The dynamical structure of political corruption networks

Luiz G. A. Alves

Institute of Mathematics and Computer Sciences
University of São Paulo

May 25, 2018 – São Carlos, SP, Brazil
Complex systems

A complex system is a system composed of interconnected parts that as a whole exhibit one or more properties not obvious from the properties of the individual parts.

“The whole is greater than the sum of its parts.”
Examples of Complex Systems
Brain

No “master-mind”

Self-organization

Evolution

Adaptation
Society

No “master-mind”

Self-organization

Evolution

Adaptation
Behind each system studied in complexity there is an intricate *wiring diagram*, or a network, that defines the interactions between the components.
We will never understand complex system unless we map out and understand the networks behind them.
What is a Complex Network?

Networks are just collections of “points” joined by “lines”
Graph theory

Graph theory: 1735, Euler

Seven bridges of Königsberg in Prussia (now Kaliningrad, Russia)
Erdős–Rényi model - 1959

\[ p = 0 \]
\[ p \cdot n = 0 \]
Watts–Strogatz model - 1998

Regular

Small-world

Random

$p = 0 \quad \text{Increasing randomness} \quad p = 1$

Six degrees of separation
Scale-free networks - 1999

**Bell Curve**
- Most nodes have the same number of links
- No highly connected nodes

**Power Law Distribution**
- Very many nodes with only a few links
- A few hubs with large number of links

![Diagram showing scale-free network characteristics](https://via.placeholder.com/150)
Emergence of Scaling in Random Networks

Albert-László Barabási* and Réka Albert

Department of Physics, University of Notre-Dame, Notre-Dame, IN 46556

Systems as diverse as genetic networks or the world wide web are best described as networks with complex topology. A common property of many large networks is that the vertex connectivities follow a scale-free power-law distribution. This feature is found to be a consequence of the two generic mechanisms that networks expand continuously by the addition of new vertices, and new vertices attach preferentially to already well connected sites. A model based on these two ingredients reproduces the observed stationary scale-free distributions, indicating that the development of large networks is governed by robust self-organizing phenomena that go beyond the particulars of the individual systems.

Scale-free networks are rare

Anna D. Broido\textsuperscript{1,}\textsuperscript{*} and Aaron Clauset\textsuperscript{2,3,4,}\textsuperscript{†}

\textsuperscript{1}Department of Applied Mathematics, University of Colorado, Boulder, CO, USA
\textsuperscript{2}Department of Computer Science, University of Colorado, Boulder, CO, USA
\textsuperscript{3}BioFrontiers Institute, University of Colorado, Boulder, CO, USA
\textsuperscript{4}Santa Fe Institute, Santa Fe, NM, USA

A central claim in modern network science is that real-world networks are typically “scale free,” meaning that the fraction of nodes with degree $k$ follows a power law, decaying like $k^{-\alpha}$, often with $2 < \alpha < 3$. However, empirical evidence for this belief derives from a relatively small number of real-world networks. We test the universality of scale-free structure by applying state-of-the-art statistical tools to a large corpus of nearly 1000 network data sets drawn from social, biological, technological, and informational sources. We fit the power-law model to each degree distribution, test its statistical plausibility, and compare it via a likelihood ratio test to alternative, non-scale-free models, e.g., the log-normal. Across domains, we find that scale-free networks are rare, with only 4\% exhibiting the strongest-possible evidence of scale-free structure and 52\% exhibiting the weakest-possible evidence. Furthermore, evidence of scale-free structure is not uniformly distributed across sources: social networks are at best weakly scale free, while a handful of technological and biological networks can be called strongly scale free. These results undermine the universality of scale-free networks and reveal that real-world networks exhibit a rich structural diversity that will likely require new ideas and mechanisms to explain.

