

# REDES COMPLEXAS: UMA NOVA FERRAMENTA NA ANÁLISE DE SÉRIES TEMPORAIS

Andriana S. L. O. Campanharo, Fernando Manuel Ramos

Programa de Pós-Graduação em Computação Aplicada  
Laboratório Associado de Computação e Matemática Aplicada (LAC)  
Instituto Nacional de Pesquisas Espaciais (INPE)  
Caixa Postal 515 – 12227-010 – São José dos Campos – SP – Brasil

{andriana, fernando}@lac.inpe.br

**Abstract.** *In this work we construct complex networks from different types of time series (e.g. random, periodic and chaotic time series), where each point is represented by a node in the associated network and the connection between nodes is determined by the correlation between the points of the time series. We investigate the statistical properties of these networks, such as the clustering coefficient and the average path length and found that time series with different dynamics exhibit distinct topological structures. These results show the complex network theory can be a powerful tool in the time series analysis.*

## 1. Introdução

A teoria das redes complexas nasceu da aplicação de medidas desenvolvidas pela teoria dos grafos e conceitos provenientes da mecânica estatística, física não-linear e sistemas complexos [Barabási and Albert 1999, Watts and Strogatz 1998]. Redes complexas são descritas por um conjunto de vértices (nós) e arestas (conexões, ligações ou *links*) e algum tipo de interação entre os mesmos. De modo geral, redes complexas possuem estruturas que são irregulares, complexas e que evoluem dinamicamente com o tempo.

Devido a sua generalidade e caráter multidisciplinar, a teoria de redes complexas abrange aplicações nas mais diversas áreas de pesquisa, como física, química, matemática, biologia, medicina, sociologia, engenharia, telecomunicações, astronomia, etc., sendo a computação responsável pelas ferramentas utilizadas na modelagem, simulação e tratamento das bases de dados [Boccaletti et al. 2006].

Recentemente, a teoria de redes complexas foi utilizada em uma aplicação, até então, inédita: na análise da dinâmica de séries temporais. A caracterização da dinâmica de um sistema físico a partir da análise de séries temporais é um tema de grande relevância em diferentes áreas da ciência e das engenharias. Por este motivo, várias “métricas” foram propostas na literatura, como, por exemplo, entropias, dimensões (e.g., dimensão de correlação) e expoentes (e.g. expoente de Lyapunov). Porém, tais técnicas possuem algumas limitações que, em alguns casos, comprometem a análise da dinâmica de séries temporais. Algumas destas limitações são: séries temporais muito longas ou muito curtas; presença de ruído; não estacionaridade; etc. Assim, a caracterização da dinâmica de séries temporais é um desafio constante, e conseqüentemente, novos métodos têm sido propostos no meio científico.

Zhang & Small [2006] e Lacasa *et al.* [2008] mostraram que, sob certas condições, é possível mapear a estrutura harmônica de uma série temporal na topologia de nós e co-

nexões de uma rede complexa, a qual pode ser então caracterizada com as ferramentas estatísticas usuais. O principal resultado obtido em ambos os estudos é que séries temporais com dinâmicas diferentes exibem estruturas topológicas de rede distintas.

A teoria de redes complexas promete ser uma ferramenta promissora na classificação de sistemas dinâmicos. Contudo, os estudos nesta área estão em sua fase inicial e o processo de mapeamento “ideal” é ainda um problema essencial a ser resolvido [Yang and Yang 2008].

## 2. Propriedades topológicas

Alguns dos conceitos que ocupam um lugar proeminente no estudo de redes complexas serão discutidos a seguir, e ainda, tais medidas serão aplicadas nas redes construídas na Seção 5, visando uma melhor caracterização das mesmas.

### 2.1. Grau de conectividade

Por definição, o grau de conectividade de um vértice  $i$ , em uma rede não-dirigida, é o número de suas conexões diretas a outros vértices, ou seja:

$$k_i = \sum_{j=1}^N A_{ij} \quad (1)$$

onde  $A$  é a matriz de adjacência associada e  $N$  é o número de vértices da rede. O grau de conectividade médio é definido como a média aritmética do grau de conectividade sobre todos os vértices, ou seja:

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_i \quad (2)$$

A caracterização topológica mais básica de um grafo  $G$  pode ser obtida em termos da distribuição  $P(k)$ , definida como a probabilidade que um vértice escolhido aleatoriamente tenha grau de conectividade  $k$ . Em redes reais, a distribuição  $P(k)$  desvia-se significativamente da distribuição de Poisson<sup>1</sup> esperada para um grafo aleatório e, em muitos casos, exibe característica de uma lei de potência com um expoente  $\gamma$  entre 2 e 3 [Boccaletti et al. 2006]. Redes com estas características são chamadas de redes *sem escala*.

