Multitype modular networks as a model of clustered social networks

Antoine Allard, Pierre-André Noël, Laurent Hébert-Dufresne, Vincent Marceau, and Louis J. Dubé

Département de Physique, de Génie Physique et d'Optique, Université Laval, Québec, Canada

Introduction

Tackling the structural complexity of social networks in an analytical framework is not an easy task and many existing models must rely on simplifying assumptions in order to be solvable.

Widely used approximation: Configuration Model (CM) [2]

- *analytical tractability*
- *explicit neglect of sub-structures (treelike assumption)*

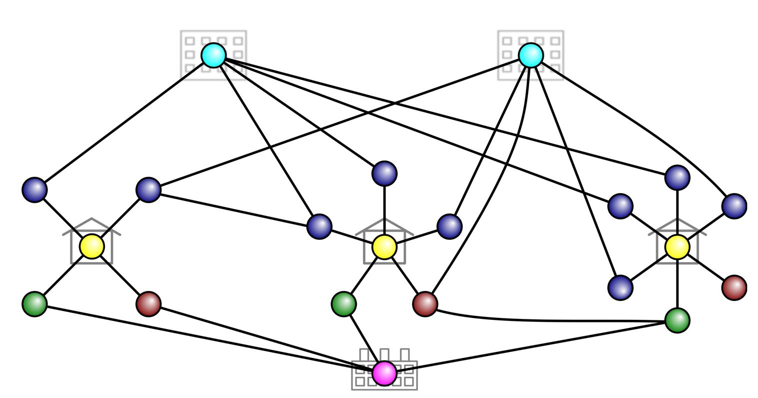
We have developped an ***analytical bond percolation*** *formalism that* ***succesfully describes topological properties of networks featuring detailed substructures****.*

Multitype modular networks

We introduce a multitype [3] and modular [4] ***generalization of the Configuration Model****.*

Multitype network

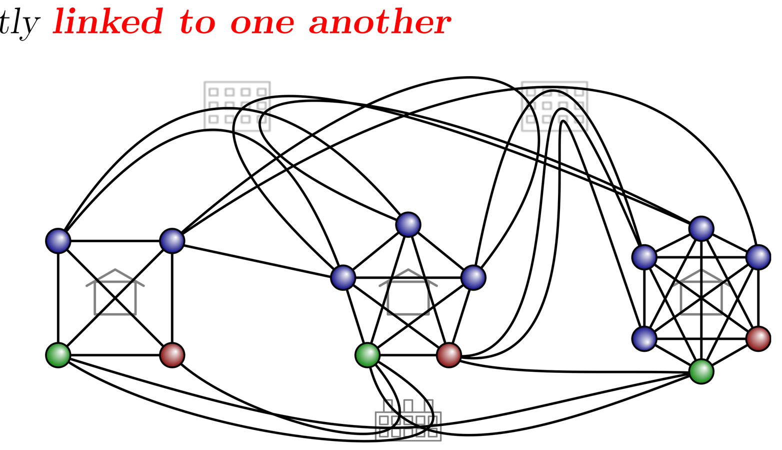
- *Composed of* ***individuals*** *or* ***groups***
- *Individuals and groups are* ***differentiated into categories*** *by assigning each to a specific type*
  (*e.g.* individuals : gender, age; groups : households, schools)
- *There are M types of individuals and Λ types of groups*
- *Individuals can be* ***linked to other individuals*** *(e.g. to model friendship) and can be* ***linked to groups they belong to***



- $P_i(\boldsymbol{k}, \boldsymbol{\xi})$ *: probability for a type-i individual to be linked to $\boldsymbol{k}$ individuals and $\boldsymbol{\xi}$ groups*
  (*i.e.* k_i type-i individuals and ξ_κ type-κ groups $\forall\, i, \kappa$)
- w_i *: fraction of individuals that are of type-i*

Collapsed network

- *Individuals sharing* ***a common group*** *have a probability to be directly* ***linked to one another***



