

# Universality of political corruption networks

Alvaro F. Martins<sup>1</sup>, Bruno R. da Cunha<sup>2,3</sup>, Quentin S. Hanley<sup>4</sup>, Sebastián Gonçalves<sup>5</sup>, Matjaž Perc<sup>6,7,8,9,\*</sup>, and Haroldo V. Ribeiro<sup>1</sup>

<sup>1</sup>Departamento de Física, Universidade Estadual de Maringá – Maringá, PR 87020-900, Brazil

<sup>2</sup>Rio Grande do Sul Superintendency, Brazilian Federal Police – Porto Alegre, RS 90160-093, Brazil

<sup>3</sup>National Police Academy, Brazilian Federal Police – Brasília, DF 71559-900, Brazil

<sup>4</sup>School of Science and Technology, Nottingham Trent University, Clifton Lane, Nottingham NG11 8NS, United Kingdom

<sup>5</sup>Instituto de Física, Universidade Federal do Rio Grande do Sul – Porto Alegre, RS 91501-970, Brazil

<sup>6</sup>Faculty of Natural Sciences and Mathematics, University of Maribor, Koroška cesta 160, 2000 Maribor, Slovenia

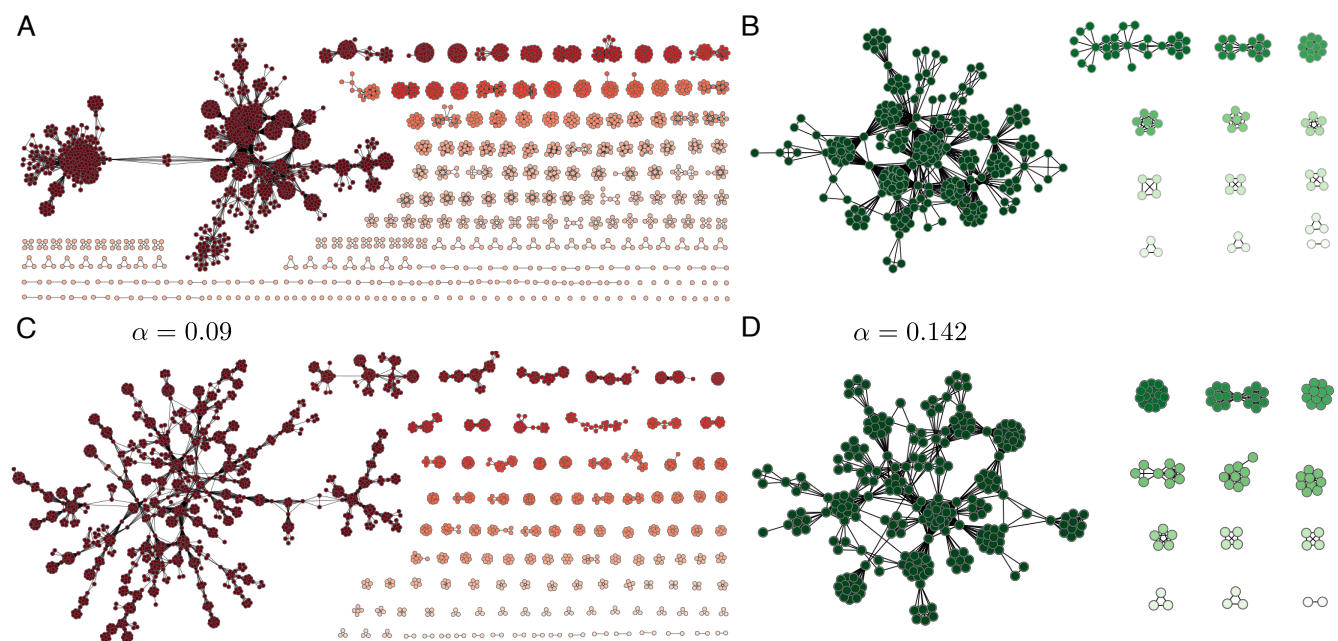
<sup>7</sup>Department of Medical Research, China Medical University Hospital, China Medical University, Taichung, Taiwan

