Mapping the Structure of Directed Networks: Beyond the Bow-Tie Diagram

Gábor Timár[1,*], Alexander V. Goltsev[1,2], Sergey N. Dorogovtsev[1,2] and José F. F. Mendes[1]
[1]Universidade de Aveiro, Portugal; [2]A. F. Ioffe Physico-Technical Institute, Russia
*gtimar@ua.pt

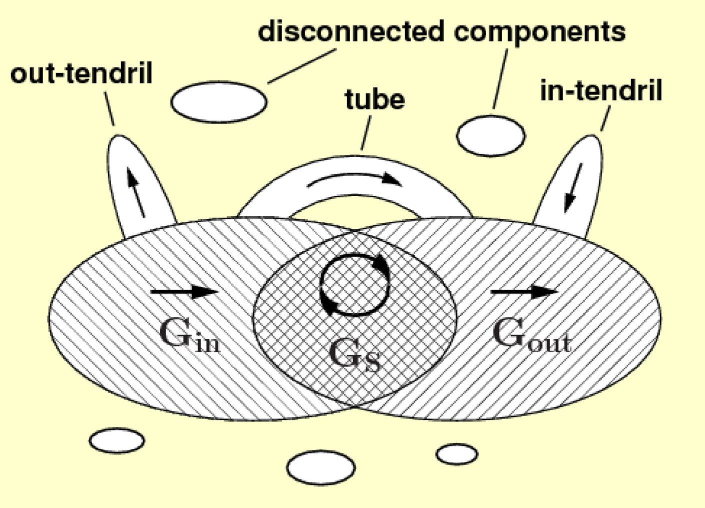
UNIVERSIDADE DE AVEIRO

VI Annual Meeting i3N, Lisboa 2017

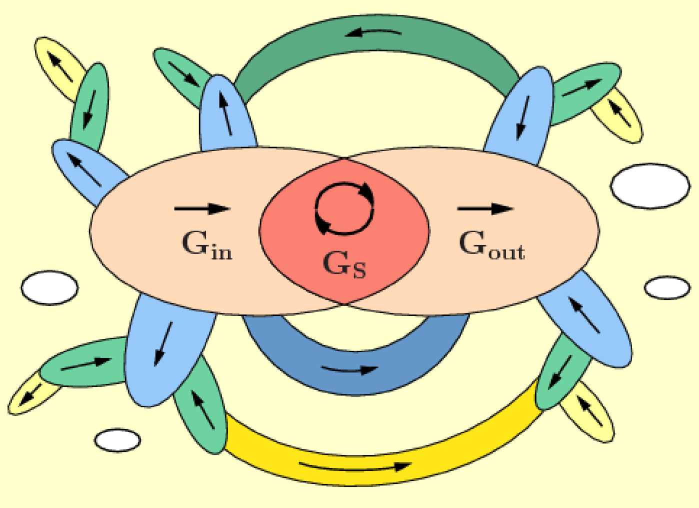
Complete decomposition of directed networks

The general picture

Many real-world networks can be represented by directed graphs, where each link connecting two nodes is assigned one of two possible directions, or both. Well-known examples are the Twitter social network, the World Wide Web (WWW), neuronal and metabolic networks, etc. Any large directed network can be partitioned into several qualitatively different subgraphs: (i) a giant strongly connected component G_S, and giant in- and out-components G_{in} and G_{out}; (ii) finite directed components (tendrils and tubes); (iii) disconnected finite components.



