IFAC

A Model of Opinion Dynamics for Community Detection in Graphs[⋆]

Irinel-Constantin Morărescu [∗] **Antoine Girard** [∗]

[∗] *Laboratoire Jean Kuntzmann, Université de Grenoble,*
BP 53, 38041 Grenoble Cedex 9, France.
irinel-constantin.morarescu@inrialpes.fr, antoine.girard@imag.fr

Abstract: In this paper, we propose a new approach to the problem of community detection in graphs. It is based on a model of opinion dynamics with decaying confidence. This model is a multi-agent system where each agent receives the opinions of its neighbors and then updates its opinion by taking a weighted average of its own opinion and those of its neighbors that are within some confidence range. The confidence ranges are getting smaller at each time step: an agent gives repetitively confidence only to the neighbors that approach sufficiently fast its own opinion. Under that constraint, global consensus may not be achieved and the agents may only reach local agreement organizing themselves in communities. A characterization of these communities is given in terms of eigenvalues of normalized Laplacian matrices of graphs. This shows that our model of opinion dynamics with decaying confidence provides an appealing approach to community detection in graphs. Experimental results show that our approach is also effective. *Copyright* © *IFAC 2010*

Keywords: Community detection, Consensus, Opinion dynamics, Algebraic graph theory.

1. INTRODUCTION

In the usual sense, communities in a graph are groups of vertices such that the concentration of edges inside one community is high and the concentration of edges between communities is comparatively low. Because of the increasing need of analysis tools for understanding complex networks in social sciences, biology, engineering or economics, the community detection problem has attracted a lot of attention in the recent years. The problem of community detection is however not rigorously defined mathematically. One reason is that community structures may appear at different scales in the graph: there can be communities inside communities. Another reason is that communities are not necessarily disjoint. Newman and Girvan (2004) proposed a formalization of the community detection problem in terms of optimization of a quality function called modularity. However, Brandes et al. (2008) have shown that this optimization problem is NP-complete. Therefore, approaches for community detection mostly rely on heuristic methods. We refer the reader to the extensive survey by Fortunato (2009) and the references therein for more details.

In this paper, we propose a new approach to the problem of community detection in graphs. It is based on a model of opinion dynamics with decaying confidence. This model is a multi-agent system where each agent receives the opinions of its neighbors and then updates its opinion by taking a weighted average of its own opinion and those of its neighbors that are within some confidence range. The confidence ranges are getting smaller at each time step: an agent gives repetitively confidence only to the neighbors that approach sufficiently fast its own opinion. Under that constraint, global consensus may not be achieved and the agents may only reach local agreement organizing themselves in groups that we call communities. Communities at different scales can be obtained by decreasing or increasing the confidence decay rate.

The paper is organized as follows. In Section 2, we introduce our model of opinion dynamics with decaying confidence and give a characterization of communities in terms of eigenvalues of the matrices defining the collective dynamics. In Section 3, we propose a new formulation of the community detection problem based on the eigenvalues of normalized Laplacian matrices of graphs. Then, we show that this problem can be solved using our model of opinion dynamics with decaying confidence. Finally, in Section 4, we show on several examples that our model of opinion dynamics not only provides an appealing approach to community detection but that it is also effective.

2. A MODEL OF OPINION DYNAMICS WITH DECAYING CONFIDENCE

In this section we present our model of opinion dynamics with decaying confidence. For brevity, most results are stated without proofs which can be found in Morărescu and Girard (2009).

2.1 Model Description

We consider a set of *n agents*, $V = \{1, \ldots, n\}$. A relation $E \subseteq V \times V$ models the *interactions* between the agents. We assume that the relation is symmetric ($(i, j) \in E$ iff $(j, i) \in E$) and anti-reflexive ($(i, i) \notin E$). V is the set of vertices and E is the set of edges of an undirected graph

$G = (V, E)$, describing the network of agents. Each agent $i \in V$ has an *opinion* modelled by a real number $x_i(t) \in \mathbb{R}$. Initially, agent i has an opinion $x_i(0) = x_i^0$ independent from the opinions of the other agents. Then, at every time step, the agents update their opinion by taking a weighted average of its opinion and opinions of other agents:

$$x_i(t+1) = \sum_{j=1}^{n} p_{ij}(t) x_j(t) \qquad (1)$$

with the coefficients $p_{ij}(t)$ satisfying

$$\forall i, j \in V, \ (p_{ij}(t) \neq 0 \iff j \in \{i\} \cup N_i(t)) \qquad (2)$$

where $N_i(t)$ denotes the *confidence neighborhood* of agent i at time t:

$$N_i(t) = \left\{ j \in V \mid ((i,j) \in E) \wedge (|x_i(t) - x_j(t)| \leq R\rho^t) \right\} \qquad (3)$$

with $R > 0$ and $\rho \in (0, 1)$ model parameters. We make the following assumption:

Assumption 1. For $t \in \mathbb{N}$, the coefficients $p_{ij}(t)$ satisfy

(1) $p_{ij}(t) \in [0, 1]$, for all $i, j \in V$.
(2) $\sum_{j=1}^{n} p_{ij}(t) = 1$, for all $i \in V$.