<sup>8</sup>Alma Mater Europaea, Slovenska ulica 17, 2000 Maribor, Slovenia

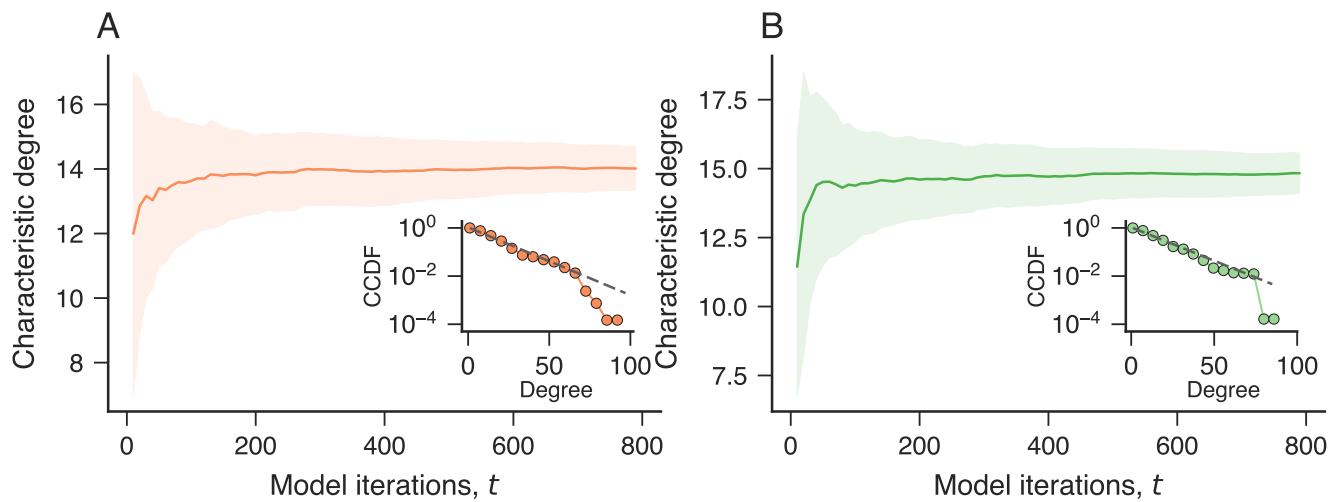
<sup>9</sup>Complexity Science Hub Vienna, Josefstädterstraße 39, 1080 Vienna, Austria

