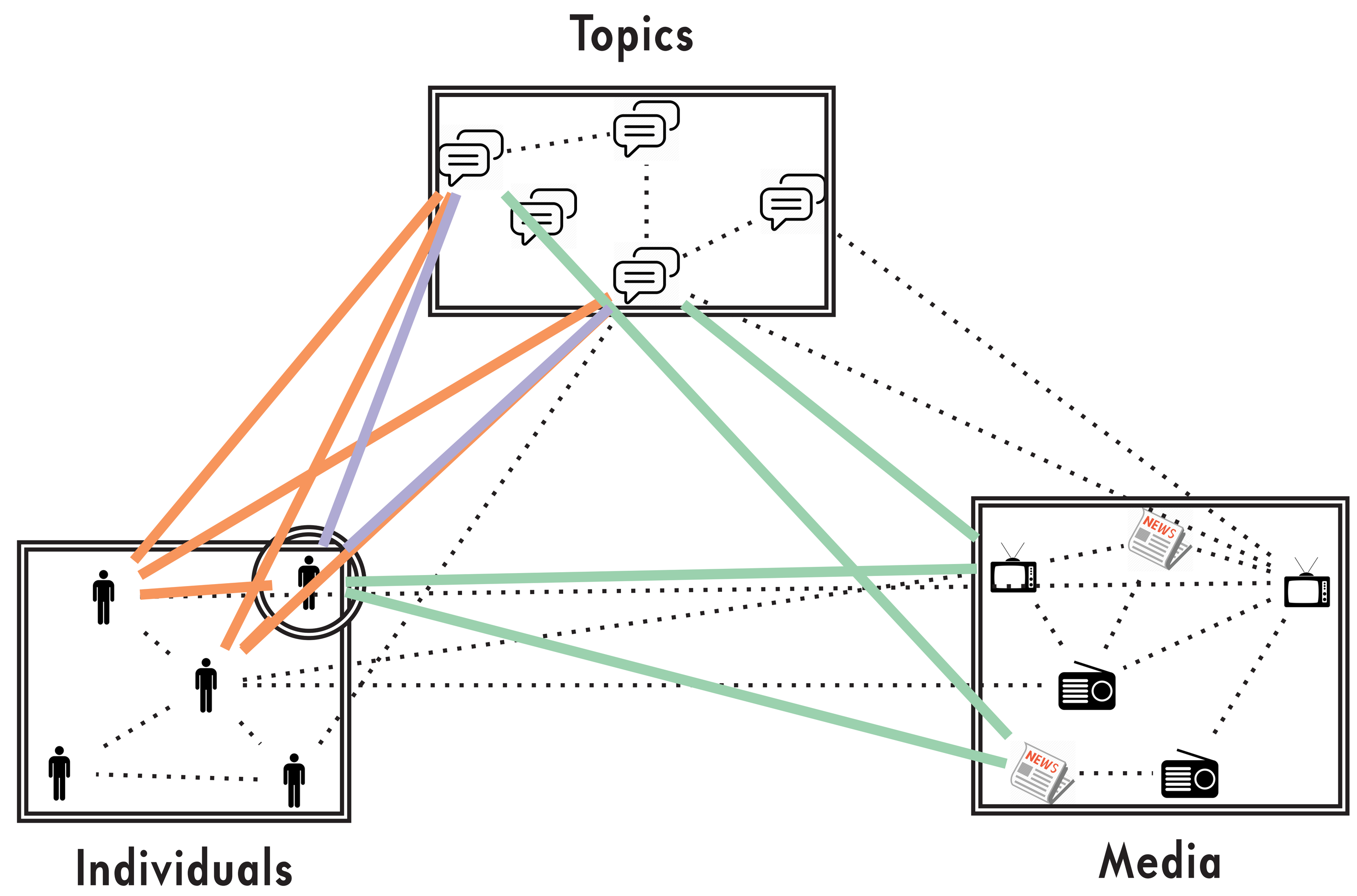
Path-Constrained Random Walks to Reveal Specific Relations in a Graph

Hong-Lan Botterman, Robin Lamarche-Perrin - firstname.lastname@lip6.fr
LIP6 - Sorbonne Université - CNRS - ISC-PIF - Complex Network Team - France

Context

People are subject to social influence as well as mass media influence. Individuals are therefore confronted to different sources of information which can have an impact on the topics they discuss. Such interactions are modeled by networks with different types of entities and links which are well adapted to reflect the different meanings of interactions.