This model can be interpreted in terms of opinion dynamics. At each time step t, agent $i \in V$ receives the opinions of its neighbors in the graph G. If the opinion of i differs from the opinion of its neighbor j more than the threshold $R\rho^t$, then i does not give confidence to j and does not take into account the opinion of j when updating its own opinion. The parameter ρ characterizes the confidence decay of the agents. Agent i gives repetitively confidence only to neighbors whose opinion converges sufficiently fast to its own opinion. 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.

Our first result states that the opinion of each agent converges to some limit value:

Proposition 1. Under Assumption 1, for all $i \in V$, the sequence $(x_i(t))_{t \in \mathbb{N}}$ is convergent. We denote x_i^* its limit. Furthermore, we have for all $t \in \mathbb{N}$,

$$|x_i(t) - x_i^*| \leq \frac{R}{1-\rho} \rho^t. \qquad (4)$$

Generally, the opinions of all the agents do not converge to a common value. Indeed, agents may only succeed in agreeing locally organizing themselves in communities that we formally define as follows:

Definition 1. Let $i, j \in V$, we say that agents i and j *asymptotically agree*, denoted $i \sim^* j$, if and only if $x_i^* = x_j^*$.

It is straightforward to verify that $\sim^*$ is an equivalence relation over V.

Definition 2. A *community* $C \subseteq V$ is an element of the quotient set $\mathscr{C} = V / \sim^*$.

Let us remark that the community structure is dependent on the initial distribution of opinions. In the following, we shall provide an algebraic characterization of these communities.

Remark 1. We assume that $\rho \in (0, 1)$, however, let us remark that for $\rho = 1$ (there is no confidence decay), with a complete graph G (every agent talks with all the

other agents) and, for all $i \in V$, equal values of non-zero coefficients $p_{ij}(t)$, our model would coincide with Krause model of opinion dynamics with bounded confidence studied in Krause (1997); Hegselmann and Krause (2002); Blondel et al. (2009).

2.2 Notations, Preliminaries and Assumptions

Let us define the set of interactions at time t, $E(t) \subseteq V \times V$ as

$$E(t) = \left\{ (i,j) \in E \mid |x_i(t) - x_j(t)| \leq R\rho^t \right\}.$$

Let us remark that $(i,j) \in E(t)$ if and only if $j \in N_i(t)$. The interaction graph at time t is then $G(t) = (V, E(t))$. Let us also define the *graph of communities* $G_{\mathscr{C}} = (V, E_{\mathscr{C}})$ where:

$$E_{\mathscr{C}} = \left\{ (i,j) \in E \mid i \sim^* j \right\}.$$

We define the vectors of opinions and of initial opinions: $x(t) = (x_1(t), \ldots, x_n(t))^\top$ and $x^0 = (x_1^0, \ldots, x_n^0)^\top$. The dynamics of the vector of opinions is then given by

$$x(t+1) = P(t)x(t)$$

where $P(t)$ is the row stochastic matrix with entries $p_{ij}(t)$. For a set of agents $I \subseteq V$, with $I = \{v_1, \ldots, v_k\}$, we define the vector of opinions $x_I(t) = (x_{v_1}(t), \ldots, x_{v_k}(t))^\top$. Given a $n \times n$ matrix A with entries a_{ij}, we define the $k \times k$ matrix A_I whose entries are the $a_{v_i v_j}$. In particular, $P_I(t)$ is the matrix with entries $p_{v_i v_j}(t)$. Let us remark that $P_I(t)$ is generally not row stochastic. However, if $I \subseteq V$ is a subset of agents such that no agent in I is connected to an agent in $V \setminus I$ in the graph $G(t)$, then it is easy to show that

$$x_I(t+1) = P_I(t)x_I(t)$$

and $P_I(t)$ is a row stochastic matrix. We make the following assumption on the matrices $P(t)$:

Assumption 2. The sequence of matrices $P(t)$ satisfy the following conditions:

(1) For all $t \in \mathbb{N}$, $P(t)$ is invertible.
(2) For all $t, t' \in \mathbb{N}$, if $G(t') = G(t)$ then $P(t') = P(t)$.

