Effective structure of complex networks and a second look at message passing approaches

Antoine Allard[1,3] and Laurent Hébert-Dufresne[2,3]

1. Institute of Complex Systems, Universitat de Barcelona, Spain
2. Vermont Complex Systems Center, University of Vermont, USA
3. Université Laval, Québec, Canada

Summary

Context

- Most theoretical approaches to model complex networks are inspired by statistical physics.
  - ✓ **Analytical** treatment.
  - ✓ Input information **intensive** in network size.
  - ✓ Useful **null models**.
  - ✗ No systematically accurate quantitative predictions for most dynamical processes on real networks.
- As a result, the current **state-of-the-art** approach—the message passing approach (MPA)—requires **extensive** input information (the whole structure).
  - ✓ Exact predictions on trees.
  - ✓ Inexact abeit **generally good quantitative predictions** on networks with loops.
  - ✗ Time and space complexity scale with the network size.
  - ✗ **No insight** on the role played by any structural property on the outcome of a dynamical process.

Main results

- We **bridge the gap between intensive and extensive approaches** with a mathematical model of networks that
  - ✓ relies solely on an **intensive** description of the structure;
  - ✓ yields **exact predictions on trees**;
  - ✓ yields **accurate predictions** comparable to the ones obtained with the MPA for networks with loops;
  - ✓ shows how **local** connection rules can constrain the **meso-scale organization** of the network;
  - ✓ offers an **intensive effective description** of networks.

Onion Decomposition (OD)

- Refines the k-core decomposition of a network in which **each core is subdivided into layers**.
- Specifies the **position of a node** into its k-shell and in the **mesoscale centrality organization** of the network.
- Characterizes the internal organization of each shell (i.e., tree vs. lattice).
- Can be **quickly obtained** for virtually any real complex networks (extracted alongside the k-core decompostion).

Characterizing each node with the **layer-degree pair** (l, d) indicates how well it is connected as well as its **"topological position"**.

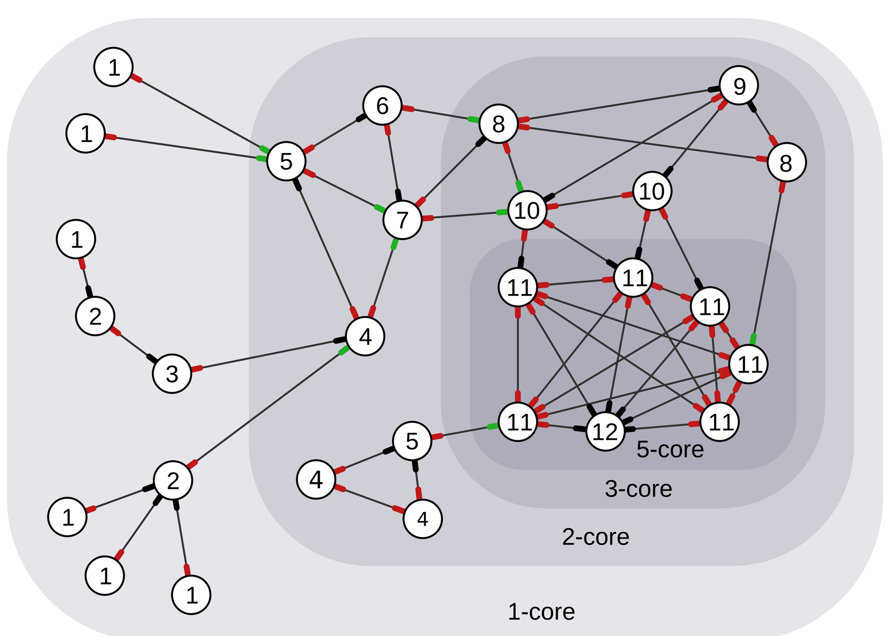

Fig. 1: Illustration of the Onion Decomposition (OD) of a simple network. The number of the layer to which each node belongs is indicated and the different k-cores are shown using increasingly darker background shades. The color of each stub according the LCCM is also shown.

