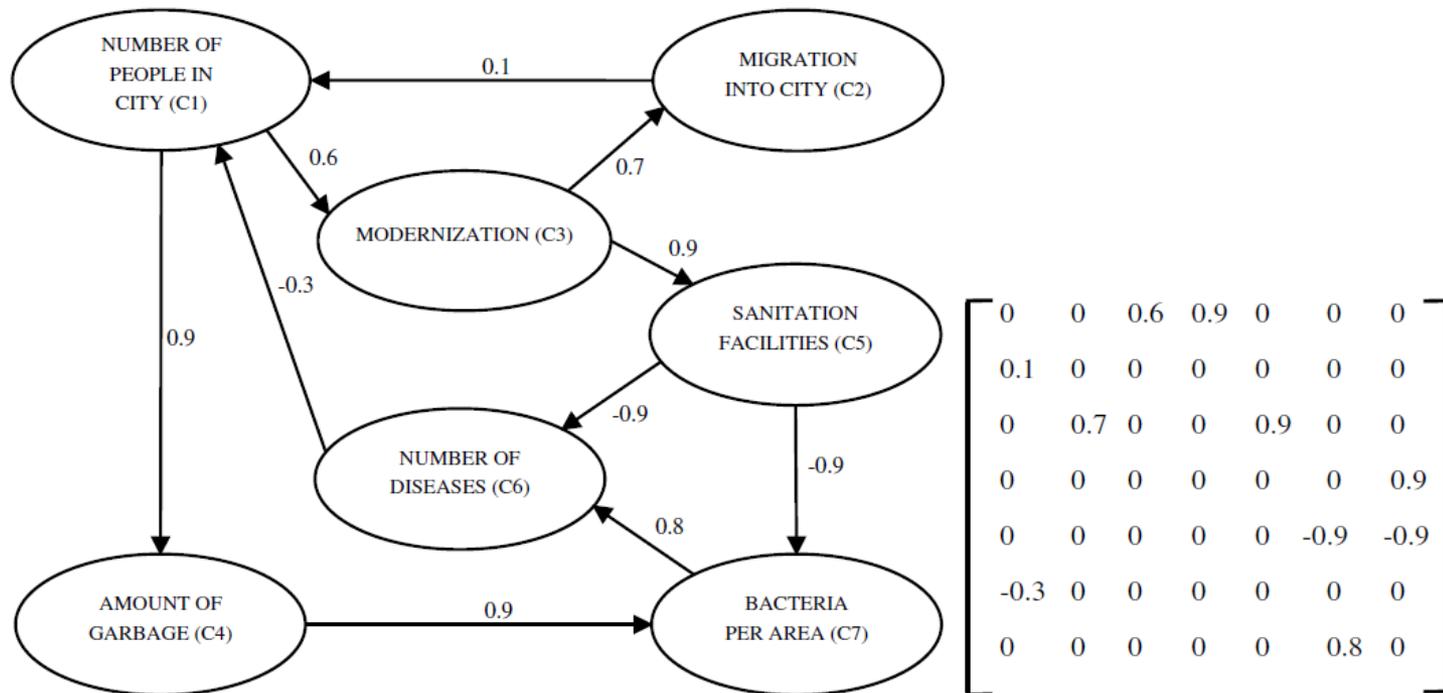
Outros modelos em rede

- Rede causal (grafo direcionado) capaz de explicar a evolução temporal simultânea das 7 variáveis.



Outros modelos em rede

- O modelo não é apenas topológico, pois é necessário indicar o grau de acoplamento entre os nós da rede, representável numa matriz de pesos W .



Outros modelos em rede

- Supondo dinâmica discreta e fornecida a topologia da rede, toma-se um vetor $A(t)$ cujos elementos correspondem ao estado de cada variável do sistema, no instante t , e o próximo estado do sistema pode ser dado na forma:

$$\hat{A}(t+1) = f[\hat{W} * \hat{A}(t)]$$

com f podendo ser linear ou não-linear. O problema de otimização resultante é dado como segue:

$$\begin{array}{ll} \underset{\hat{W}}{\text{minimize}} & J(\hat{W}) \\ \text{subject to} & -1 \leq w_{ij} \leq 1, i, j = 1, \dots, N, \end{array}$$

$$J(\hat{W}) = \frac{1}{T} \sum_{t=1}^T \mathcal{E}(t),$$

$$\mathcal{E}(t) = \frac{1}{2N} \|A(t) - \hat{A}(t)\|^2,$$

onde T é o horizonte de tempo considerado.