\*email: matjaz.perc@gmail.com

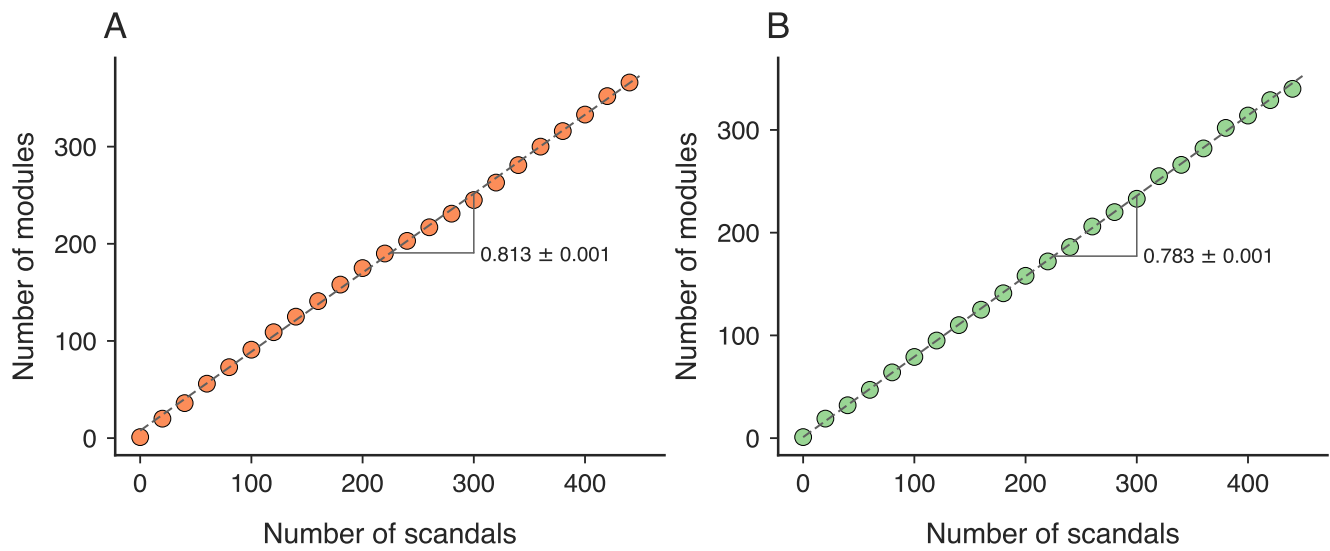
## Supplemental Materials



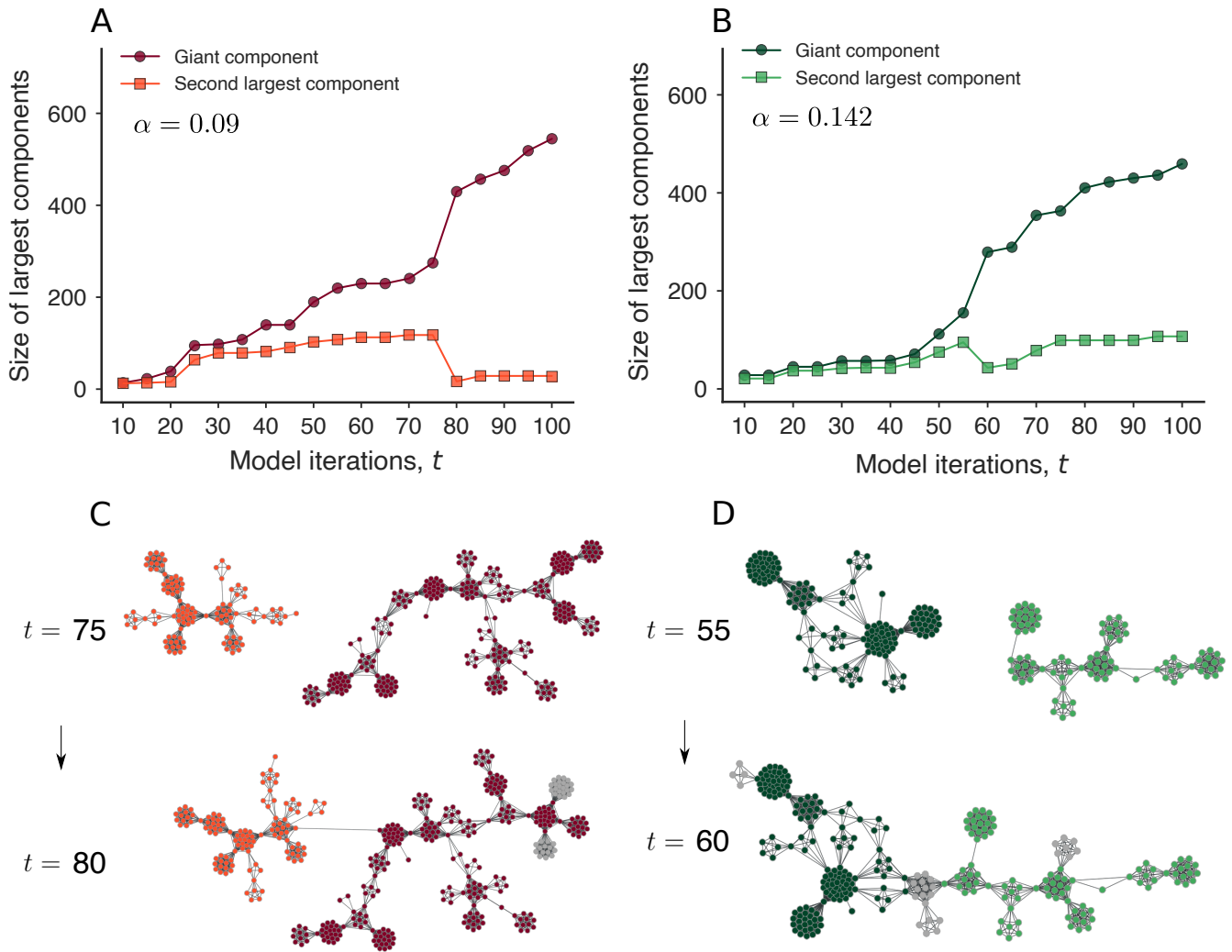
**Figure S1.** Visual comparison between simulated and empirical corruption networks. The Spanish and Brazilian corruption networks (panels A and B, respectively) compared to typical networks generated by our model using the empirical values of the recidivism rate of Spain ( $\alpha = 0.09$ , panel C) and Brazil ( $\alpha = 0.14$ , panel D). Both simulated networks have been generated using the exact number of political scandals of the empirical data. Node color stands for network components.



