

Comparando algoritmos para conversão de séries temporais em grafos

Vinicius de Paula Silva¹

Departamento de Computação, Universidade Federal de Ouro Preto, Brasil

Leonardo N. Ferreira²

Max Planck Institute for Human Development, Alemanha

Leonardo B. L. Santos³

Centro Nacional de Monitoramento e Alertas de Desastres Naturais, Brasil

Vander L. S. Freitas⁴

Departamento de Computação, Universidade Federal de Ouro Preto, Brasil

A constante evolução do estudo sobre redes complexas possibilitou seu uso em áreas como biologia, medicina e neurociência [1]. A partir do mapeamento do sistema, e conseqüentemente do espaço topológico gerado, pode-se utilizar métricas de rede para sua caracterização. Um exemplo de aplicação é a análise de séries temporais via o formalismo de redes [4].

O presente estudo investiga algoritmos para conversão de séries temporais em redes e avalia quais deles melhor capturam a incidência de ruído entre as séries. Os algoritmos escolhidos são: *Dynamical Characterization with the Top Integral Function* (DCTIF) [2], *Dynamical Characterization with Symbolic Dynamics* (DCSD) [2] e *Visibility Graph* (VG) [4].

Para avaliar a capacidade dos algoritmos em capturar características das séries, propõe-se o seguinte experimento: gera-se séries temporais sintéticas e aplica-se perturbações a elas. A hipótese inicial é que as redes geradas a partir da série e suas correspondentes perturbadas sejam diferentes. Ainda, essa diferença aumenta conforme a intensidade da perturbação cresce. A comparação das redes é feita a partir de diferentes algoritmos encontrados em [5].

Para efeito de comparação, utiliza-se também técnicas estatísticas tradicionais já consolidadas que comparam diretamente as séries temporais geradas: *Dynamic Time Warping* [3], Informação Mútua e Correlação de Pearson, que quantificam a similaridade/distância entre duas séries.

Como estudo de caso utiliza-se séries temporais geradas a partir da seguinte função: $f(t) = \frac{\sin(t)+5}{10} + \xi(t)$, com t variando de 0 a 30, com incremento $\Delta t = 0.1$, isto é, considerando-se 300 pontos. $\xi(t)$ é uma fonte de ruído com distribuição uniforme, de intensidade σ (Figura 1). Gera-se sempre uma série temporal com $\sigma = 0$ (sem ruído) e outra com $\sigma > 0$. Quantifica-se as diferenças entre as séries usando as técnicas estatísticas tradicionais e; também as redes geradas por elas, por meio dos algoritmos DCTIF, DCSD e VG, usando diferentes formas de comparar as redes (Portrait Divergence, NetLSD, GCD-11). Foram realizados testes com $\sigma \in [0.001, 0.4]$. Por conta da natureza estocástica do problema, 1000 séries são geradas para cada σ e o resultado médio é computado.

¹vinicius.silva1@aluno.ufop.edu.br

²ferreira@leonardonascimento.com

³santoslbl@gmail.com

⁴vander.freitas@ufop.edu.br

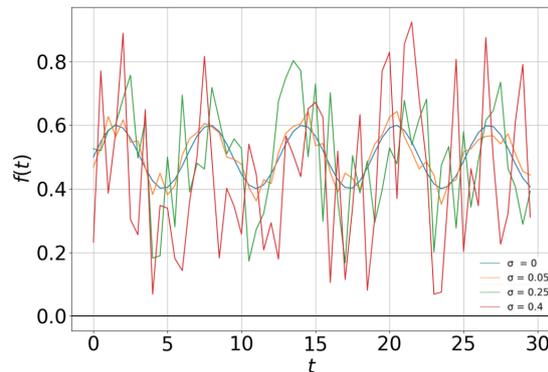


Figura 1: Séries temporais geradas com diferentes intensidades σ .

O algoritmo baseado em redes que melhor conseguiu captar diferenças entre as séries foi o VG, e o algoritmo de comparação de redes NetLSD detectou maior variação entre as redes geradas, com as diferenças aumentando logaritmicamente com σ . O algoritmo DCTIF utilizado em conjunto com o NetLSD não foi capaz de reconhecer diferenças entre as séries a partir de $\sigma \approx 0.1$. Já o DCSD com o Portrait Divergence não identifica as mudanças nas séries a partir de $\sigma \approx 0.05$. Com relação às técnicas estatísticas tradicionais, todas conseguem capturar as mudanças, em especial o DTW, que apresenta um crescimento linear. Em suma, tanto no algoritmo VG, em conjunto com o NetLSD, quanto na correlação de Pearson e na Informação Mútua, as diferenças entre as séries originais e as ruidosas aumentam abruptamente quando a intensidade do ruído é baixa, e depois mais lentamente. A partir de uma dada intensidade ($\sigma > \sigma_c$), as três abordagens supracitadas capturam o mesmo nível de similaridade, isto é, sem discernir um σ do outro.

Agradecimentos

Os autores agradecem ao apoio da Universidade Federal de Ouro Preto e a Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig), processo 6.32/2021.

Referências

- [1] Albert-László Barabási e Márton Pósfai. **Network science**. Cambridge: Cambridge University Press, 2016.
- [2] Vander L. S. Freitas, Juliana C. Lacerda e Elbert E. N. Macau. “Complex Networks Approach for Dynamical Characterization of Nonlinear Systems”. Em: **International Journal of Bifurcation and Chaos** 29.13 (2019), p. 1950188. DOI: 10.1142/S0218127419501888.
- [3] Toni Giorgino. “Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package”. Em: **Journal of Statistical Software** 31.7 (2009), pp. 1–24. DOI: 10.18637/jss.v031.i07.
- [4] Lucas Lacasa et al. “From time series to complex networks: The visibility graph”. Em: **Proceedings of the National Academy of Sciences** 105.13 (2008), pp. 4972–4975. DOI: 10.1073/pnas.0709247105.
- [5] Mattia Tantardini et al. “Comparing methods for comparing networks”. Em: **Scientific Reports** 9.1 (nov. de 2019), p. 17557. ISSN: 2045-2322. DOI: 10.1038/s41598-019-53708-y. URL: <https://doi.org/10.1038/s41598-019-53708-y>.