Outros modelos em rede

- A solução, portanto, está associada à necessidade de proposição de uma topologia para a rede de acoplamentos e um grau de acoplamento para cada aresta da rede proposta.
- Trata-se de um problema de otimização extremamente difícil de resolver.
- Algumas áreas envolvidas:
 - **Mapas cognitivos fuzzy** (Papageorgiou, E. Learning algorithms for fuzzy cognitive maps – A review study, IEEE Transactions on Systems, Man, and Cybernetics, Part C, 42(2):150–163, 2012.)
 - **Redes causais** (Pearl, J. Causality: Models, reasoning, and inference, Cambridge University Press, 2000) (Glymour, C. & Cooper, G. (Eds.) Computation, causation, and discovery, MIT Press, 1999)

Outros modelos em rede

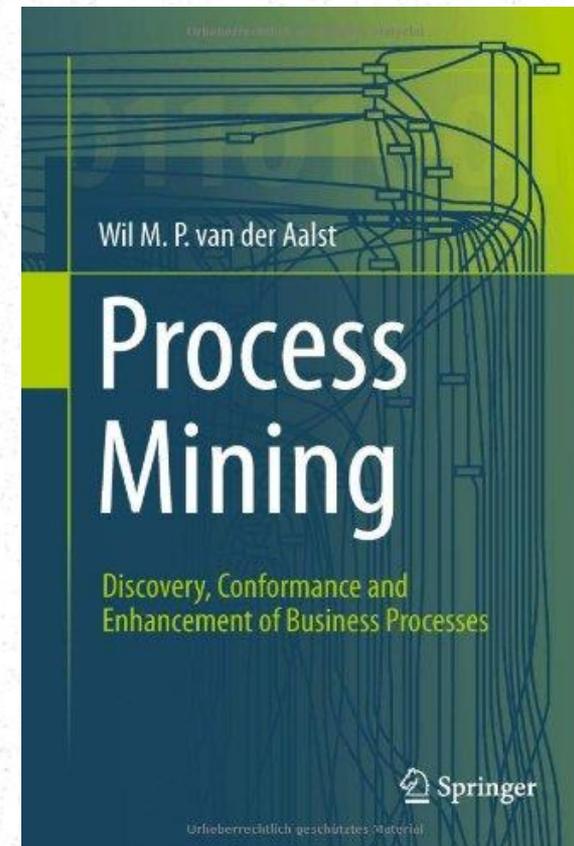
- A realização de intervenções no processo, com o subsequente monitoramento dos seus efeitos, pode auxiliar na obtenção de redes causais mais realísticas (Steyvers, M.; Tenenbaum, J.B.; Wagenmakers, E.-J.; Blum, B. Inferring causal networks from observations and interventions, *Cognitive Science*, vol. 27, pp. 453-489, 2003.).
- No contexto de mapas cognitivos fuzzy, propostas híbridas envolvendo algoritmos evolutivos e buscas locais envolvendo técnicas de gradiente (especificamente para a definição dos campos não-nulos da matriz de pesos) têm produzido bons resultados (Madeiro, S.S. & Von Zuben, F.J. Gradient-based Algorithms for the Automatic Construction of Fuzzy Cognitive Maps. *Proceedings of the 11th International Conference on Machine Learning and Applications*, pp. 344-349, 2012.).

Mineração de processos (Process mining)

- Técnicas de mineração de processos permitem extrair informações de logs de eventos. Por exemplo, o monitoramento de um sistema de gerenciamento de fluxo de trabalho ou os logs de transação de um sistema de planejamento de recursos empresariais podem ser usados para descobrir modelos que descrevem processos, organizações e produtos.
- Mineração de processos está diretamente relacionada com BAM (Business Activity Monitoring), BOM (Business Operations Management), BPI (Business Process Intelligence) e data/workflow mining.
- Mas diferente das técnicas clássicas de mineração de dados, o foco está no processo em si e nas relações entre seus componentes, ou seja, acabam sempre resultando **modelos em rede**.

