Percolation on clustered and correlated random graphs

General formalism and applications

Antoine Allard, Laurent Hébert-Dufresne, Jean-Gabriel Young and Louis J. Dubé

Département de physique, de génie physique, et d'optique, Université Laval, Québec, Canada

Summary

– We present a very general model of **site and bond percolation** on random graphs that can be **clustered** and **correlated** in various ways.

– **Percolation properties are solved exactly** for small graphs and for "infinite size" graphs.

– **Theoretical laboratory**: the **very general nature** of the model permits to **investigate a wide range of questions** related to the **influence of the structure of networks on dynamical processes** taking place on them (e.g., disease propagation, robustness). Two applications illustrate our findings.

A *stub matching* scheme

– Each node belongs to a type ($\mathcal{N}$: set of types of nodes).

– Nodes of type i occupy a fraction w_i of the nodes.

– Each node has k_α stubs of type α ($\mathcal{E}$: set of types of stubs).

– The joint degree distribution $P_i(k_1,\ldots,k_{|\mathcal{E}|}) \equiv P_i(\boldsymbol{k})$ prescribes the number of stubs of each type that nodes of type i have.

– Hyperedges are formed by matching stubs according to a set of arbitrary rules $\mathcal{R}$. **Hyperedges** can be any kind of **arbitrary motif** and can have a **fixed or random** structure.

– The rules $\mathcal{R}$ define $R_\alpha(n_1,\ldots,n_{|\mathcal{N}|}) \equiv R_\alpha(\boldsymbol{n})$, the distribution of the composition (i.e., n_i nodes of type i for each $i \in \mathcal{N}$) of the hyperedge reached from a stub of type α. They also define $\{\kappa_{\mu\nu}\}$, the matrix prescribing the types of stubs that can be matched to form an hyperedge.

– Graphs are generated by drawing nodes with stubs according to $P_i(\boldsymbol{k})$, and by randomly matching stubs according the rules $\mathcal{R}$. This process yields a **maximally random graph ensemble** subjected to the constraints defined above.

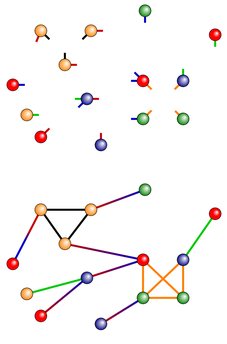

Fig. 1: **Illustration of the stub matching scheme. (top)** Nodes are drawn randomly according to $P_i(\boldsymbol{k})$ and w_i. **(bottom)** Stubs are matched according to the rules $\mathcal{R}$. For instance, three black stubs are matched to form a triangle, and a blue and a red stub are matched to form a directed edge (from the blue to the red stub). There are 4 types of nodes, 5 types of stubs, and 4 different kinds of hyperedges.

Solving percolation

Small arbitrary motifs

To mathematically describe the percolation on the graphs generated by the stub matching scheme, **bond and/or site percolation on each hyperedge generated by the rules $\mathcal{R}$ must be solved beforehand** [1].

– The distribution of the **number of nodes of each type that can be reached** from a node of type i through a stub of type α is

$$Q_{i\alpha}(\boldsymbol{l}|\boldsymbol{n}) = \sum_{\boldsymbol{b}=\boldsymbol{\delta}_i}^{\boldsymbol{n}} W_{i\alpha}(\boldsymbol{l}|\boldsymbol{b}) \prod_{j\in\mathcal{N}} \binom{n_j-\delta_{ij}}{b_j-\delta_{ij}} p_{\alpha j}^{b_j-\delta_{ij}} \left(1-p_{\alpha j}\right)^{n_j-b_j} .$$

where $W_{i\alpha}(\boldsymbol{l}|\boldsymbol{b})$ is obtained by **iterating**

$$W_{i\alpha}(\boldsymbol{l}|\boldsymbol{b}) = W_{i\alpha}(\boldsymbol{l}|\boldsymbol{l}) \prod_{j\in\mathcal{N}} \binom{b_j-\delta_{ij}}{l_j-\delta_{ij}} \prod_{k\in\mathcal{N}} (1-T_{\alpha jk})^{l_j(b_k-l_k)}$$

$$W_{i\alpha}(\boldsymbol{l}|\boldsymbol{l}) = 1 - \sum_{\boldsymbol{r}<\boldsymbol{l}} W_{i\alpha}(\boldsymbol{r}|\boldsymbol{l})$$

from the **initial condition** $W_{i\alpha}(\boldsymbol{\delta}_i|\boldsymbol{\delta}_i) = 1$ (one node of type i).

– The parameters $\boldsymbol{n}$ (hyperedge composition), $\{p_{\alpha j}\}$ (**node occupation probabilities**) and $\{T_{\alpha jk}\}$ (**edge occupation probabilities**) are specified by the rules $\mathcal{R}$.