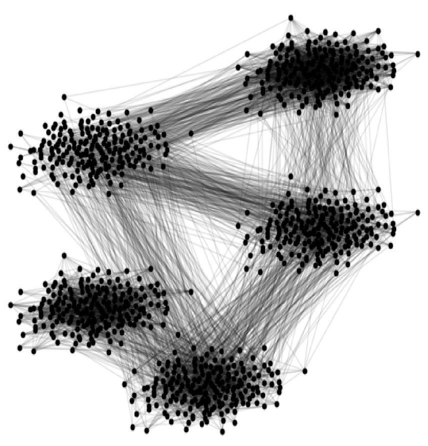
The question is whether the “Individuals-Topics” interactions in social media are random events or whether they are related to other existing interactions between the nodes. We propose a path-constrained random walk to perform statistical regression on networks and thus to unveil the full network topology.



Heterogeneous Information Network

A 7-tuple $G = (V, E, w, \mu_s, \mu_t, \mathcal{V}, \mathcal{E})$

- V set of nodes
- E set of links
- $w : E \to \mathbb{R}$ weight function
- $\mu_s : E \to V$ source node function
- $\mu_t : E \to V$ target node function
- $\mathcal{V}$ partition of V, set of node types
- $\mathcal{E}$ partition of E, set of link types



Furthermore, two functions

$\phi_V : V \to \mathcal{V}$ node type function
$\psi_E : E \to \mathcal{E}$ link type function

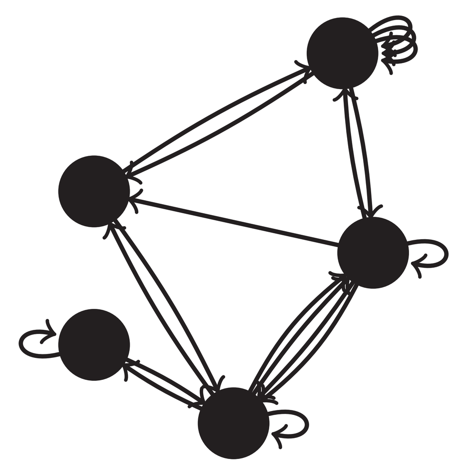
such that $\forall\, e_1, e_2 \in E,\ (\psi_E(e_1) = \psi_E(e_2))$

$$\Rightarrow\ (\phi_V(\mu_s(e_1)) = \phi_V(\mu_s(e_2)) \wedge \phi_V(\mu_t(e_1)) = \phi_V(\mu_t(e_2)))$$

Network Schema

Let the HIN $G = (V, E, w, \mu_s, \mu_t, \mathcal{V}, \mathcal{E})$ and the associated functions ϕ_V and ψ_E.
The associated HIN schema $T_G = (\mathcal{V}, \mathcal{E}, \nu_s, \nu_t)$

$$\nu_s : \mathcal{E} \to \mathcal{V} : E^* \mapsto \nu_s(E^*) := \phi_V(\mu_s(e))$$
$$\nu_t : \mathcal{E} \to \mathcal{V} : E^* \mapsto \nu_t(E^*) := \phi_V(\mu_t(e))$$
$$e \in \psi_E^{-1}(E^*)$$
$$\psi_E^{-1} : \mathcal{E} \to 2^E : E^* \mapsto \{e \in E \,|\, \psi_E(e) = E^*\}$$



Meta Path

A meta path $\mathcal{P}$ of length $n-1 \in \mathbb{N}$ is $\mathcal{P} = V_1 \xrightarrow{E_1} V_2 \cdots \xrightarrow{E_{n-1}} V_n$, where $V_i \in \mathcal{V}, E_i \in \mathcal{E}$.
The inverse meta path is $\mathcal{P}^{-1} = V_n \xrightarrow{E_{n-1}} \cdots V_2 \xrightarrow{E_1} V_1$.
Each meta path is a set of paths.