Let us remark that the first assumption can be enforced, for instance, by choosing $p_{ii}(t) > 1/2$ for all $i \in V$, for all $t \in \mathbb{N}$, in that case $P(t)$ is a strictly diagonally dominant matrix and therefore it is invertible. The second assumption states that $P(t)$ only depends on the graph $G(t)$, then we shall write $P(t) = P(G(t))$ where $P(G')$ is the matrix associated to a subgraph G' of G.

We now make a last assumption. From Proposition 1, we know that the opinion of each agent converges to its limit value no slower than $O(\rho^t)$. This is an upper bound, the convergence to the limit value is actually often slightly faster. Let $X^0 \subseteq \mathbb{R}^n$ be the subset of vectors of initial opinions such that if $x^0 \in X^0$ then there exists $\underline{\rho} < \rho$ and $M \geq 0$ such that for all $i \in V$, for all $t \in \mathbb{N}$,

$$|x_i(t) - x_i^*| \leq M\underline{\rho}^t.$$

Let us remark that numerical experiments show that in practice $x^0 \in X^0$. This observation motivates the following assumption:

Assumption 3. The vector of initial opinions x^0 is an element of X^0.

It should be noted that unlike Assumptions 1, 2, it is generally not possible to check a priori whether Assumption 3 holds.

2.3 Algebraic Characterization of Communities

Let us consider a community $C \in \mathscr{C}$. Then it is clear that, in the graph $G_{\mathscr{C}}$, no agent in C is connected to an agent in $V \setminus C$. Therefore, it follows that $P_C(G_{\mathscr{C}})$ is a row stochastic matrix. Then, let $\lambda_1(P_C(G_{\mathscr{C}})), \ldots, \lambda_{|C|}(P_C(G_{\mathscr{C}}))$ denote the eigenvalues of $P_C(G_{\mathscr{C}})$ with $\lambda_1(P_C(G_{\mathscr{C}})) = 1$ and

$$|\lambda_1(P_C(G_{\mathscr{C}}))| \geq |\lambda_2(P_C(G_{\mathscr{C}}))| \geq \cdots \geq |\lambda_{|C|}(P_C(G_{\mathscr{C}}))|.$$

The following theorem gives a characterization of the communities in terms of the eigenvalues $\lambda_2(P_C(G_{\mathscr{C}}))$ for $C \in \mathscr{C}$.

Theorem 1. Under Assumptions 1 and 2, for almost all vectors of initial opinions $x^0 \in X^0$, for all communities $C \in \mathscr{C}$, such that $|C| \geq 2$, $|\lambda_2(P_C(G_{\mathscr{C}}))| < \rho$.

Proof. We only state the main ideas of the proof that is too long to be stated here completely. The details can be found in Morărescu and Girard (2009). First, under Assumption 1, and since $x^0 \in X^0$, it is possible to show that there exists $T \in \mathbb{N}$ such that for all $t \geq T$, $G(t) = G_{\mathscr{C}}$. Let us assume that there exists $C \in \mathscr{C}$ such that $|\lambda_2(P_C(G_{\mathscr{C}}))| \geq \rho$. Then, since no agent in C is connected to an agent in $V \setminus C$ in $G_{\mathscr{C}}$, it follows from Assumption 2 that for all $t \geq T$, $x_C(t) = P_C(G_{\mathscr{C}})x_C(t)$. Therefore, the rate of convergence of $x_C(t)$ is $|\lambda_2(P_C(G_{\mathscr{C}}))| \geq \rho$ except if $x_C(T)$ and thus $x(T)$ belongs to a specific subspace of zero measure $H_C(G_{\mathscr{C}})$. However, $x^0 \in X^0$ implies that the rate of convergence of $x_C(t)$ is smaller than $\rho < \rho$. Thus $x(T)$ necessarily belongs to $H_C(G_{\mathscr{C}})$. Then, Assumption 2 allows us, by going backward in time, to show that the initial conditions leading to $H_C(G_{\mathscr{C}})$ at time T are included in a set of zero measure (consisting of a countable union of subspaces) that is independent of $G_{\mathscr{C}}$, C and T. $\square$

A stronger version of Theorem 1 would state that the algebraic characterization of communities holds for almost all $x^0 \in \mathbb{R}^n$. To prove this result, we need to establish that $\mathbb{R}^n \setminus X^0$ is a set of zero measure, at least for generic values of ρ. We were not able to prove this result so far; however, experimental results tend to show that it holds in practice. In the following, we use Theorem 1 to address the problem of community detection in graphs.