Further information

Multi-scale structure and topological anomaly detection via a new statistic: The onion decomposition
Scientific Reports **6**, 31708 (2016)

Percolation and the effective structure of complex networks
arXiv:1804.09633 (2018)

Analytical predictions of percolation on real complex networks: A cautionary tale
coming soon (2018)

Layered and correlated configuration model (LCCM)

Local rules enforcing meso-scale organization

A node of coreness k belonging to the l-th layer must

1. have exactly k links with nodes in layers $l' \geq l$ if layer l is the first layer in the k-shell (i.e., layer $l-1$ is in the k'-shell with $k' < k$).
2. Otherwise, it must have at least $k+1$ links to nodes in layers $l' \geq l-1$ and at most k links to nodes in layers $l' \geq l$.

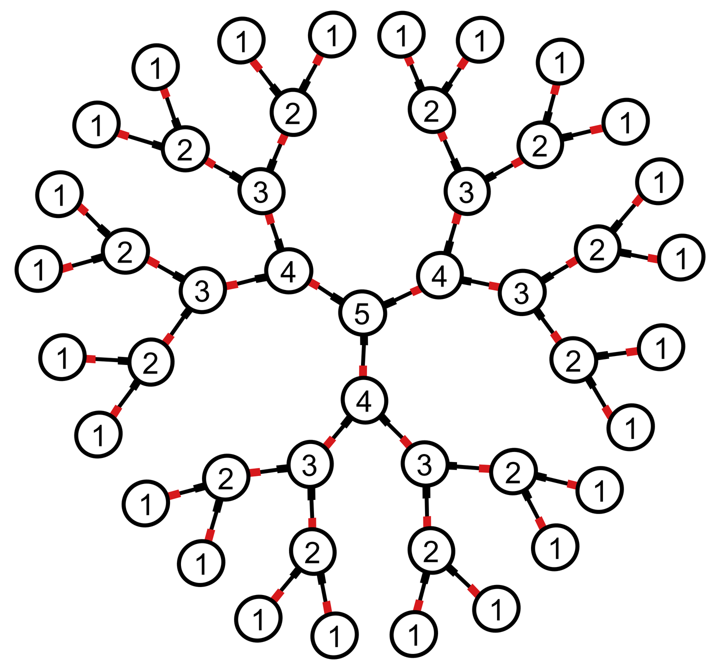

Fig. 2: Compression of structural information of a perfect tree. Whereas the MPA would assigns a unique label to each node, the LCCM is able reproduce the structure using only intensive information.

Effective random network ensemble

- **Rewiring the links using a degree-preserving procedure** while ensuring that the above connection rules are respected at all time allows to explore the ensemble of **all networks with the same layer-degree sequence**.
- Doing so while preserving the density of links between and within each layer-degree class (two-point correlations) of nodes defines the **Layered and Correlated Configuration Model**.
- The LCCM is a subset of the Correlated Configuration Model (CCM) and of the Configuration Model (CM).

Solving percolation on the LCCM

We solve bond/site percolation on the LCCM using a multitype probability generating function formalism [*General and exact approach to percolation on random graphs*, Phys. Rev. E 92, 062807 (2015)]. This approach

- identifies nodes using the layer-degree pair;
- classifies half-links (stubs) as red, black or green according to whether they point towards nodes in layer $l' \geq l$, $l' = l-1$ or $l' < l-1$, respectively. Links can only consist of combinations red-red, red-black and red-green;
- provides exact expression for the **relative size of the extensive component** S and the **percolation threshold** p_c.

The meso-scale information of the OD

- **improves the predictions** obtained using previous models (CCM, CM);
- **drastically changes the nature** of the predictions (e.g., inflection points);
- predicts a connected network when supposed to;
- predicts a threshold with a **relative error below 1.5%** for 75% of the 111 real networks tested;
- provides **exact predicitons on trees**.