### 2.2. Coeficiente de agrupamento

Também conhecido como transitividade, o agrupamento é uma propriedade típica de redes de conhecimento, onde dois indivíduos com um amigo em comum são também conhecidos um do outro [Wasserman and Faust 1994]. Esta tendência inerente ao agrupamento é quantificada pelo coeficiente de agrupamento [Watts and Strogatz 1998].

Suponha que um vértice  $i$  tenha  $k_i$  vizinhos, logo  $k_i(k_i - 1)/2$  ligações podem existir entre eles (isto ocorre quando todos os vizinhos de  $i$  estão conectados a todos os outros vizinhos de  $i$ ). A razão entre o número  $E_i$  de ligações que realmente existem

---

<sup>1</sup>A distribuição de Poisson é um tipo de distribuição discreta de probabilidade, cuja forma analítica é dada por  $P(k) = \frac{\lambda^k e^{-\lambda}}{k!}$ , onde  $\lambda$  é o parâmetro da distribuição.

pelo número total de ligações possíveis fornece o valor do coeficiente de agrupamento do vértice  $i$ :

$$C_i = \frac{2E_i}{k_i(k_i - 1)} \quad (3)$$

Definimos o coeficiente de agrupamento da rede como a média dos  $C_i$  para todo  $i$ . Pela relação (3) podemos observar que o coeficiente de agrupamento assume valores pertencentes ao intervalo  $[0, 1]$ .

### 2.3. Comprimento do caminho

O comprimento do caminho médio  $L$  é definido como o número de ligações no menor caminho entre dois vértices, medidos sobre todos os pares de vértices da rede. Existem vários métodos numéricos que calculam o menor caminho entre quaisquer pares de vértices em um grafo, tais como os algoritmos de Dijkstra, Floyd-Warshall, Bellman-Ford, etc. Neste trabalho o valor de  $L$  foi obtido com base no algoritmo de Floyd-Warshall.

### 3. Séries Temporais Utilizadas

O processo de mapeamento proposto na Seção 4 foi aplicado em três conjuntos distintos de séries temporais. O primeiro conjunto se refere à uma série aleatória obtida a partir de uma distribuição uniforme entre  $[0, 1]$ . O segundo conjunto se refere à uma série puramente periódica, obtida através da relação:

$$y_i = \sin(2\pi\omega i) \quad (4)$$

com  $i = 1, 2, \dots, n$ . E, finalmente o terceiro tipo se refere a série temporal oriunda da evolução temporal da variável de estado  $x$  do sistema de equações diferenciais ordinárias de Lorenz [Lorenz 1963]

$$\begin{aligned} dx/dt &= \sigma(-x + y) \\ dy/dt &= rx - y - xz \\ dz/dt &= -b + xy \end{aligned} \quad (5)$$

com  $\sigma = 10$ ,  $r = 28$  e  $b = 8/3$ . Neste caso, a série temporal obtida é denominada caótica.

### 4. Processo de Mapeamento Proposto

Neste trabalho, propomos um mapeamento inédito “série temporal - rede complexa” baseado na função de autocorrelação da série temporal em estudo, definida como:

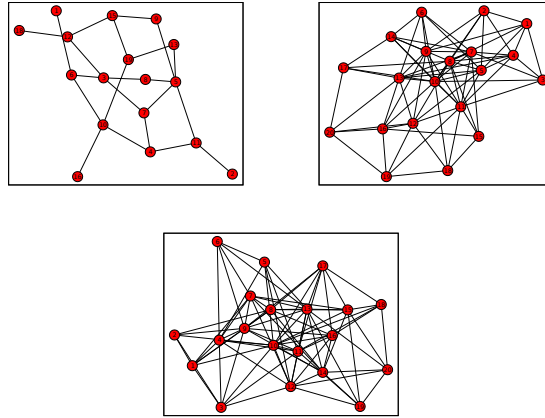
$$A(\tau) = \frac{1}{N - \tau} \sum_{n=1}^{N-\tau} \frac{(x_n - \bar{x})(x_{n+\tau} - \bar{x})}{\sigma^2} \quad (6)$$

onde  $x_i$  com  $i \in 1, \dots, N$  são os valores assumidos pela série temporal,  $N$  é o seu número de pontos,  $\bar{x}$  é a sua média e  $\sigma^2$  é a sua variância.  $\tau$  é uma escala que assume valores inteiros e positivos entre 0 e  $N - 1$  e  $A(\tau) \in [-1, 1]$ .