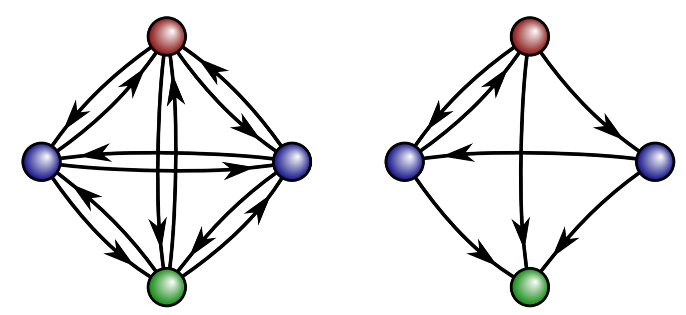
- *Although the groups shape the structure, they do not appear in the resulting network*

Sub-component size distribution

To ***take advantage of the tractability of the CM****, we must first solve independently for the size distribution of each sub-component type.*

Multitype Clusters

- *Fully connected cluster composed of $\boldsymbol{n}$ nodes (fig. a)*
  (*i.e.* n_i type-i nodes $\forall\, i$)
- *In a type-κ group, $i \to j$ edges exist independently with probability $p_{\kappa ij}$ (fig. b).*
  ($i \to j$: edge followed from a type-i node to a type-j node)



(a) (b)

- *Let us define for $\kappa = 1, \ldots, \Lambda$.*

$$\mathbf{p}_\kappa \equiv \begin{bmatrix} p_{\kappa 11} & p_{\kappa 12} & \cdots & p_{\kappa 1M} \\ p_{\kappa 21} & p_{\kappa 22} & \cdots & p_{\kappa 2M} \\ \vdots & \vdots & \ddots & \vdots \\ p_{\kappa M1} & p_{\kappa M2} & \cdots & p_{\kappa MM} \end{bmatrix}.$$

Component Size Distribution

- $Q_{i\kappa}(\boldsymbol{k}|\boldsymbol{n}; \mathbf{p}_\kappa)$ *: probability of finding a component of $\boldsymbol{k}$ nodes in a type-κ group composed of $\boldsymbol{n}$ nodes and reached from a type-i node.*
  ($\boldsymbol{k}$ nodes: k_i nodes of type-i $\forall\, i$)
- $Q_{i\kappa}(\boldsymbol{k}|\boldsymbol{n}; \mathbf{p}_\kappa)$ *is obtained recursively using*

$$Q_{i\kappa}(\boldsymbol{k}|\boldsymbol{n}; \mathbf{p}_\kappa) = Q_{i\kappa}(\boldsymbol{k}|\boldsymbol{k}; \mathbf{p}_\kappa) \prod_{j,l} \left[\binom{n_j - \delta_{ij}}{k_j - \delta_{ij}} (1 - p_{\kappa lj})^{k_l (n_j - k_j)} \right]$$

and

$$Q_{i\kappa}(\boldsymbol{k}|\boldsymbol{k}; \mathbf{p}_\kappa) = 1 - \sum_{\substack{\boldsymbol{l} = \boldsymbol{\delta_i} \\ |\boldsymbol{l}| < |\boldsymbol{k}|}}^{\boldsymbol{k}} Q_{i\kappa}(\boldsymbol{l}|\boldsymbol{k}; \mathbf{p}_\kappa)$$

from the starting value $Q_{i\kappa}(\boldsymbol{\delta_i}|\boldsymbol{\delta_i}; \mathbf{p}_\kappa) = 1$.