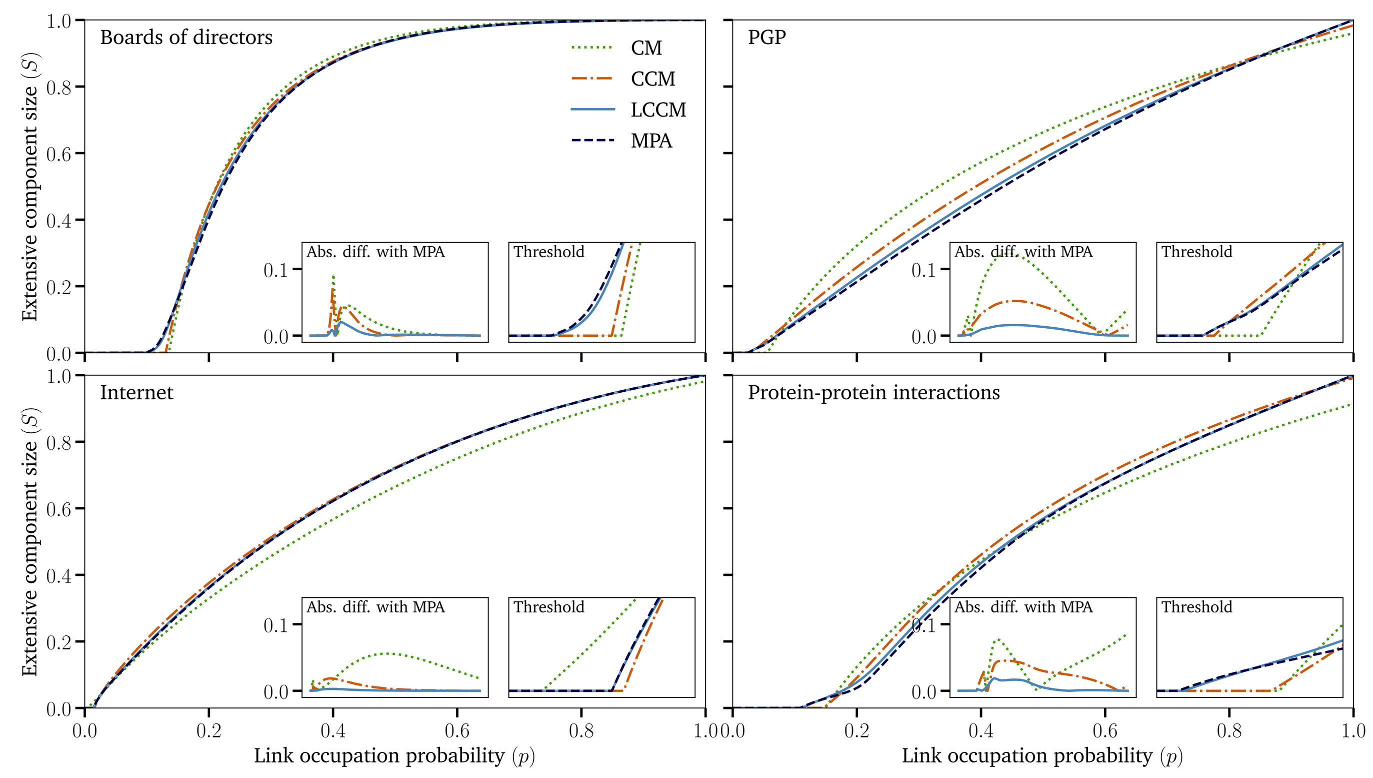

Fig. 3: Relative size of the extensive component predicted by the LCCM with the CM, the CCM and the MPA for 4 representative real network datasets. The insets show the absolute value of the difference between the MPA and the CM, the CCM and the LCCM as a function of p, as well as an enlargement of the region around the percolation threshold. The largest connected component was used for all dataset.

Second look at the Message Passing Approach

Absence of phase transition

Solving percolation using the MPA relies on the **assumption that the state of second neighbors are independent**. This assumption

- has been proven to be very accurate for a wide range of real complex networks;
- implies exact results on tree-like networks (no loop).

However, the predictions in this exact regime are drastically different than one could naively expect: there is **no phase transition**.

- **No extensive component**.
- **No independence of scale** at the "expected" threshold.