Characterizing networks
Structure

\[ V = [m_1, m_2, \ldots, m_k] \]

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean geodesic distance</td>
<td>( \ell )</td>
</tr>
<tr>
<td>Global efficiency</td>
<td>( E )</td>
</tr>
<tr>
<td>Harmonic mean distance</td>
<td>( \bar{h} )</td>
</tr>
<tr>
<td>Vulnerability</td>
<td>( V )</td>
</tr>
<tr>
<td>Network clustering coefficient</td>
<td>( C ) and ( \bar{C} )</td>
</tr>
<tr>
<td>Weighted clustering coefficient</td>
<td>( C^n )</td>
</tr>
<tr>
<td>Cyclic coefficient</td>
<td>( \Theta )</td>
</tr>
<tr>
<td>Maximum degree</td>
<td>( k_{\text{max}} )</td>
</tr>
<tr>
<td>Mean degree of the neighbors</td>
<td>( k_{\text{nn}}(k) )</td>
</tr>
<tr>
<td>Degree-degree correlation coefficient</td>
<td>( \tau )</td>
</tr>
<tr>
<td>Assortativity coefficient</td>
<td>( \tilde{Q}, \tilde{Q} )</td>
</tr>
<tr>
<td>Bipartivity degree</td>
<td>( b ) and ( \beta )</td>
</tr>
<tr>
<td>Degree Distribution entropy</td>
<td>( H(i) )</td>
</tr>
<tr>
<td>Average search information</td>
<td>( S )</td>
</tr>
<tr>
<td>Access information</td>
<td>( A_i )</td>
</tr>
<tr>
<td>Hide information</td>
<td>( H_i )</td>
</tr>
<tr>
<td>Target entropy</td>
<td>( T )</td>
</tr>
<tr>
<td>Road entropy</td>
<td>( R )</td>
</tr>
<tr>
<td>Betweenness centrality</td>
<td>( B_i )</td>
</tr>
<tr>
<td>Central point dominance</td>
<td>( CPD )</td>
</tr>
<tr>
<td>( l )th moment</td>
<td>( M_l )</td>
</tr>
<tr>
<td>Modularity</td>
<td>( Q )</td>
</tr>
<tr>
<td>Participation coefficient</td>
<td>( P_i )</td>
</tr>
<tr>
<td>( z )-score</td>
<td>( z_i )</td>
</tr>
<tr>
<td>Significance profile</td>
<td>( SP_i )</td>
</tr>
<tr>
<td>Subgraph centrality</td>
<td>( SC )</td>
</tr>
<tr>
<td>Hierarchical clustering coefficient</td>
<td>( C_{rs} )</td>
</tr>
<tr>
<td>Convergence ratio</td>
<td>( \psi_d(i) )</td>
</tr>
<tr>
<td>Divergence ratio</td>
<td>( \psi_d(i) )</td>
</tr>
<tr>
<td>Edge reciprocity</td>
<td>( \phi ) and ( \rho )</td>
</tr>
<tr>
<td>Matching index of edge ((i, j))</td>
<td>( \mu_{ij} )</td>
</tr>
</tbody>
</table>
Unveiling the mechanisms of corruption
Dataset:

- We have collected data from Brazilian news magazines and daily newspapers.
  - Wikipedia;
  - Veja;
  - Folha de São Paulo;
  - O Estado de São Paulo.
Caveats:

- Bias from news;
- The list considers people investigated for corruption. This does not mean that the person was found guilty;
- Judicial proceedings take years or even decades;
Dynamics of the number of people involved in corruption scandals
Dynamics of the number of people involved in corruption scandals

Exponential distribution
Dynamics of the number of people involved in corruption scandals

4 yrs periodic-like behavior
Complex network representation of corruption scandals

Two individuals are connected if they are investigated in the same scandal!
Complex network representation of corruption scandals

• We have considered the latest stage of our data (2014) to build the network:
  
  - 404 nodes: individuals investigated for corruption;
  
  - 3549 edges: two nodes are connected if they are investigated under the same corruption scandal;
27 modules, being 14 of them within the giant component
Degree

\[ k_i = \sum_{j=1}^{N} A_{ij}, \]

\[
A = \begin{bmatrix}
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 0
\end{bmatrix}
\]
Netcarto: identifying the nodes’ role in the corruption network

\[ Z_i = \frac{k_i - \bar{k}_{s_i}}{\sigma_{k_{s_i}}} \]

\[ P_i = 1 - \sum_{s=1}^{N_M} \left( \frac{k_{is}}{k_i} \right)^2 \]

