

## ANÁLISE ESTATÍSTICA DAS REDES DE CORRUPÇÃO BRASILEIRA

Alvaro Franco Martins (PIBIC/CNPq), Haroldo Valentin Ribeiro (Orientador),  
e-mail: [alvarought@gmail.com](mailto:alvarought@gmail.com)

Universidade Estadual de Maringá / Centro de Ciências Exatas/Maringá, PR.

### FÍSICA/FÍSICA ESTATÍSTICA E TERMODINÂMICA

**Palavras-chave:** Corrupção, Sistemas Complexos, Redes Complexas

#### Resumo:

Analizamos escândalos de corrupção brasileiros que foram bem documentados ao longo dos últimos 28 anos. Inicialmente, o presente trabalho envolveu a coleta desses dados através da linguagem de programação *Python* e sua biblioteca *Scrapy*, útil para extração automatizada de dados pela internet. Após essa coleta de dados, começou-se a análise da rede formada pelas pessoas envolvidas nos escândalos de corrupção. Esse estudo foi feito por meio de técnicas de Física Estatística de Sistemas Complexos, observando diversos aspectos dessa rede. Por fim, foi aplicado um algoritmo de predição de possíveis ligações futuras entre indivíduos não conectados na rede, na qual uma taxa de aproximadamente 25% de precisão foi obtida.

#### Introdução

Nosso projeto aborda o problema social da corrupção do ponto de vista dos sistemas complexos, utilizando-se de ferramentas de redes complexas. Uma rede complexa é composta por um conjunto de objetos (os vértices) e um conjunto de ligações entre esses objetos (os links). A palavra complexo refere-se à característica de emergência desses sistemas, na qual à medida que dados e a interação aumentam, esse sistema começa a adquirir características distintas daquelas observadas isoladamente. Exemplos de redes complexas são bastante diversos e incluem teias alimentares, redes sociais, redes de aeroportos, ou até mesmo colaborações científicas. Nossa rede é formada pelos agentes envolvidos/mencionados em corrupção, na qual cada vértice é uma pessoa e as ligações entre elas ocorrem quando estão envolvidas no mesmo escândalo. Essa rede é bastante intrincada e apresenta diversas propriedades interessantes associadas a sua dinâmica de crescimento.

Nos focamos em entender a dinâmica dessa rede e como novas ligações poderiam ser preditas ao analisar sua evolução. Além disso, neste projeto,

fizemos inicialmente um estudo mais detalhado da biblioteca *Scrapy*, a qual foi utilizada para fazer uma coleta automatizada de dados. Esses dados relacionados aos escândalos de corrupção ocorridos no Brasil foram coletados a partir de páginas da internet de revistas de notícias e jornais brasileiros publicamente acessíveis. Os dados reunidos abrangem 28 anos de corrupção no Brasil (1987-2014) referentes a 65 casos bem noticiados, contendo mais de 400 nomes de pessoas envolvidas com corrupção.

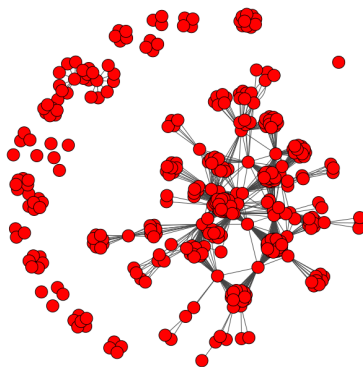
Após a obtenção, estudamos esses dados utilizando conceitos e métodos de redes complexas e também de séries temporais. No contexto de redes complexas, inicialmente, fizemos uma abordagem introdutória de sistemas complexos com o objetivo de estudar e analisar os dados obtidos. Foi possível ilustrar comportamentos e padrões das informações obtidas. Um exemplo desses comportamentos está associado com a forma pela qual a maior componente da rede de corrupção brasileira evolui com o tempo, observando crescimentos repentinos nos anos associados às mudanças importantes de poder político no país, isto é, anos de eleição.

## Materiais e métodos

O presente trabalho foi desenvolvido por meio da linguagem de programação *Python*, no ambiente *Jupyter Notebook*, utilizando biblioteca *Scrapy*, a qual permite uma obtenção massiva e automatizada de dados pela internet. A abordagem desses dados incluiu a verificação detalhada do que era verídico e o que constava em domínio público sobre cada escândalo e seus envolvidos. Após feito isso, a análise se tornou mais sistemática a fim de estudar as possíveis relações mais intrínsecas do sistema.