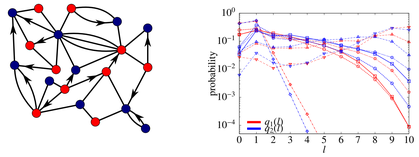

Fig. 2: **(left) Example of an hyperedge that can be handled by this approach.** There are two types of nodes, and edges may be simple, multiple or directed. **(right) Validation of the analytical framework.** Probability distribution $q_j(l)$ of reaching l nodes of type j from a random initial node [i.e., projections of $Q_{i\alpha}(\boldsymbol{l}|\boldsymbol{n})$]. Six different sets of site/bond occupation probabilities were considered. Lines: theory; symbols: simulations.

Infinite size random graphs

Exact mapping to a graph ensemble with a locally **treelike structure** through the use of the generating functions [2]

$$\theta_{i\alpha}(\boldsymbol{x}) = \sum_{\boldsymbol{n}=\boldsymbol{0}}^{\infty} \frac{n_i R_\alpha(\boldsymbol{n})}{\langle n_i\rangle_{R_\alpha}} \sum_{\boldsymbol{l}=\boldsymbol{\delta}_i}^{\boldsymbol{n}} Q_{i\alpha}(\boldsymbol{l}|\boldsymbol{n}) \prod_{j\in\mathcal{N}} x_{\alpha j}^{l_j-\delta_{ij}}$$

where $\langle n_i\rangle_{R_\alpha} = \sum_{\boldsymbol{n}=\boldsymbol{0}}^{\infty} n_i R_\alpha(\boldsymbol{n})$.

– The relative size of the **giant component** is

$$\mathcal{S} = 1 - \sum_{i\in\mathcal{N}} w_i \sum_{\boldsymbol{k}=\boldsymbol{0}}^{\infty} P_i(\boldsymbol{k}) \prod_{\alpha\in\mathcal{E}} \left[\theta_{i\alpha}(\boldsymbol{a})\right]^{k_\alpha}$$

where $\boldsymbol{a} \equiv \{a_{\mu i}\}_{\mu\in\mathcal{E},i\in\mathcal{N}}$ is the **stable fixed point** of the system

$$a_{\mu i} = \sum_{\nu\in\mathcal{E}} \kappa_{\mu\nu} \sum_{\boldsymbol{k}=\boldsymbol{0}}^{\infty} \frac{k_\nu P_i(\boldsymbol{k})}{\langle k_\nu\rangle_{P_i}} \prod_{\alpha\in\mathcal{E}} \left[\theta_{i\alpha}(\boldsymbol{a})\right]^{k_\alpha-\delta_{\alpha\nu}}$$

with $\langle k_\nu\rangle_{P_i} = \sum_{\boldsymbol{k}=\boldsymbol{0}}^{\infty} k_\nu P_i(\boldsymbol{k})$.

– The **percolation threshold** corresponds to the point where $\boldsymbol{a} = \boldsymbol{1}$ becomes an unstable fixed point of the system of equations above.

I. Hard-core random networks

We generate a maximally random graph ensemble with an arbitrary degree distribution and an **arbitrary k-core structure**, and use it to **model bond percolation on real networks**.

– The type of a node corresponds to its coreness, c.

– $C_{cc'}$:number of edges leaving a node of type c to a node of type c'.

– K_{ck}:number of nodes of type c that have a degree k.

– A node of type c with a degree k has c stubs of type $2c$ that **contribute** to its coreness and $k-c$ stubs of type $2c-1$ that **do not contribute**.

– **Stubs are matched randomly** under the following **constraints**:

- contrib. stubs must lead to nodes with equal or higher c;
- non-contrib. stubs must lead to nodes with equal or lower c;
- two non-contributing stubs cannot be paired together.

Our random graph model **performs better at predicting bond percolation on real networks** than widely used models with the advantage of **requiring less information** $\left[\mathcal{O}(k_{\max}^{3/2})\right.$ instead of $\left.\mathcal{O}(k_{\max}^{2})\right]$.

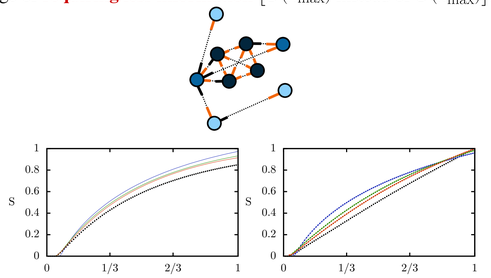

Fig. 3: **(top) Network construction procedure.** Nodes of coreness 1, 2 and 3 (from lighter to darker) have contributing stubs (orange) and non-contributing stubs (black) that are randomly matched under the constraints outlined above. **(bottom) Predicitons of the size of the giant component** (orange) as a function of the occupation probability of edges, T, using the matrices $\{C_{cc'}\}$ and $\{K_{ck}\}$ extracted from the MathSci (left) and PGP (right) networks [3]. They are compared to the results of bond percolation on the real datasets (black), and to the predictions obtained by considering the degree distribution (blue), and the degree correlations (green) solely. Lines: theory; dots: simulations.

