# A Model of Opinion Dynamics for Community Detection in Graphs

Irinel Constantin Morarescu, Antoine Girard

LABORATOIRE JEAN KUNTZMANN
MATHÉMATIQUES APPLIQUÉES • INFORMATIQUE

UNIVERSITE JOSEPH FOURIER GRENOBLE 1
SCIENCES.TECHNOLOGIE.SANTÉ

## Community Detection in Graphs

**Communities in a graph are groups of vertices such that the concentration of edges inside communities is high and the concentration of edges between communities is comparatively low.** Community detection has been motivated by the increasing need of understanding complex network structures in social sciences, biology, engineering, economics...

Let $G = (V, E)$ be a graph where $V = \{1, \dots, n\}$ and $E \subseteq V \times V$ is a symmetric relation over $V$. Let $L(G)$ be the normalized Laplacian of the graph $G$. **The second smallest eigenvalue $\mu_2(L(G))$ of the normalized Laplacian can serve as an algebraic measure of the connectivity. We propose a formulation of community detection in terms of eigenvalues of normalized Laplacians.**

Let $\mathscr{C}$ be a partition of the set of vertices $V$. For all $C \in \mathscr{C}$, with $|C| \geq 2$, let $G_C = (C, E_C)$ be the subgraph of $G$ consisting of the set of vertices $C$ and of the set of edges of $G$ between elements of $C$. Let us define the following measure associated to the partition $\mathcal{C}$

$$\underline{\mu_2}(\mathscr{C}) = \min_{C \in \mathscr{C}, |C| \geq 2} \mu_2(L(G_C)).$$

**Problem 1** *Given a real number $\delta \in (0, 1]$, find a partition $\mathscr{C}$ of $V$ such that $\underline{\mu_2}(\mathscr{C}) > \delta$.*

If $\delta \geq \mu_2(L(G))$, **we want to find groups of vertices that are more densely connected together than the global graph. This coincides with the notion of community.** The larger $\delta$ the more densely connected the communities. This makes it possible to search for communities at different scales.

## Opinion Dynamics with Decaying Confidence

**We consider an opinion dynamics over the graph** $G$ inspired by Krause model [2,3]. Let $x_i(t) \in \mathbb{R}$ denote the opinion of agent $i \in V$, agent $i$ updates its opinion by taking into account the opinions of its neighbors that are within some confidence range:

$$x_i(t+1) = \begin{cases} x_i(t) + \frac{\alpha \sum_{j \in N_i(t)} (x_j(t) - x_i(t))}{|N_i(t)|} & \text{if } N_i(t) \neq \emptyset \\ x_i(t) & \text{if } N_i(t) = \emptyset \end{cases}$$

where $N_i(t) = \{j \in V | \ ((i,j) \in E) \wedge (|x_i(t) - x_j(t)| \leq R\rho^t)\}$, $\alpha \in (0, 1/2)$, $R > 0$, $\rho \in (0, 1)$.

Parameter $\rho$ characterizes the confidence decay of the agents. This model can be interpreted in terms of negotiations where **agent $i$ requires that, at each negotiation round, the opinion of agent $j$ moves significantly towards its own opinion in order to keep negotiating with $j$.**

**Proposition 1** *For all $i \in V$, the sequence $x_i(t)$ is convergent. We denote its limit $x_i^*$, we have for all $t \in \mathbb{N}$, $|x_i(t) - x_i^*| \leq R\rho^t/(1-\rho)$. We say that $i$ and $j$ belongs to the same community if $x_i^* = x_j^*$. The set of communities $\mathscr{C}$ is a partition of $V$.*

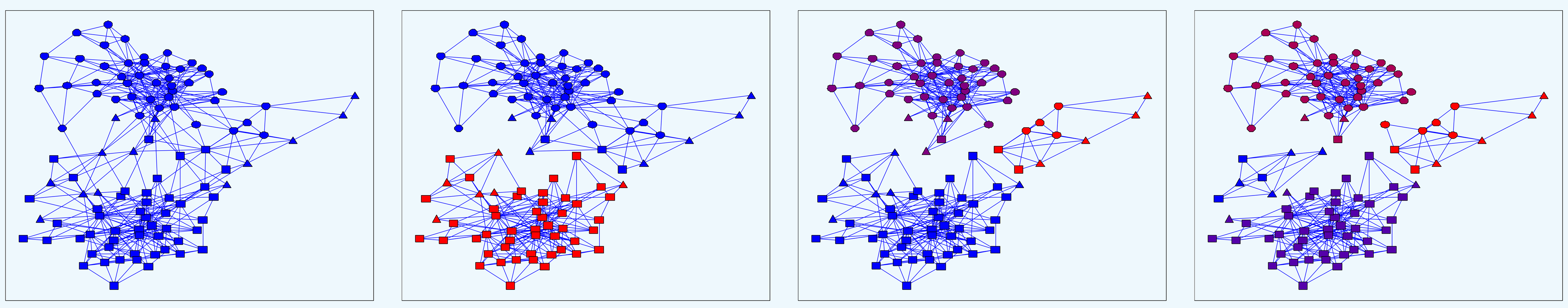
$O(\rho^t)$ is an upper-bound for the convergence speed, in practice the convergence is slightly faster. Let $X^0 \subseteq \mathbb{R}^n$ be the set of initial opinions such that if $x^0 \in X^0$ then there exists $\underline{\rho} < \rho$ such that for all $i \in V$, $|x_i(t) - x_i^*| = O(\underline{\rho}^t)$. Numerical experiments show that in practice $x^0 \in X^0$. The question whether $\mathbb{R}^n \setminus X^0$ is a set of zero measure is open...