## Resultados e Discussão

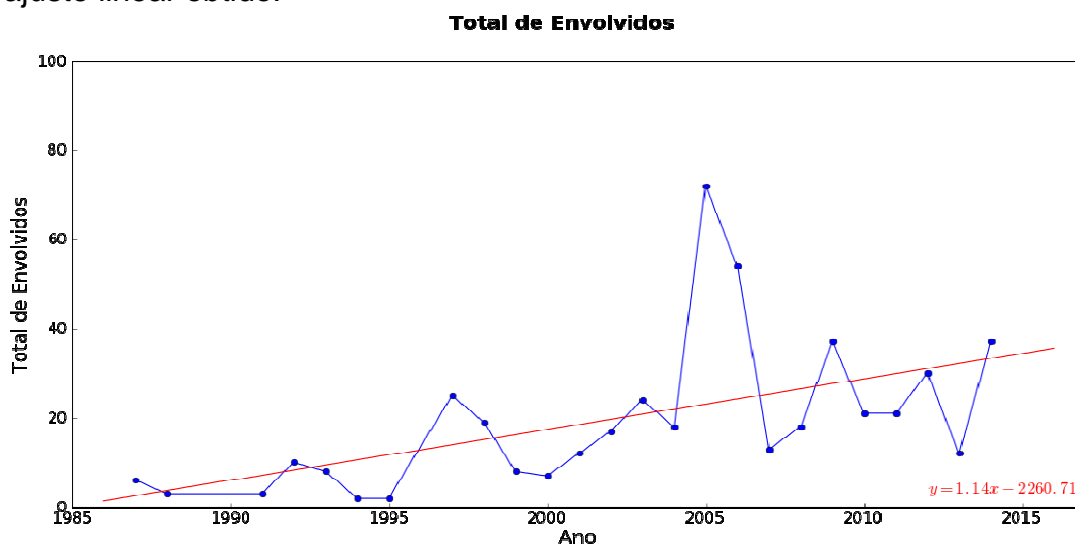
A Figura 1 ilustra a forma geral da rede de corrupção brasileira obtida a partir de nossos dados.



**Figura 1** – Forma geral da rede de corrupção brasileira.

Uma característica notória dessa rede indica que a corrupção ocorre em pequenos grupos que raramente são compostos por mais de oito pessoas.

Além disso, estudamos como a maior componente da rede de corrupção brasileira evolui com o tempo, observando crescimentos repentinos nos anos associados às mudanças importantes de poder político, ou seja, anos de eleição. A Figura 2 ilustra os dados empíricos, listando o número de envolvidos em escândalos de corrupção em cada ano analisado. No contexto de séries temporais, uma análise simples que realizamos foi a investigação da evolução do número de pessoas envolvidas com corrupção ao longo dos anos. Nesse caso, estudamos alguns dos métodos de regressão linear. A Figura 2 mostra esse número em função do tempo e o ajuste linear obtido.



**Figura 2** – Número de envolvidos em casos de corrupção por ano. Podemos notar que de quatro em quatro anos, os dados apresentam um máximo e depois um mínimo.

Apesar do comportamento complexo dos dados, podemos concluir que existe uma leve tendência de aumento no número de pessoas envolvidas com corrupção ao longo do tempo. De fato, nossa análise de regressão indica que existe uma tendência estatisticamente significativa de aumento de aproximadamente uma pessoa por ano.

## Conclusões

Ao estudar a evolução da rede de corrupção, mostramos que sua distribuição de grau exibe duas variações abruptas em anos que estão associadas às mudanças importantes no poder político que rege o Brasil. Além disso, também conseguimos mostrar que algoritmos para prever ligações podem acertar links futuros entre indivíduos da rede. Baseada na evolução temporal da rede, nossos resultados demonstram uma chance em torno de 25% de acerto para a predição de links futuros. A análise desses

dados rendeu a publicação de um artigo na revista *Journal of Complex Networks*.

## Agradecimentos

Agradeço ao CNPq pelo apoio financeiro.

## Referências

Stéfan van der Walt, S. Chris Colbert and Gaël Varoquaux. **The NumPy Array: A Structure for Efficient Numerical Computation**, Computing in Science & Engineering, 13, 22-30 (2011), DOI:10.1109/MCSE.2011.37.

Jones E, Oliphant E, Peterson P, et al. **SciPy: Open Source Scientific Tools for Python**, 2001-, Disponível em: <<https://scipy.org/>> Acesso em: 18 jul. 2018.

### **Web Scraping in Python using Scrapy (with multiple examples)**

Disponível em: <<https://www.analyticsvidhya.com/blog/2017/07/web-scraping-in-python-using-scrapy/>> Acesso em: 18 jul. 2018.

Wes McKinney. **Data Structures for Statistical Computing in Python**, Proceedings of the 9th Python in Science Conference, 51-56 (2010).

Haroldo V. Ribeiro, Luiz G. A. Alves, Alvaro F. Martins, Ervin K. Lenzi, Matjaz Perc. **The dynamical structure of political corruption networks**. Journal of Complex Networks 1, 15 (2018).

World Bank. **Anti-corruption** (2016). Disponível em: <<http://www.worldbank.org/en/topic/governance/brief/anti-corruption>> Acesso em: 18 jul. 2018.

Transparency International. **Global corruption report** (2009). Disponível em: <[http://files.transparency.org/content/download/107/431/file/2009\\_GCR\\_EN.pdf](http://files.transparency.org/content/download/107/431/file/2009_GCR_EN.pdf)> Acesso em: 18 jul. 2018.

Banfield, E. C. **Corruption as a feature of governmental organization**. In Here the People Rule, 147–170 (Springer, 1985).