Details of the mapping from the matrices $\{C_{cc'}\}$ and $\{K_{ck}\}$ extracted from real network datasets to the stub matching scheme:

$$w_c = \frac{\sum_{k=c}^{k_{\max}} K_{ck}}{\sum_{c'=1}^{c_{\max}} \sum_{k'=c'}^{k_{\max}} K_{c'k'}} \,;\; P_c(\ldots,k_{2c-1}=k-c,k_{2c}=c,\ldots) = \frac{K_{ck}}{\sum_{k'=c}^{k_{\max}} K_{ck'}}$$

$$\kappa_{2c',2c-1} = \kappa_{2c-1,2c''} = 1 \quad \text{with } c' < c \le c''$$

$$\kappa_{2c,2c-1} = 1 - \kappa_{2c,2c} = \frac{C_{cc} - \left[Ncw_c - \sum_{c''=c+1}^{c_{\max}} C_{cc''}\right]}{\left[Ncw_c - \sum_{c''=c+1}^{c_{\max}} C_{cc''}\right]}$$

$$R_{2c}(\ldots,n_c=1,\ldots,n_{c'}=1,\ldots) = \frac{C_{cc'}}{Nw_c c}$$

$$R_{2c}(\ldots,n_c=2,\ldots) = \frac{Nw_c c - \sum_{c''=c+1}^{c_{\max}} C_{cc''}}{Nw_c c}$$

$$R_{2c-1}(\ldots,n_{c'}=1,\ldots,n_c=1,\ldots) = \frac{C_{cc'}}{\sum_{c''=1}^{c_{\max}} C_{cc''} - Nw_c c}$$

$$R_{2c-1}(\ldots,n_c=2,\ldots) = \frac{\sum_{c''=c}^{c_{\max}} C_{cc''} - Nw_c c}{\sum_{c''=1}^{c_{\max}} C_{cc''} - Nw_c c}$$

II. Clustering regimes

Conjecture A [4]: clustering has **opposite effects** on the bond percolation threshold and on the size of giant component depending of the **density of triangles** in a graph.

– $\bar{c}(k)$: average number of triangles to which a node of degree k belongs.

– **Weak** clustering, $\bar{c}(k) < (1-k)^{-1}$, leads to a **higher percolation threshold** and to a **smaller giant component** than for an equivalent unclustered graph.

– **Strong** clustering, $\bar{c}(k) > (1-k)^{-1}$, leads to a **lower percolation threshold** and to a **larger giant component** than for an equivalent unclustered graph.

Using our model, we design a graph that **qualifies for the strong regime, but whose percolation properties are those of the weak regime**.

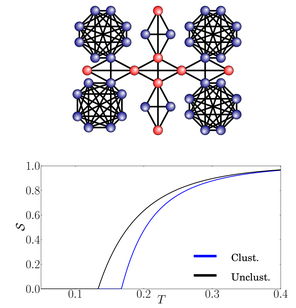

Fig. 4: **(top) A graph contradicting Conjecture A** (only a fraction of the edges is shown). Each red node (degree 6) belongs to a motif of 4 red nodes, and to a motif of 2 red nodes and 2 blue nodes. Each blue node (degree 10) belongs to a motif of 8 blue nodes, and to a motif of 2 red nodes and 2 blue nodes. Hence $\bar{c}(6) = \frac{9}{15} > \frac{1}{5}$ and $\bar{c}(10) = \frac{24}{45} > \frac{1}{9}$, which qualifies this graph for the strong clustering regime. **(bottom) Size of the giant component**, $\mathcal{S}$, as a function of the occupation probability of edges, T, for the clustered graph shown above (blue) and for its unclustered random counterpart (black). In contrast to Conjecture A, here clustering **increases** the percolation threshold and **decreases** the size of the giant component.

This **counterexample** strongly suggests that the criterion on $\bar{c}(k)$ **may be a necessary condition but not a sufficient one**.
We propose **Conjecture B**: weak and strong clustering regimes can be uniquely determined from the existence of an **effective local treelike structure**. Investigation is under way to validate this new conjecture.

[1] **A. Allard, L. Hébert-Dufresne, *et al.*, *EPL* 98, 16001 (2012)**

[2] **A. Allard, L. Hébert-Dufresne, *et al.*, *J. Phys. A* 45, 405005 (2012)**

[3] **L. Hébert-Dufresne, A. Allard *et al.*, arXiv:1208.5768 (2013)**

[4] M.Á. Serrano and M. Boguñá, *Phys. Rev. Lett.* 97, 088701 (2006)

IRSC CIHR Instituts de recherche en santé du Canada / Canadian Institutes of Health Research

NSERC CRSNG

Fonds de recherche sur la nature et les technologies Québec