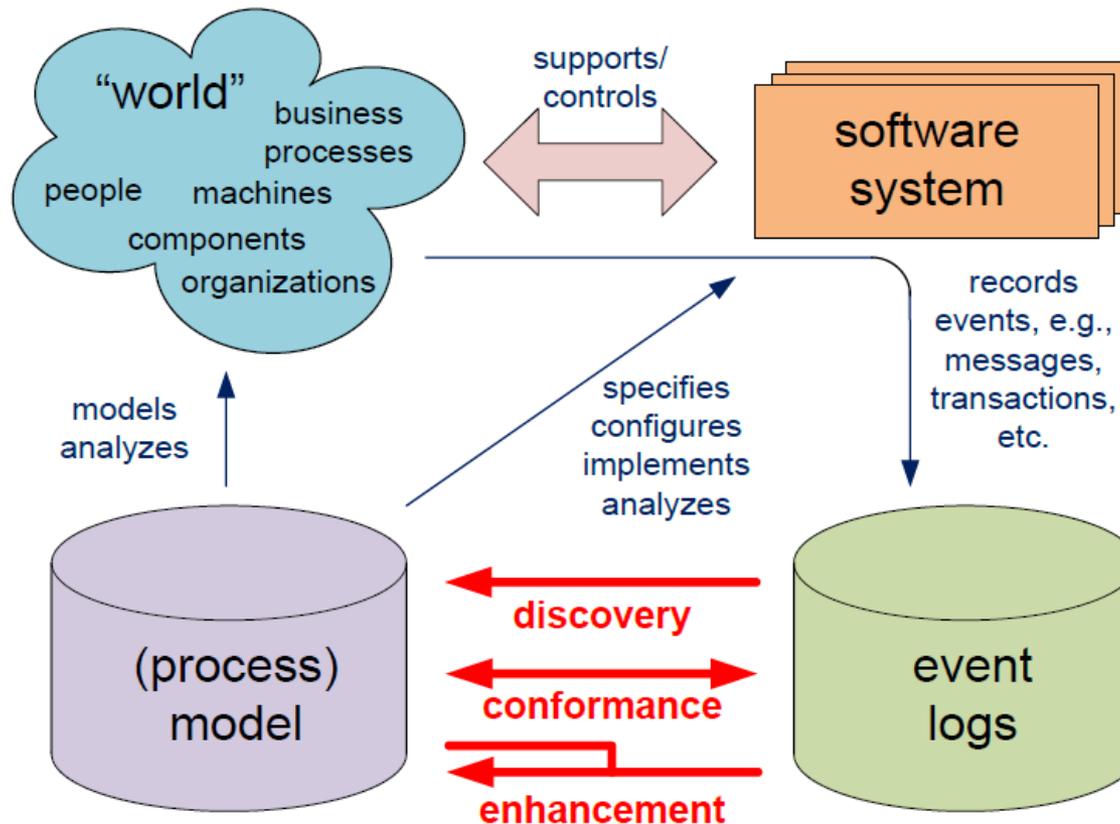
Mineração de processos (Process mining)

- Mineração de processos é uma disciplina emergente baseada em modelos guiados por processos e mineração de dados. Com essa técnica, as corporações não só podem se beneficiar plenamente das informações armazenadas em seus sistemas, mas também podem verificar a conformidade dos processos, detectar pontos de estrangulamento e prever problemas de execução.
- O primeiro livro-texto na área é: van der Aalst, W.M.P. Process Mining - Discovery, Conformance and Enhancement of Business Processes, Springer, 2011.



Mineração de processos (Process mining)

The three main types of process mining:
discovery, conformance, and enhancement



ProM

- <http://www.processmining.org/prom/start>
- ProM (which is short for Process Mining framework) is an Open Source framework for process mining algorithms. ProM provides a platform to users and developers of the process mining algorithms that is easy to use and easy to extend. Our mission is to become the de facto standard process mining platform in the academic world by establishing an active, recognized community of contributors and users, and to create awareness for the power of process mining technology by promoting applications and industrial uptake. Our vision is to actively advance the state-of-the-art of process mining technology by developing methods that really work, by creating an open community, and by providing a stable and easily extensible platform, which optimally supports process mining.

Bibliografia

- Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D.U. Complex networks: Structure and dynamics, *Physics Reports*, 424(4):175-308, 2006.
- Costa, L.D.F., Rodrigues, F.A., Travieso, G. & Villas Boas, P.R. Characterization of complex networks: A survey of measurements, *Advances in Physics*, 56(1): 167-242, 2007.
- Dorogovtsev, S.N. & Mendes, J.F.F. *Evolution of Networks: From Biological Nets to the Internet and WWW*, Oxford University Press , 2003.
- Erdős, P. & Rényi, A. On random graphs. *Publicationes Mathematicae (Debrecen)*, 6:290–297, 1959.
- Newman, M.E.J. Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E* 64(1):016131, 2001a.

Bibliografia

- Newman, M.E.J. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E*, 64(1):016132, 2001b.
- Newman, M.E.J. The structure and function of complex networks, *SIAM Review*, vol. 45, pp. 167-256, 2003.
- Milgram, S. The small world problem. *Psychology Today*, 1:61-67, 1967.
- Sole, R.V., Satorras, P.R., Smith, E. & Kepler, T.B. A model of large-scale proteome evolution. *Advances in Complex Systems*, 5(1):43-54, 2002.
- Watts, D.J. & Strogatz, S.H. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440-442, 1998.

Leitura Adicional

- Barabási, A.-L. "Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life", Plume Books, 2003.
- Buchanan, M. "Nexus: Small Worlds and the Groundbreaking Theory of Networks", W. W. Norton & Company, 2003.
- Johnson, S. "Emergence: The Connected Lives of Ants, Brains, Cities, and Software", Scribner, 2002.
- Strogatz, S. "Sync: The Emerging Science of Spontaneous Order", Hyperion, 2003.
- Watts, D.J. "Six Degrees: The Science of a Connected Age", W. W. Norton & Company, 2003.
- Watts, D.J. "Small Worlds: The Dynamics of Networks between Order and Randomness", Princeton Studies in Complexity, Princeton University Press, 2003.