Dada uma série temporal, cada um de seus pontos corresponde um nó da rede associada. No processo de mapeamento proposto, o valor absoluto da função de autocorrelação (6) expressa a probabilidade de conexão para cada par de nós da rede. Assim, dois pontos de uma série temporal  $x(n)$  e  $x(n + \tau)$  que apresentam uma correlação elevada (em módulo) terão maior probabilidade de terem seus nós correspondentes conectados e vice-versa.

## 5. Resultados

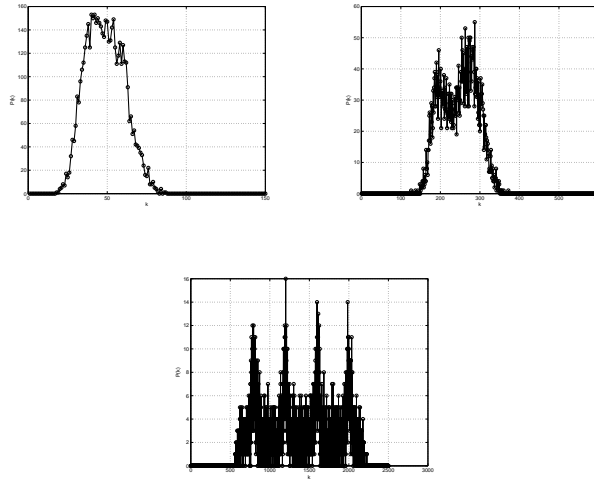
O mapeamento apresentado na Seção 4 foi aplicado nos diferentes tipos de séries temporais em estudo, e desta forma, foram obtidas redes cujas características foram investigadas com base no grau de conectividade e nos valores de  $C$  e de  $L$ . A Figura 1 mostra exemplos de redes obtidas a partir do mapeamento proposto, para cada uma das séries em estudo.



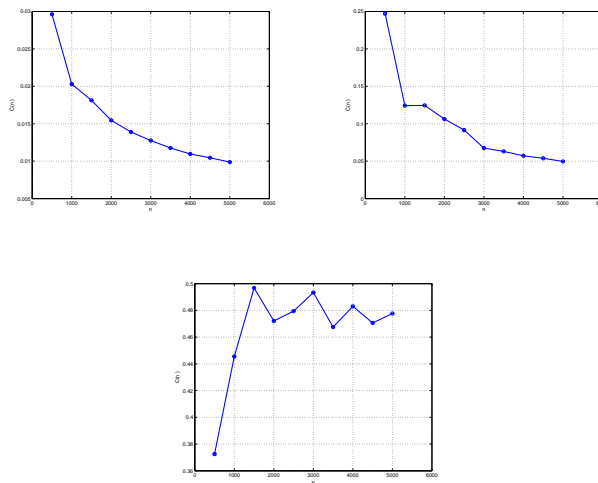
**Figura 1. Exemplos de redes geradas a partir das séries temporais aleatória, caótica e periódica, respectivamente. Para a construção de tais redes foram consideradas séries temporais com 20 pontos cada e 20 nós.**

A Figura 2 mostra as distribuições dos graus de conectividade associados a cada uma destas séries. Podemos observar pela distribuição do grau de conectividade da rede associada à série aleatória, que a maior parte de seus nós possuem o mesmo grau de conectividade. Nós com grau de conectividade superior ou inferior à média ocorrem em número reduzido. O mesmo comportamento pode ser observado na distribuição do grau de conectividade da rede associada à série caótica. A distribuição do grau de conectividade da rede associada à série periódica apresenta um padrão periódico.