Air transportation network

Netcarto: identifying the nodes’ role in the corruption network
Degree distribution

\[ P(k) = \frac{N_k}{N}, \quad k = 0, 1, \ldots \]
Degree distribution

\[ P(k) = \frac{e^{-\langle k \rangle} \langle k \rangle^k}{k!} \]

\[ P(k) \approx k^{-\gamma} \]
Corruption network: Exponential and invariant over time

Exponential distribution
Similarities with the crime organizations

• Terrorism network:


• Drug traffic network:

Network evolution
The characteristic degree exhibits abrupt changes over the years.
Changes in the size of the largest component of the corruption network over time are caused by a coalescence of network modules.
Collor case:

*Impeachment of Fernando Affonso Collor de Mello*
Cayman Dossier:
Set of papers of companies on the Cayman Islands for money laundry.

Fig. 4. Changes in the size of the largest component of the corruption network over time are caused by a coalescence of network modules. (A) Evolving behaviour of the fraction of nodes belonging to the main component of the time-varying network (size of the largest cluster) over the years. (B) The growth rate of the size of the largest cluster (that is, the derivative of the curve of the previous plot) over the years. In both plots, the shaded regions indicate the term of each Brazilian President from 1986 to 2014 (names and parties are shown in the plot). We note the existence of three abrupt changes between the years 1991–1992, 1997–1998 and 2004–2005. (C) Snapshots of the changes in the complex network between the years in which the abrupt changes in the main component took place. We note that between 1991 and 1992, the abrupt change was simply caused by the appearance of the corruption scandal ‘Caso Collor’, that became the largest component of the network in 1992. The abrupt change between 1997 and 1998 is caused not only by the appearance of three new corruption cases, but also due to the coalescence of two of these new cases with previous network modules. The change between 2004 and 2005 is also caused by the coalescence of previous network components with new corruption cases. In these plots, the modules are coloured with the same colours between consecutive years and new nodes are shown in black. The names of all scandal are shown in the plots.
“Mensalão” – big monthly payment: Vote-buying of parliamentarians

Fig. 4. Changes in the size of the largest component of the corruption network over time are caused by a coalescence of network modules. (A) Evolving behaviour of the fraction of nodes belonging to the main component of the time-varying network (size of the largest cluster) over the years. (B) The growth rate of the size of the largest cluster (that is, the derivative of the curve of the previous plot) over the years. In both plots, the shaded regions indicate the term of each Brazilian President from 1986 to 2014 (names and parties are shown in the plot). We note the existence of three abrupt changes between the years 1991–1992, 1997–1998 and 2004–2005. (C) Snapshots of the changes in the complex network between the years in which the abrupt changes in the main component took place. We note that between 1991 and 1992, the abrupt change was simply caused by the appearance of the corruption scandal 'Caso Collor', that became the largest component of the network in 1992. The abrupt change between 1997 and 1998 is caused not only by the appearance of three new corruption cases, but also due to the coalescence of two of these new cases with previous network modules. The change between 2004 and 2005 is also caused by the coalescence of previous network components with new corruption cases. In these plots, the modules are coloured with the same colours between consecutive years and new nodes are shown in black. The names of all scandal are shown in the plots.
Predicting missing links in complex networks
General overview of the methods

• A link can exist between two nodes if they are similar in some properties.
Some measures for predicting missing links

- Number of common neighbors:
  - Nodes sharing a large number of neighbors are likely to be connected.
Some measures for predicting missing links

- Short paths between nodes:
  - How soon two random walkers are expected to meet at the same node when starting at two other nodes.
Random walks
Some measures for predicting missing links

• Hierarchical structures:
  - Based on the existence of hierarchical structures within the network.

From the top-10 links predicted by SimRank, more than 25% appeared later in the network.
The role of network science in mitigating corruption

• How could we use these results to prevent corruption?
In summary:

• Corruption runs in small groups;
• The network has hubs and a modular structure;
• Growth through the coalescence of modules;
• Predicting partners in future scandals;
Perspectives
Getting more data
Future analysis
Collaborators

Matjaž Perc

Haroldo V Ribeiro

Ervin K. Lenzi

Alvaro F. Martins
Thank you!

@lgaalves
lgaalves@usp.br
@lgaalves
https://lgaalves.github.io/