($\boldsymbol{\delta_i} \equiv [\delta_{i1}, \ldots, \delta_{iM}]$ and $|\boldsymbol{k}| \equiv \sum_i k_i$)

Probability Generating Function (PGF)

- $R_\kappa(\boldsymbol{n})$ *: size distribution of type-κ groups*
- *The component size distribution in type-κ groups reached from a type-i node is generated by:*

$$\Theta_{i\kappa}(\boldsymbol{x}; \mathbf{p}_\kappa) = \sum_{\boldsymbol{n}=\boldsymbol{0}}^{\infty} \frac{n_i R_\kappa(\boldsymbol{n})}{\langle n_i \rangle_{R_\kappa}} \left[\sum_{\boldsymbol{k}=\boldsymbol{\delta_i}}^{\boldsymbol{n}} Q_{i\kappa}(\boldsymbol{k}|\boldsymbol{n}; \mathbf{p}_\kappa) \prod_{l=1}^{M} x_l^{k_l - \delta_{\kappa l}} \right]$$

($\langle n_i \rangle_{R_\kappa} \equiv \sum_{\boldsymbol{n}} n_i R_\kappa(\boldsymbol{n})$)

Bond percolation

Using propagation arguments and a PGF formalism [1,3], we obtain topological properties of the network ensemble.

- T_{ij} *: $i \to j$ edge occupation probability (elements of $\mathbf{T}$)*
- $\Theta_{i\kappa}(\boldsymbol{x}; \mathbf{p}_\kappa)$ *becomes $\Theta_{i\kappa}(\boldsymbol{x}; \mathbf{p}_\kappa, \mathbf{T})$ using $T_{ij} p_{\kappa ij}$ instead of $p_{\kappa ij}$*

Percolation threshold

- $\Gamma_{(*j)}^{(\times n)}$ *: average number of neighbouring* [type-n nodes reached following an $\times \to n$ edge] *of a* [type-j node previously reached by an $* \to j$ edge] *(computed using $\Theta_{i\kappa}(\boldsymbol{x}; \mathbf{p}_\kappa, \mathbf{T})$ and $P_j(\boldsymbol{k}, \boldsymbol{\xi})$)*
  ($\times$ and $*$ may refer to group or node types)
- $\mathbf{A}$ *: propagation matrix giving the average number of new nodes reached after a node-to-node translation on the network*
  (built using $\Gamma_{(*j)}^{(\times n)}$ $\forall \times, *$)
- *The* ***phase transition happens at $\det(\mathbf{A} - \mathbf{I}) = \mathbf{0}$****, marking the point where the giant component first appears*

Probability of reaching the giant component

The probability that a randomly chosen node leads to the giant component is given by

$$\mathcal{P} = 1 - \sum_i w_i \left[\sum_{\boldsymbol{k},\boldsymbol{\xi}} P_i(\boldsymbol{k}, \boldsymbol{\xi}) \prod_{l,\nu} \left[1 + (\overrightarrow{a_{jl}} - 1) T_{jl} \right]^{k_l} \left[\Theta_{j\nu}(\overrightarrow{\boldsymbol{b_\nu}}; \mathbf{p}, \mathbf{T}) \right]^{\xi_\nu} \right]$$

($\overrightarrow{\boldsymbol{b_\nu}} \equiv [\overrightarrow{b_{\nu 1}}, \ldots, \overrightarrow{b_{\nu M}}]$)

where $\overrightarrow{a_{ij}}$ and $\overrightarrow{b_{\mu j}}$ are respectively the probability that an outgoing $i \to j$ and $\mu \to j$ edge does not lead to the giant component. Those quantities are obtained by solving