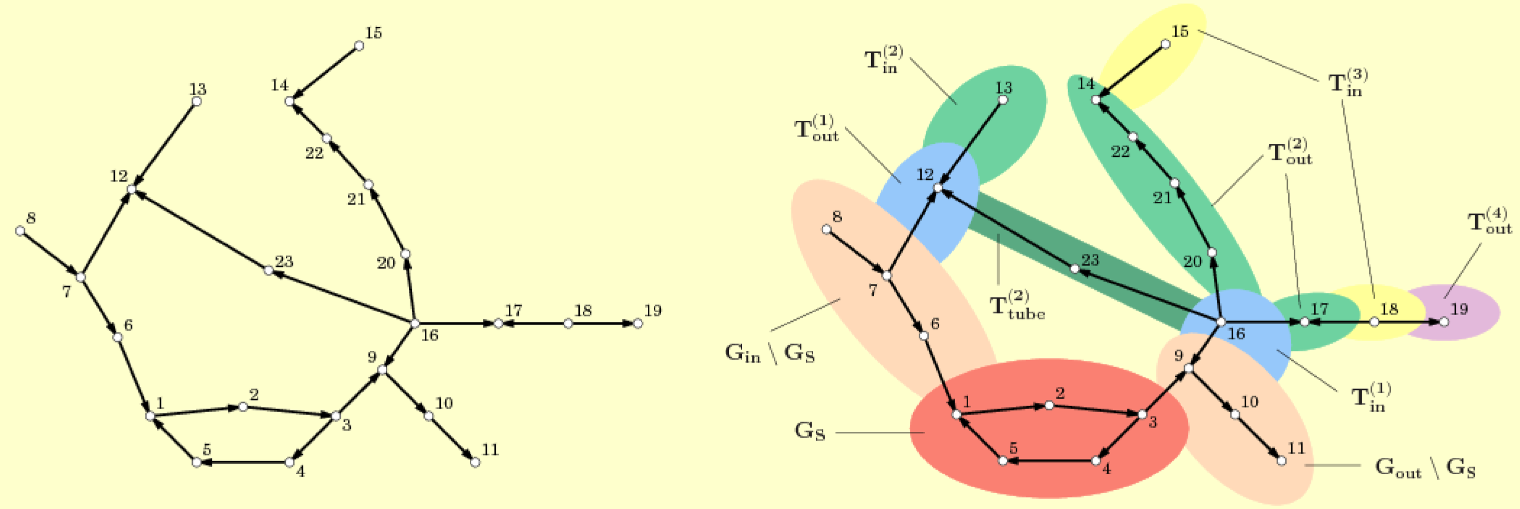
The classical "Bow-tie" diagram of Broder et al. [1] representing the structure of directed networks. G_S is the largest set of nodes in which every node can be reached from any other node. G_{in} contains all nodes from which G_S is reachable. G_{out} consists of all nodes that can be reached from G_S. These "giant" components constitute the core of the network and they are usually much larger than other components. "Tendrils" contain additional nodes that can be reached from G_{in} or ones from which G_{out} is reachable (empty domains with arrows). A "tube" is an in- and out-tendril simultaneously. There may also be disconnected clusters (empty ovals).
The tendrils of this picture are only a first layer, a large fraction of nodes may still be attached to this first layer. All these nodes have so far been ignored!!



Our improved representation of directed networks [2]. This picture provides a complete decomposition, accounting for all nodes in the network. We introduce "tendril layers" (domains colored blue, green, yellow).
1. tendril layer: attached to G_{in} and G_{out}
2. tendril layer: attached to 1. tendril layer
3. tendril layer: attached to 2. tendril layer
There may be tubes in any of the tendril layers (simultaneous in- and out-tendrils).
There may be an arbitrarily large number of tendril layers, depending on the size and structure of the network.

See the opposite panel for the decomposition of a small example directed network.

A simple example



Giant components (the core of the network):

- $G_S = \{1, 2, 3, 4, 5\}$ Largest set where every node is reachable from every other node.
- $G_{in} \setminus G_S = \{6, 7, 8\}$ Nodes from which G_S is reachable, but are not in G_S.
- $G_{out} \setminus G_S = \{9, 10, 11\}$ Nodes that can be reached from G_S, but are not in G_S.

Tendril layers:

1. layer
   - $T_{in}^{(1)} = \{16\}$ Additional nodes from which G_{out} can be reached.
   - $T_{out}^{(1)} = \{12\}$ Additional nodes that can be reached from G_{in}.
2. layer
   - $T_{in}^{(2)} = \{13, 23\}$ Additional nodes from which $T_{out}^{(1)}$ can be reached.
   - $T_{out}^{(2)} = \{14, 17, 20, 21, 22, 23\}$ Additional nodes that can be reached from $T_{in}^{(1)}$.
   - $T_{tube}^{(2)} = \{23\}$ Nodes that belong simultaneously to $T_{in}^{(2)}$ and $T_{out}^{(2)}$.
3. layer
   - $T_{in}^{(3)} = \{15, 18\}$ Additional nodes from which $T_{out}^{(2)}$ can be reached.
4. layer
   - $T_{out}^{(4)} = \{19\}$ Additional nodes that can be reached from $T_{in}^{(3)}$.