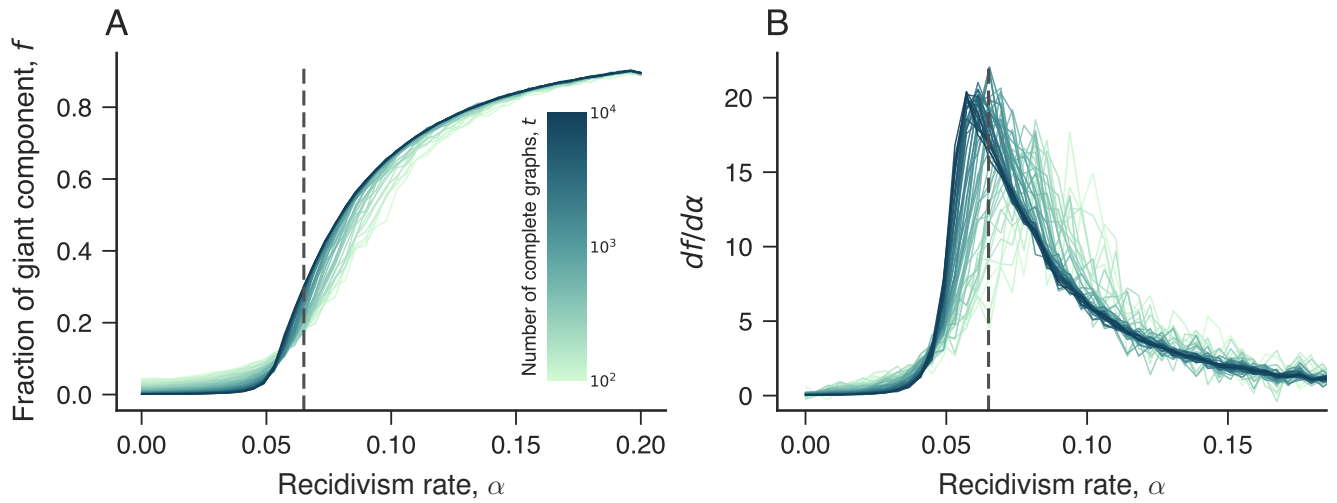
**Figure S2.** Our model replicates the exponential degree distribution and the approach to a constant value of the characteristic degree observed in empirical networks. (A) Average value of the maximum-likelihood estimate of the characteristic degree obtained from one hundred simulations using the Spanish recidivism rate ( $\alpha = 0.09$ ) as a function of the number of model iterations  $t$  (that is, number of scandals added to the network). The shaded region stands for 95% bootstrap confidence intervals. The inset shows the complementary cumulative degree distribution after  $t = 800$  model iterations, where the dashed line indicates the exponential distribution adjusted to data. Panel (B) shows the same quantities for simulations using the Brazilian recidivism rate ( $\alpha = 0.14$ ).