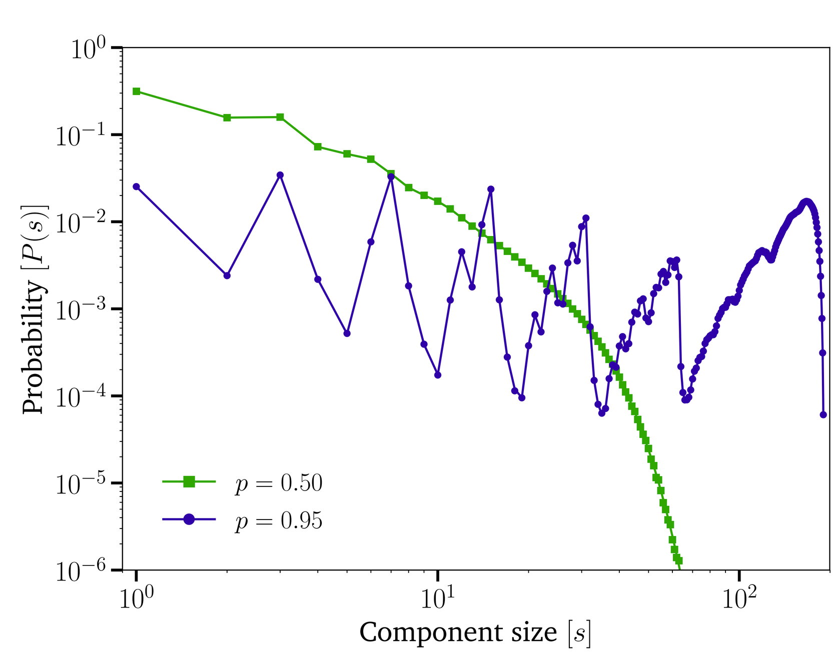

Fig. 4: Distribution of the size of the non-extensive components predicted by MPA (lines) compared with the results of numerical simulations (markers) for a Cayley tree with a coordination number equal to 3 and 7 generations ($N = 190$). Bethe lattice with the same coordination number percolates at $p = 0.5$.

What is the MPA percolating on?

- The explicit use of the adjacency matrix suggests that the MPA percolates directly on the actual network structure.
- On the contrary, it percolates on the ensemble of random networks defined by **L-cloning** with $L \to \infty$ (See Fig. 5).
- If the original network is a tree, the MPA sees an infinite number of finite trees ⇒ **no phase transition**.
- If the original network contains loops, its 2-core is stretched into a "heteregeneous" Bethe lattice ⇒ **phase transition**.

$$\text{Extensive comp. size:} \quad S = \frac{1}{N}\sum_{i=1}^{N}\left(1 - \prod_{j=1}^{N} u_{ij}^{a_{ij}}\right)$$

$$\text{where} \quad u_{ij} = (1-p) + p\prod_{l=1}^{N} u_{jl}^{(1-\delta_{il})a_{jl}}$$

where N is the number of nodes, p is the bond occupation probability, and $\{a_{ij}\}$ are the entries of the adjacency matrix.

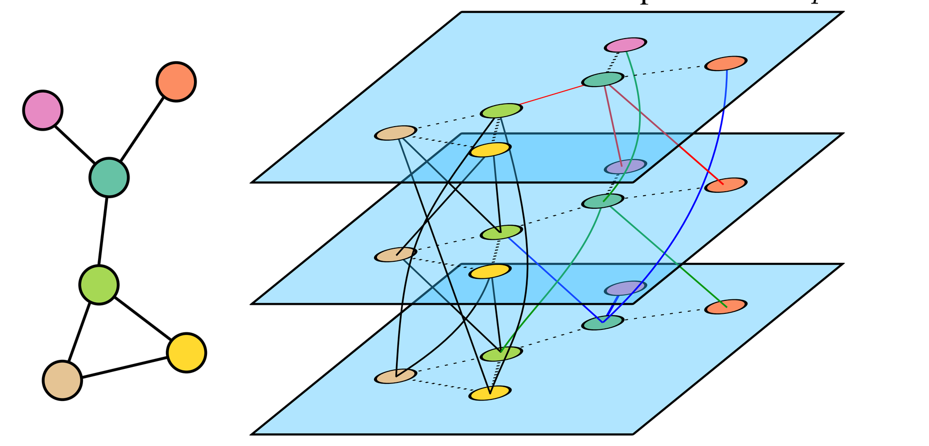

Fig. 5: Illustration of the L-cloning procedure introduced in *Network cloning unfolds the effect of clustering on dynamical processes*, Phys. Rev. E 91, 052807 (2015) with $L = 3$ for the small network shown on the left.