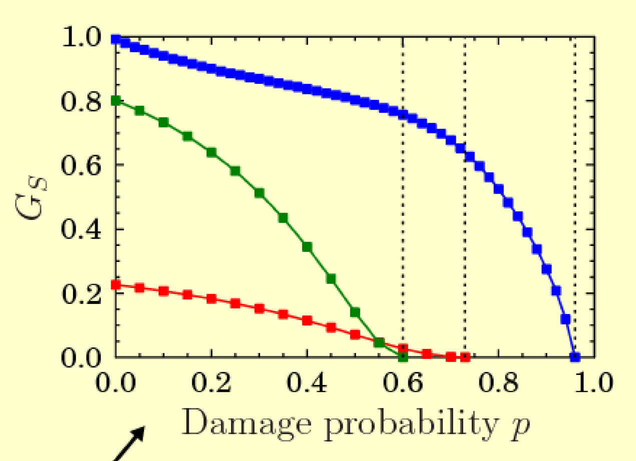
Randomly damaged directed networks

What happens to the structure when each link is destroyed with a probability p?

Technological networks are always prone to failures and biological networks may be affected by diseases. What is more, nodes or links may be destroyed on purpose to alter the functioning of a network. The first step in understanding the structural changes in damaged networks is to assume random damage. Let each link be removed with probability p. As p is increased, the core of the network - the giant components - decreases in size, and more tendril layers appear. At a certain critical value of p (the percolation critical point p_c) the giant directed components disappear and the network disintegrates into small, disconnected directed components. Here we show these typical behaviors for various important real-world directed networks, compared with random networks.

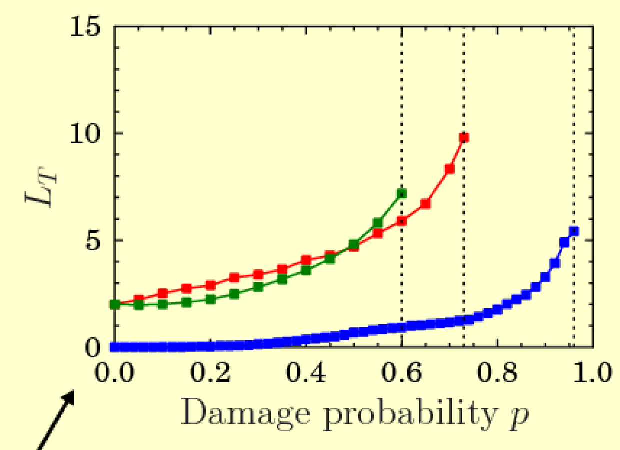
Gnutella file-sharing network (62586 nodes, 147892 links)
C. elegans neural network (495 nodes, 7938 links)"
Random network (10000 nodes, 25000 links)

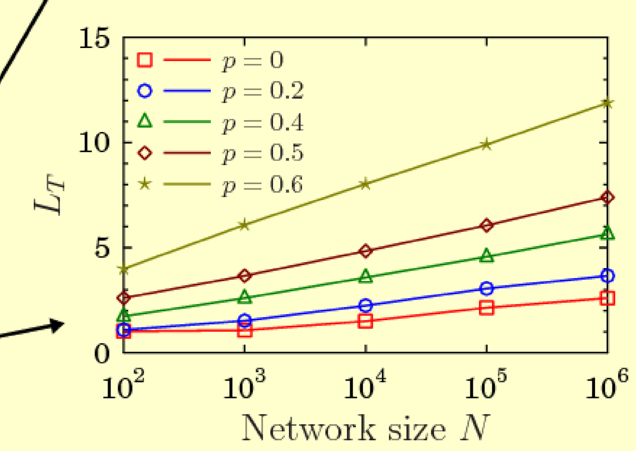
Relative size of giant strongly connected component, G_S