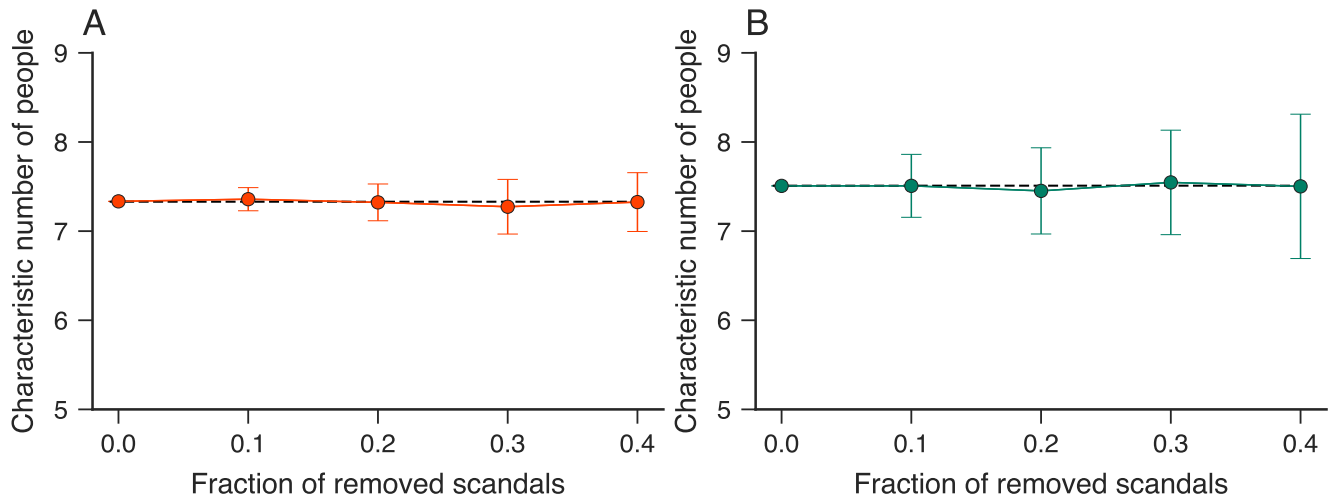
**Figure S3.** Our model replicates the linear association between the number of network modules and the number of scandals observed during the growth of empirical networks. Panels (A) and (B) show the average association between the number of network modules and the total of scandals estimated from one hundred simulations with the Spanish and Brazilian recidivism rates, respectively. In both panels, the dashed lines represent linear models adjusted to data and the values within these panels indicate the average ratio between network modules and scandals during the growth of simulated networks.



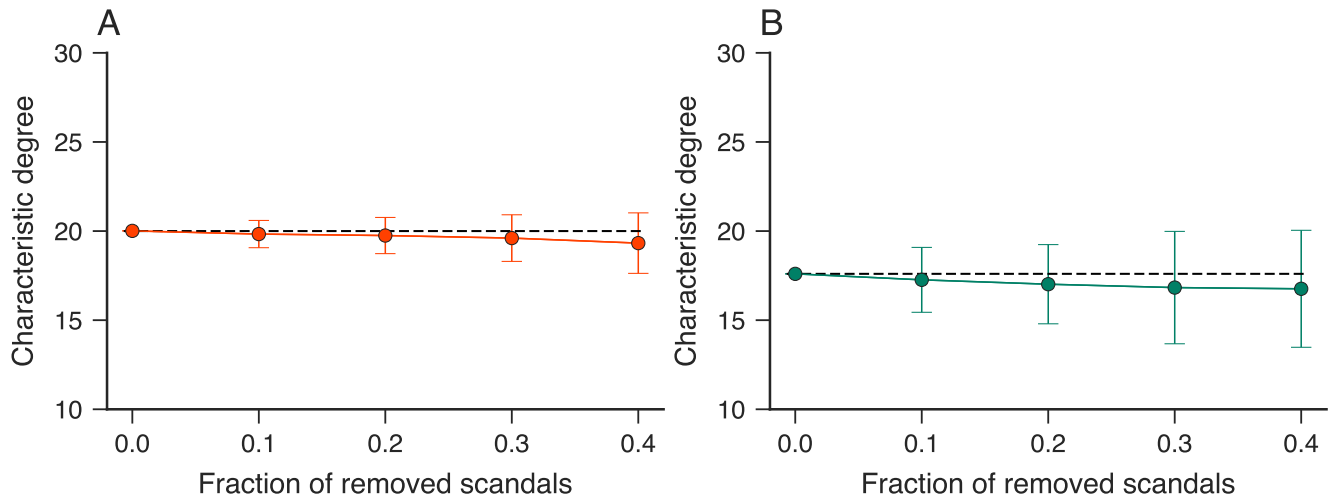
**Figure S4.** Our model replicates the coalescence of network components observed during the growth of empirical corruption networks. Panels (A) and (B) depict the evolution of the size of the giant component (circle) and second-largest component (square) of typical simulated networks using the Spanish and Brazilian recidivism rates, respectively. Similarly to the empirical networks, we observe that the simulated behavior exhibits abrupt changes caused by the emergence of new scandals involving recidivist agents. Panels (C) and (D) show snapshot visualizations of simulated networks before and after the abrupt changes observed in panels (A) and (B), respectively. In these snapshots, new nodes are colored in gray.