Em seguida, calculamos os valores médios de  $C$  em função do número de nós  $n$  (conforme mostra a Figura 3) com base em dez amostras de cada uma das séries em estudo. De maneira similar às análises anteriores, consideramos redes com número de nós variáveis e, desta forma, investigamos o comportamento de  $C$  em função das dimensões das redes. Para a série aleatória, as redes são fracamente agrupadas e o valor de  $C$  decresce em função do número de nós  $n$ . Para a série caótica, as redes obtidas são também fracamente agrupadas. Para a série periódica, podemos observar que com o incremento de  $n$  as redes obtidas vão se tornando mais agrupadas.



**Figura 2.** Distribuição do grau de conectividade  $P(k)$  versus  $k$  associado às séries aleatória, caótica e periódica, respectivamente. Em todos os casos foram consideradas séries temporais com 5000 pontos e, desta forma, 5000 nós em cada rede.

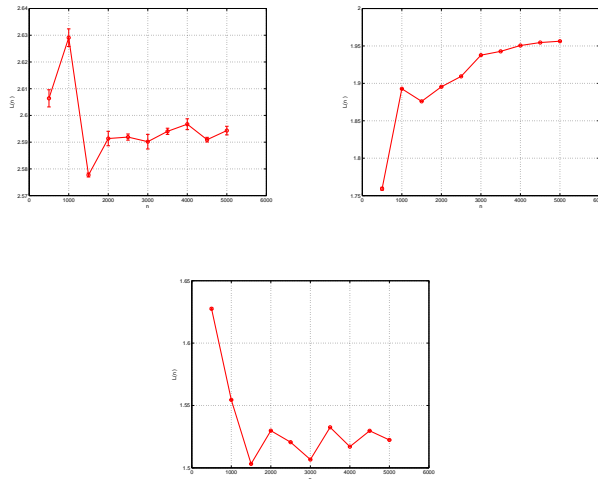


**Figura 3.** Valores médios de  $C$  em função do número do número de nós  $n$  associados às séries aleatória, caótica e periódica, respectivamente. Em todos os casos foram consideradas séries temporais com número de pontos variáveis, ou seja, entre 500 e 5000 pontos.

Finalmente, calculamos os valores de  $L$  em função do número de nós  $n$  (conforme mostra a Figura 4) com base em dez amostras de cada uma das séries em estudo. Para a série aleatória, o valor de  $L$  é pequeno para qualquer valor de  $n$ . Para a série caótica, o valor inicial de  $L$  é pequeno, porém o mesmo cresce com o incremento de  $n$ . Para a série periódica, podemos observar que o valor inicial de  $L$  é pequeno e que este valor diminui com o incremento de  $m$ .

## 6. Conclusões

Pelo processo de codificação proposto podemos observar que séries aleatórias produzem redes aleatórias em virtude dos baixos valores de  $C$  e de  $L$ . Observamos ainda, que



**Figura 4. Valores médios de  $L$  em função do número de nós  $n$  associados as séries aleatória, caótica e periódica, respectivamente. Em todos os casos foram consideradas séries temporais com número de pontos variáveis, ou seja, entre 500 e 5000 pontos.**

séries periódicas produzem redes *pequeno mundo* em virtude do valor elevado de  $C$  e do baixo valor de  $L$ . Contudo, as redes obtidas a partir das séries caóticas apresentaram características específicas, o que impossibilitou classificá-las em padrões de redes usuais, tais como redes aleatórias, pequeno mundo ou livres de escala.

Os processos de mapeamentos sugeridos neste trabalho ainda estão em fase exploratória, e desta forma, serão aprimorados futuramente. Além do estudo de diferentes estratégias de correspondência “série temporal - rede complexa”, buscamos ainda compreender o papel das características dos dados (tamanho da série, resolução temporal, etc.), do ruído, e da dinâmica subjacente.

## Referências

- Barabási, A. L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 289.
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D. U. (2006). Complex networks: *Physics Reports*, 424.
- Lacasa, L., Luque, B., Ballesteros, F., Luque, J., and Nuño, J. C. (2008). From time series to complex networks: The visibility graph. *Proceedings of the National Academy of Sciences*, 105.
- Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of Atmospheric Sciences*, 20.
- Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press, New York.
- Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of “small world” networks. *Letters to Nature*, 393.
- Yang, Y. and Yang, H. (2008). Complex network-based time series analysis. *Physica A*, 387.
- Zhang, J. and Small, M. (2006). Complex network from pseudoperiodic time series: Topology versus dynamics. *Physical Review Letters*, 96.