critical points p_c are marked with dashed lines

Number of tendril layers, L_T



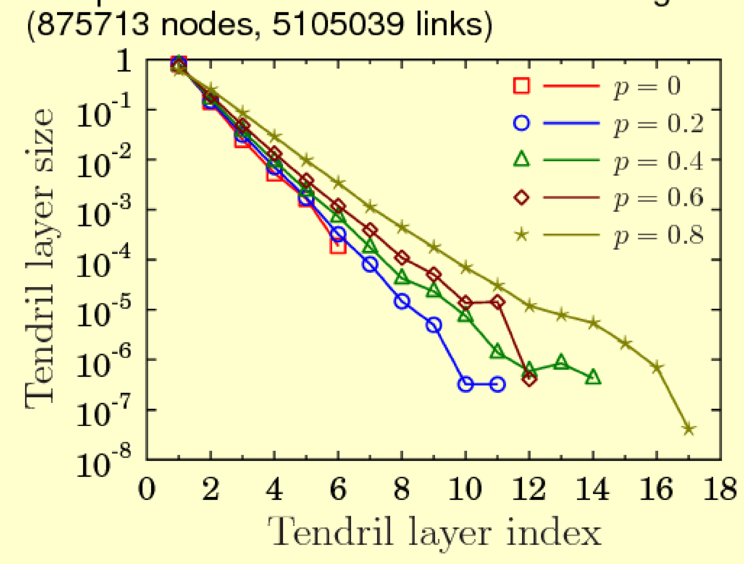


- The sizes of the giant components go to zero as we approach the critical point p_c. (Here only the behavior of the giant strongly connected component G_S is shown.)
- The number of tendril layers increases in damaged networks. The number of layers has a maximum at the critical point p_c.
- The number of tendril layers increases logarithmically with network size N for random networks. The increase becomes faster as the damage increases.

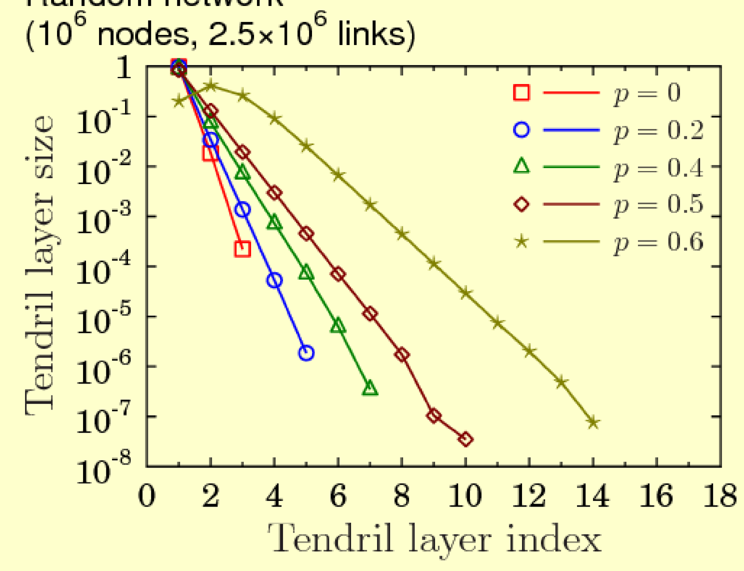
Size sequence of tendril layers

- Each successive tendril layer is smaller by a given fraction (exponential decay).
- Tendril layer sizes decay slower as the damage increases.

Sample of the World Wide Web from Google (875713 nodes, 5105039 links)



Random network (10^6 nodes, 2.5×10^6 links)



(y-axes show relative sizes of tendril layers)

Summary

- We introduce a complete decomposition of directed networks, improving on a long-standing incomplete description.
- We show that directed networks have a complex structure: a hierarchy of tendril layers.
- The number and sizes of tendril layers increase as we approach the percolation critical point.
- We hope our scheme will help reveal as yet undiscovered structures in important real-world directed networks such as Twitter, the WWW, neuronal networks and metabolic networks.

Acknowledgment

This work was supported by the FET proactive IP project MULTIPLEX 317532 and the grant PEST UID/CTM/50025/2013.

[1] A. Broder et al.,"Graph structure in the web", Computer Networks **33**, 309 (2000)
[2] G. Timar et al., "Mapping the Structure of Directed Networks: Beyond the Bow-Tie Diagram", Physical Review Letters **118**, 078301 (2017)

This work is funded by National Funds throught FCT - Portuguese Foundation for Science and Technology, Reference UID/CTM/50025/2013 and FEDER funds through the COMPETE 2020 Programme under the project number POCI-01-0145-FEDER-007688

INSTITUTE FOR NANOSTRUCTURES, NANOMODELLING AND NANOFABRICATION i3N
FACULDADE DE CIÊNCIAS E TECNOLOGIA UNIVERSIDADE NOVA DE LISBOA
universidade de aveiro
Universidade do Minho Escola de Engenharia
Cofinanciado por: COMPETE 2020 PORTUGAL 2020 UNIÃO EUROPEIA Fundo Europeu de Desenvolvimento Regional
FCT Fundação para a Ciência e a Tecnologia MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR