Abstract

Corruptive behaviour in politics limits economic growth, embezzles public funds and promotes socioeconomic inequality in modern democracies. We analyze well-documented political corruption scandals in Brazil over the past 27 years, focusing on the dynamical structure of networks where two individuals are connected if they were involved in the same scandal. Our research reveals that corruption runs in small groups that rarely comprise more than eight people, in networks that have hubs and a modular structure that encompasses more than one corruption scandal. We observe abrupt changes in the size of the largest connected component and in the degree distribution, which are due to the coalescence of different modules when new scandals come to light or when governments change. We show further that the dynamical structure of political corruption networks can be used for successfully predicting partners in future scandals. We discuss the important role of network science in detecting and mitigating political corruption.

Figure 1: Demography and evolving behaviour of corruption scandals in Brazil. A) The number of people involved in each corruption scandal in chronological order (from 1987 to 2014). B) Cumulative probability distribution (on a log-linear scale) of the number of people involved in each corruption scandal (red circles). C) Time series of the number of people involved in corruption scandals by year (red circles). The alternating gray shades indicate the term of each general election that took place in Brazil between 1987 and 2017. D) Autocorrelation function of the time series of the yearly number of people involved in scandals (red circles).

Figure 2: Complex network representation of people involved in corruption scandals. Complex network of people involved in all corruption cases in our dataset (from 1987 to 2014). Each vertex represents a person and the edges among them occur when two individuals appear (at least once) in the same corruption scandal. Node sizes are proportional to their degrees and the colour code refers to the modular structure of the network. There are 27 significant modules, and 14 of them are within the giant component (indicated by the red dashed loop).

Figure 3: Characterization of nodes based on the within-module degree ($Z$) and participation coefficient ($P$). Each dot in the $Z$-$P$ plane corresponds to a person in the network and the different shaded regions indicate the different roles according to the network cartography proposed by Guimerà and Amaral. The majority of nodes (97.5%) are classified as ultra peripheral (R1) or peripheral (R2), and the remaining are modular connection (R3, three nodes), provincial hubs (R4, three nodes), and connector hubs (R6, two nodes).

Figure 4: The vertex degree distribution is exponential, invariant over time, and the characteristic degree exhibits abrupt changes over the years.

Figure 5: Changes in the size of the largest component of the corruption network over time are caused by a coalescence of network modules.

Figure 6: Predicting missing links between people in the corruption network may be useful for investigating and mitigating political corruption. We tested the predictive power of eleven methods for predicting missing links in the corruption networks. These methods are based on local similarity measures (degree product, association strength, cosine, Jaccard, resource allocation, Adams-Azar, and common neighbors), global path- and random walk-based similarity measures (rooted PageRank and SimRank), and on the hierarchical structure of networks (hierarchical random graph – HRV). To assess the accuracy of these methods, we applied each algorithm to snapshots of the corruption network in a given year (excluding 2014), ranked the top-10 predictions, and verified whether these predictions appear in future snapshots of the network. The bar plot shows the fraction of correct predictions for each method. We also included the predictions of a random model where missing links are predicted by chance (error bars are 95% bootstrap confidence intervals).