**Community detection can be solved using our opinion dynamics model** as follows:

**Proposition 2** *Let $\rho = 1 - \alpha\delta$, for almost all vectors of initial opinions $x^0 \in X^0$, the set of communities $\mathscr{C}$ obtained by the opinion dynamics model is a solution to Problem 1.*

## Experimental Results

**We consider a network of 105 books on politics.** Each vertex represents a book on American politics bought from Amazon.com. An edge between two vertices means that these books are frequently purchased by the same buyer. We represented below the most frequently obtained partitions of the books network for $\delta = 0$, $\delta = 0.1$, $\delta = 0.15$, $\delta = 0.2$. The shape of the vertices represent the political alignment of the book (liberal, conservative, centrist).



Let us remark that **even though the information on the political alignment of the books is not used by the algorithm, our approach allows to uncover this information.** Indeed, for $\delta = 0.1$, we obtain 2 communities that are essentially liberal and conservative. For $\delta = 0.2$, we then obtain 4 communities: liberal, conservative, centrist-liberal, centrist-conservative.

## Comparative Study

Some formalizations of the community detection problem have been proposed in terms of optimization of quality functions such as modularity [4]. **Modularity $\mathsf{Q}(\mathscr{C})$ of a partition $\mathscr{C}$ measures how well the partition reflects the community structure of a graph**:

$$\begin{aligned} \mathsf{Q}(\mathscr{C}) &= \text{fraction of edges within classes of } \mathscr{C} \\ &\quad - \text{expected fraction of such edges.} \\ &= \frac{1}{|E|} \sum_{C \in \mathscr{C}} \sum_{i,j \in C} \left( a_{ij} - \frac{d_i d_j}{|E|} \right) \end{aligned}$$

where $a_{ij}$ are the coefficients of the adjacency matrix of $G$ and $d_i$ is the degree of vertex $i$.

The higher the modularity, the better the partition reflects the community structure of the graph. **Community detection can be formulated as modularity maximization.** Modularity maximization is NP-complete, **approaches for community detection rely mostly on heuristic methods** [1,5].

In the table below, we give a comparative summary of the modularity of the partition obtained for three different networks by our approach and by the algorithms presented in [1,5].

| Network | Karate | Books | Blogs |
|---|---|---|---|
| # of vertices | 34 | 105 | 1222 |
| Our approach | 0.417 | 0.523 | 0.426 |
| [5] | 0.419 | 0.526 | 0.426 |
| [1] | 0.419 | 0.527 | 0.427 |

Though slightly smaller, **the modularity of the partition we obtain is comparable to that of other partitions which is actually surprising since our approach, contrarily to [1,3] does not try to maximize modularity**.

## References

1. V.D. Blondel, J.-L. Guillaume, R. Lambiotte and E. Lefebvre, Fast unfolding of communites in large networks, *J. of Statistical Mechanics: Theory and Experiment*, vol. 1742-5468(08), pp. 10008+12, 2008.
2. V.D. Blondel, J.M. Hendrickx and J.N. Tsitsiklis, On Krause's consensus formation model with state-dependent connectivity. *IEEE Trans. on Automatic Control*, vol. 54(11), pp. 2506-2517, 2009.
3. R. Hegselmann and U. Krause, Opinion dynamics and bounded confidence models, analysis and simulation. *Journal of Artificial Societies and Social Simulation*, vol. 5(3), 2002.
4. M.E.J. Newman and M. Girvan, Finding and evaluating community structure in networks, *Physical Review E*, vol. 69(2):026113, pp. 1-15, 2004.
5. M.E.J. Newman, Modularity and community structure in networks, *Proc. Natl. Acad. Sci. USA*, vol. 103, pp. 8577-8582, 2006.