Method

- Path-constrained random walk
  Let $G = (V, E, w, \mu_s, \mu_t, \mathcal{V}, \mathcal{E})$ and $\mathcal{P} = V_1 \xrightarrow{E_1} \cdots \xrightarrow{E_{n-1}} V_n$.
  $\forall\, v_n \in V_n,\ v_1 \in V_1,$

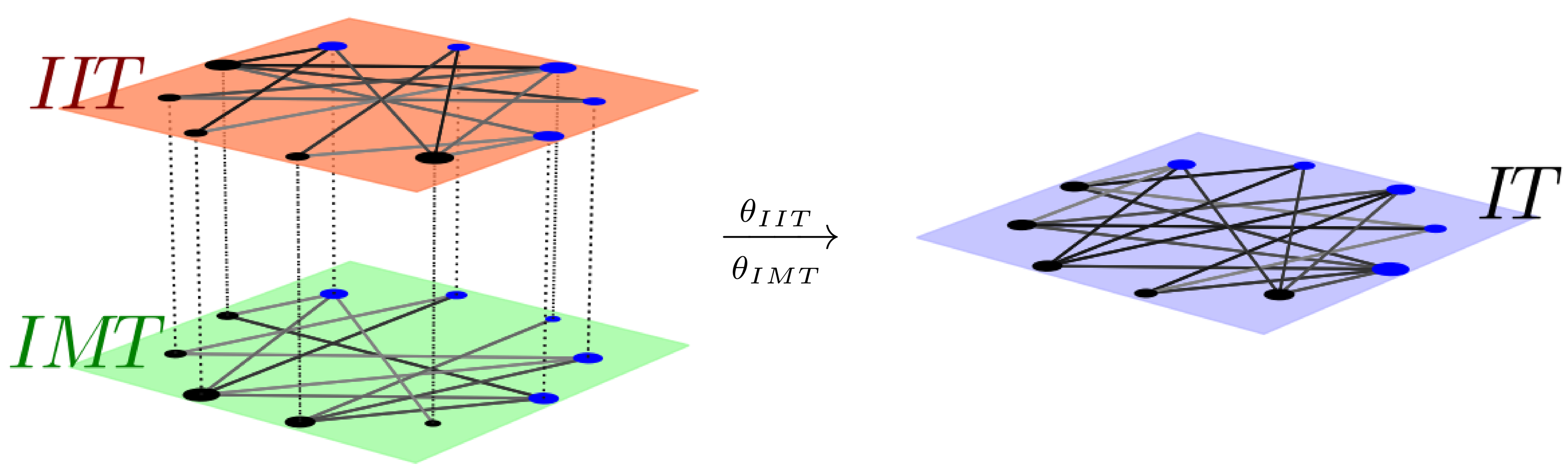
$$\mathbb{P}\big((v_n|v_1)\,|\,\mathcal{P}\big) = \sum_{v_{n-1} \in V_{n-1}} \frac{w_{E_{n-1}}(v_{n-1}, v_n)}{\sum_j w_{E_{n-1}}(v_{n-1}, v_j)} \times \mathbb{P}\Big((v_{n-1}|v_1)\,|\,\mathcal{P}^{1,n-1}\Big)$$

$\mathbb{P}\big((v_2|v_1)\,|\,\mathcal{P}^{1,2}\big)$ basis of recurrence and $\mathcal{P} =: \mathcal{P}^{1,n}$

- Ordinary least squares

$$\Theta^* = \underset{\Theta}{\operatorname{argmin}} \sum_{v_1 \in V_1} \sum_{v_n \in V_n} \sum_{\mathcal{P} \in \mathcal{E}} \Big(\mathbb{P}\big((v_n|v_1)\,|\,\mathcal{P}\big)\,\theta_{\mathcal{P}} - \mathbb{P}\big((v_n|v_1)\,|\,E_c\big)\Big)^2$$

Each metapath gives a bipartite network which is an explanatory variable of the model.



Future Work

- Application to artificial graphs
  → Possibility to derive analytical expressions
- Application to real data
  → Modeling the data by heterogeneous information network
  → Analyzing the possible correlations between the links

References

[1] C. Shi, X. Kong, Y. Huang, P. S. Yu, and B. Wu. Hetesim: A general framework for relevance measure in heterogeneous networks. *IEEE Transactions on Knowledge & Data Engineering*, 26(10):2479–2492, Oct. 2014.

[2] C. Meng, R. Cheng, S. Maniu, P. Senellart, and W. Zhang. Discovering meta-paths in large heterogeneous information networks. In *Proceedings of the 24th International Conference on World Wide Web, WWW '15*, pages 754–764, Republic and Canton of Geneva, Switzerland, 2015. International World Wide Web Conferences Steering Committee.

Acknowledgements

This work is funded in part by the European Commission H2020 FETPROACT 2016-2017 program under grant 732942 (ODYCCEUS), by the ANR (French National Agency of Research) under grant ANR-15-CE38-0002 (AlgoDiv)