($\mu \to j$: edge followed from a type-j node to a type-μ group)

$$\overrightarrow{a_{ij}} = \sum_{\boldsymbol{k},\boldsymbol{\xi}} \frac{k_i P_j(\boldsymbol{k}, \boldsymbol{\xi})}{\langle k_i \rangle_{P_j(\boldsymbol{k},\boldsymbol{\xi})}} \prod_{l,\nu} \left[1 + (\overrightarrow{a_{jl}} - 1) T_{jl} \right]^{k_l - \delta_{il}} \left[\Theta_{j\nu}(\overrightarrow{\boldsymbol{b_\nu}}; \mathbf{p}, \mathbf{T}) \right]^{\xi_\nu}$$

$$\overrightarrow{b_{\mu j}} = \sum_{\boldsymbol{k},\boldsymbol{\xi}} \frac{\xi_\mu P_j(\boldsymbol{k}, \boldsymbol{\xi})}{\langle \xi_\mu \rangle_{P_j(\boldsymbol{k},\boldsymbol{\xi})}} \prod_{l,\nu} \left[1 + (\overrightarrow{a_{jl}} - 1) T_{jl} \right]^{k_l} \left[\Theta_{j\nu}(\overrightarrow{\boldsymbol{b_\nu}}; \mathbf{p}, \mathbf{T}) \right]^{\xi_\nu - \delta_{\mu\nu}}$$

Giant Component Size and Composition

The fraction of the network occupied by type-i nodes that belong to the giant component is given by

$$\mathcal{S}_i = w_i \left[1 - \sum_{\boldsymbol{k},\boldsymbol{\xi}} P_i(\boldsymbol{k}, \boldsymbol{\xi}) \prod_{l,\nu} \left[1 + (\overleftarrow{a_{jl}} - 1) T_{lj} \right]^{k_l} \left[\Theta_{j\nu}(\overleftarrow{\boldsymbol{b_\nu}}; \mathbf{p}^\dagger, \mathbf{T}^\dagger) \right]^{\xi_\nu} \right]$$

where $\overleftarrow{a_{ij}}$ and $\overleftarrow{b_{\mu j}}$ are respectively the probability that an incoming $j \to i$ and $j \to \mu$ edge does not link to the giant component. Those quantities are obtained by solving

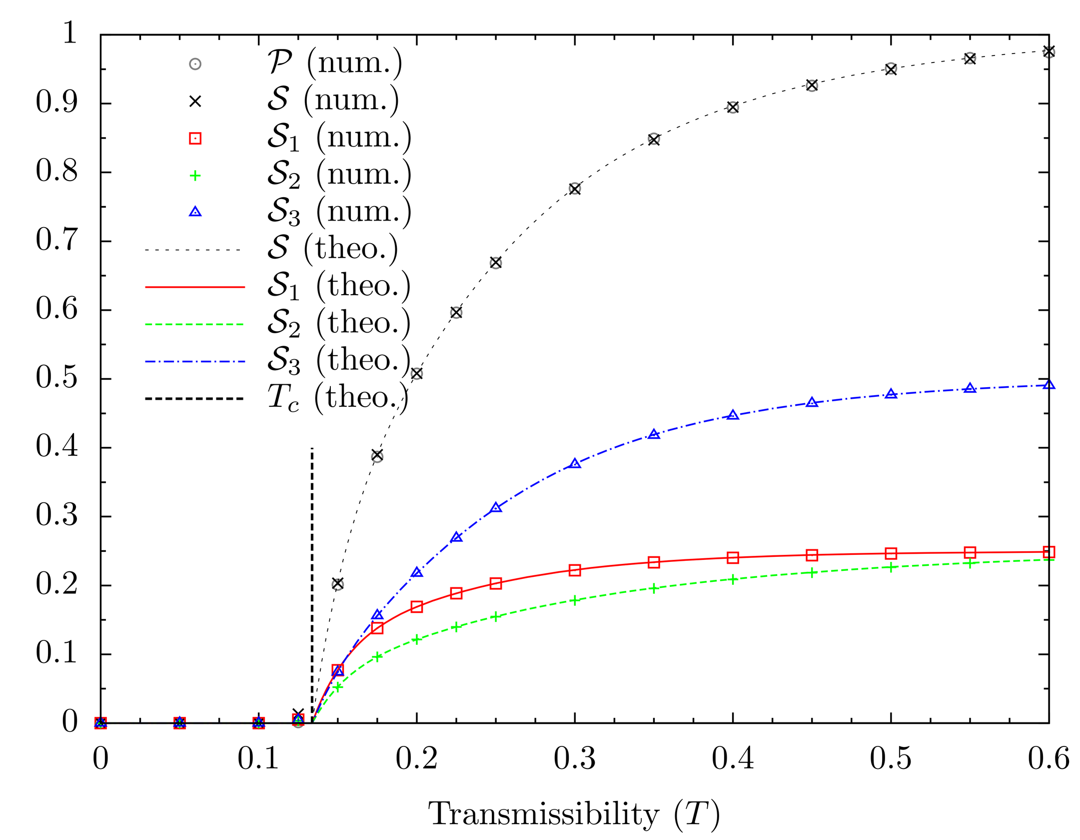
$$\overleftarrow{a_{ij}} = \sum_{\boldsymbol{k},\boldsymbol{\xi}} \frac{k_i P_j(\boldsymbol{k}, \boldsymbol{\xi})}{\langle k_i \rangle_{P_j(\boldsymbol{k},\boldsymbol{\xi})}} \prod_{l,\nu} \left[1 + (\overleftarrow{a_{jl}} - 1) T_{lj} \right]^{k_l - \delta_{il}} \left[\Theta_{j\nu}(\overleftarrow{\boldsymbol{b_\nu}}; \mathbf{p}^\dagger, \mathbf{T}^\dagger) \right]^{\xi_\nu}$$