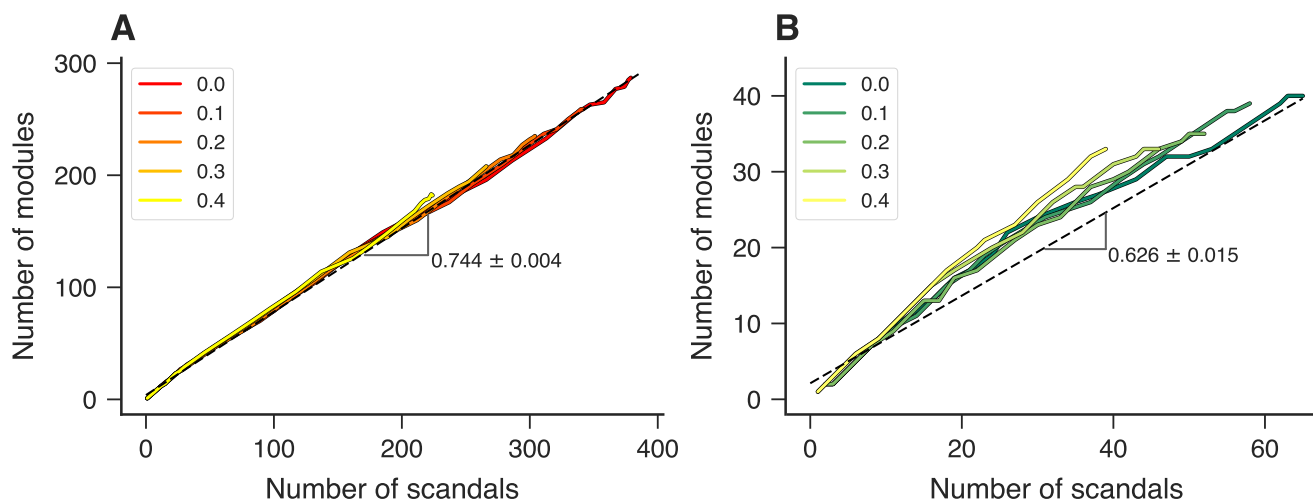
**Figure S5.** Model behavior near the criticality for different network sizes. (A) Average fraction of the giant component of simulated networks  $f$  as a function of the recidivism rate  $\alpha$ . (B) The derivative of  $f$  with respect to  $\alpha$ . In both panels, the different curves represent the average model behavior over one hundred realizations (with  $s_c = 7$ ,  $\beta = 12$ , and  $p = 0.025$ ) and after including  $t$  complete graphs (the value of  $t$  is indicated by the color code). The dashed vertical line indicates the critical recidivism rate. We observe that the transition does not become more abrupt when increasing the network size, suggesting that our model displays a continuous (second-order) phase transition.



**Figure S6.** Robustness of the size distribution of scandals against randomly removing a fraction of scandals from our datasets. Estimates of the characteristic number of people involved in political corruption scandals in (A) Spain and (B) Brazil as a function of the fraction of removed scandals. The dashed lines indicate the values when considering all scandals in our datasets. The markers represent average values estimated from one hundred realizations of the randomly removing process, and the error bars stand for one standard deviation of these values. In both cases, the characteristic number of people remains very stable even after removing 40% of the scandals in our datasets.

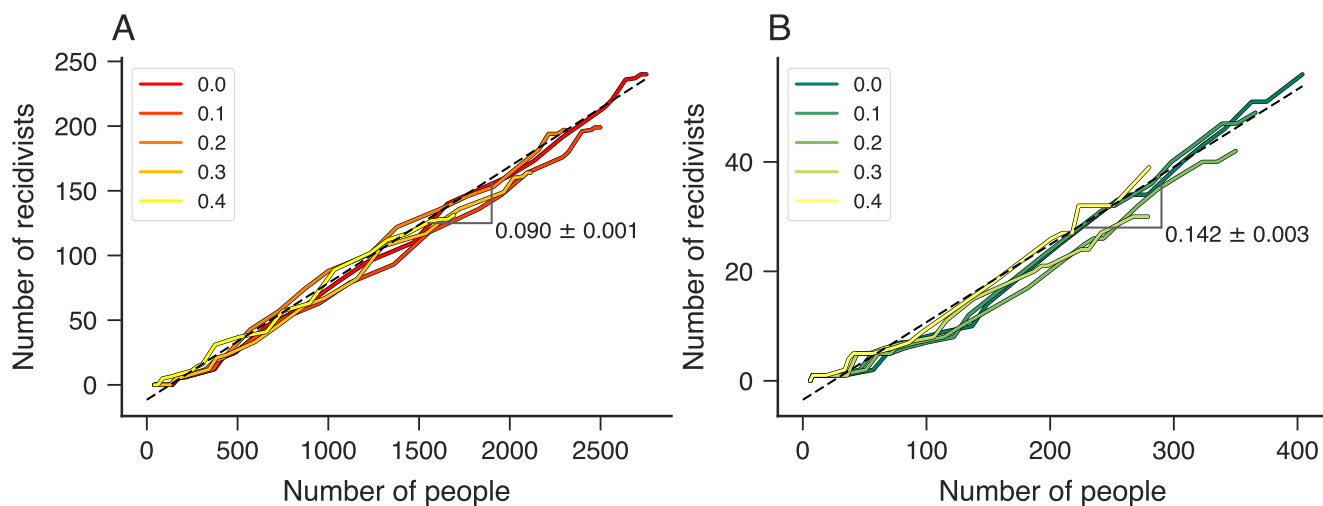


**Figure S7.** Robustness of the degree distribution of corruption networks against randomly removing a fraction of scandals from our datasets. Estimates of characteristic degree for the latest stage of the (A) Spanish and (B) Brazilian corruption networks as a function of the fraction of removed scandals. The dashed lines indicate the values when considering all scandals in our datasets. The markers represent average values estimated from one hundred realizations of the randomly removing process, and the error bars stand for one standard deviation of these values. In both cases, the characteristic degree remains very stable even after removing 40% of the scandals in our datasets.



**Figure S8.** Robustness of the association between the number of network modules and the total of political scandals against randomly removing a fraction of scandals from our datasets. The curves show the relationships between the number of modules and the number of scandals for each year of the (A) Spanish and (B) Brazilian networks under different fractions of removed scandals (indicated by the color code). The dashed lines represent a linear model adjusted to the relationship obtained when considering all scandals in our datasets. We note the linear association remains a good approximation for the empirical behavior even after removing 40% of the scandals from our datasets.





**Figure S9.** Robustness of the association between the number of recidivist agents and the total number of people against randomly removing a fraction of scandals from our datasets. The curves show the relationships between the number of recidivist agents and the total number of people for each year of the (C) Spanish and (D) Brazilian corruption networks under different fractions of removed scandals (indicated by the color code). The dashed lines represent a linear model adjusted to the relationship obtained when considering all scandals in our datasets. We note that linear association remains a good approximation for the empirical behavior even after removing 40% of the scandals in our datasets.