$$\overleftarrow{b_{\mu j}} = \sum_{\boldsymbol{k},\boldsymbol{\xi}} \frac{\xi_\mu P_j(\boldsymbol{k}, \boldsymbol{\xi})}{\langle \xi_\mu \rangle_{P_j(\boldsymbol{k},\boldsymbol{\xi})}} \prod_{l,\nu} \left[1 + (\overleftarrow{a_{jl}} - 1) T_{lj} \right]^{k_l} \left[\Theta_{j\nu}(\overleftarrow{\boldsymbol{b_\nu}}; \mathbf{p}^\dagger, \mathbf{T}^\dagger) \right]^{\xi_\nu - \delta_{\mu\nu}}.$$

Numerical validation

In order to validate our formalism, we compare its predictions with the results of extensive numerical simulations with

- *3 types of nodes and 3 types of groups*
- $\mathbf{w} = [0.25, 0.25, 0.50]$
- $P_i(\boldsymbol{k}, \boldsymbol{\xi})$ *: every node is linked to one type-1 group, every type-1 node and half of the type-2 nodes are linked to one type-2 group, and every type-3 node is linked to one type-3 group*
- *Each group type have its own size distribution ($R_\kappa(\boldsymbol{n})$) and edge density ($\mathbf{p}_\kappa$)*
- *Uniform transmissibility matrix ($T_{ij} = T \;\forall\, i, j$)*
- 10^5 *nodes (near T_c) and 10^4 nodes (elsewhere)*
- *At least 10^4 simulations for each value of T*



Conclusion

We have introduced a generalized multitype network model that takes into account detailed ***social clustering****. While the underlying structure is* ***analyticaly tractable*** *due to its treelike topology, the modular approach permits the existence of* ***closed loops****. This model will allow to push further our understanding of the impact of* ***non-trivial sub-structures*** *on the global topology of networks and of their influence on* ***propagation dynamics****.*

[1] Allard *et al.*, ***Heterogeneous bond percolation on multitype networks with an application to epidemic dynamics***, **Phys. Rev. E 79, 036113 (2009)**.

[2] Molloy and Reed, *A Critical Point for Random Graphs with a Given Degree Sequence*, Random Struct. Alg. **6**, 161 (1995).

[3] Newman *et al.*, *Random graphs with arbitrary degree distributions and their applications*, Phys. Rev. E **64**, 026118 (2001).

[4] Newman, *Properties of highly clustered networks*, Phys. Rev. E **68**, 026121 (2003).