3. COMMUNITY DETECTION IN GRAPHS VIA OPINION DYNAMICS

The community detection problem has attracted a lot of attention in the recent years. In the usual sense, communities in a graph are groups of vertices such that the concentration of edges inside communities is high with respect to the concentration of edges between communities. Some formalizations of the community detection problem have been proposed in terms of optimization of quality functions such as modularity (Newman and Girvan (2004)).

The modularity of a partition measures how well the partition reflects the community structure of a graph. More precisely, let $G = (V, E)$ be an undirected graph, let $\mathcal{P}$ be a partition of V. Essentially, the modularity $\mathsf{Q}(\mathcal{P})$ of the partition $\mathcal{P}$ is the proportion of edges within the classes of the partition minus the expected proportion of such edges (see Newman and Girvan (2004) for more details). The

higher the modularity, the better the partition reflects the community structure of the graph. Thus, it is reasonable to formulate the community detection problem as modularity maximization. However, Brandes et al. (2008) have shown that this optimization problem is NP-complete. Therefore, approaches for community detection rely mostly on heuristic methods. Newman (2006) proposed a modularity optimization algorithm based on spectral relaxations. Blondel et al. (2008) presented a hierarchical combinatorial approach for modularity optimization. This algorithm which can be used for very large networks, is currently the one that obtains the partitions with highest modularity.

In the following section, we propose an alternative formulation of the community detection problem using a measure of connectivity of graphs given by the eigenvalues of their *normalized Laplacian matrix*.

3.1 Problem Formulation

Let $G = (V, E)$ be an undirected graph with $V = \{1, \ldots, n\}$, with $n \geq 2$. For a vertex $i \in V$, the degree $d_i(G)$ of i is the number of neighbors of i in G. The normalized Laplacian of the graph G is the matrix $L(G)$ given by

$$L_{ij}(G) = \begin{cases} 1 & \text{if } i = j \text{ and } d_i(G) \neq 0, \\ \dfrac{-1}{\sqrt{d_i(G)d_j(G)}} & \text{if } (i,j) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Let us review some of the properties of the normalized Laplacian matrix (see e.g. Chung (1997)). $\mu_1(L(G)) = 0$ is always an eigenvalue of $L(G)$, it is simple if and only if G is connected. All other eigenvalues are real and belong to the interval $[0, 2]$. The second smallest eigenvalue of the normalized Laplacian matrix is denoted $\mu_2(L(G))$. It can serve as an algebraic measure of the connectivity: $\mu_2(L(G)) = 0$ if the graph G has two distinct connected components, $\mu_2(L(G)) = n/(n-1)$ if the graph is the complete graph (for all $i, j \in V$, $i \neq j$, $(i,j) \in E$), in the other cases $\mu_2(L(G)) \in (0, 1]$.

Let $\mathcal{P}$ be a partition of the set of vertices V. For all $I \in \mathcal{P}$, with $|I| \geq 2$, let $G_I = (I, E_I)$ be the subgraph of G consisting of the set of vertices I and of the set of edges of G between elements of I (i.e. $E_I = E \cap (I \times I)$). Let $L(G_I)$ denote the normalized Laplacian matrix of the graph G_I. Let us define the following measure associated to the partition $\mathcal{P}$

$$\underline{\mu_2}(\mathcal{P}) = \min_{I \in \mathcal{P}, |I| \geq 2} \mu_2(L(G_I)).$$

Essentially, $\underline{\mu_2}(\mathcal{P})$ measures the connectivity of the less connected component of $G_{\mathcal{P}}$. We now propose a formulation of the community detection problem:

Problem 1. Given a graph $G = (V, E)$ and a real number $\delta \in (0, 1]$, find a partition $\mathcal{P}$ of V such that for all $I \in \mathcal{P}$, such that $|I| \geq 2$, $\mu_2(L(G_I)) > \delta$ (i.e. $\underline{\mu_2}(\mathcal{P}) > \delta$).

If $\mu_2(L(G)) > \delta$, it is sufficient to choose the trivial partition $\mathcal{P} = \{V\}$. If $\delta \geq \mu_2(L(G))$, then 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 δ the more densely connected the communities. This makes it possible to search for communities at different scales of the graph.

Let us remark that Problem 1 generally has several solutions. Actually, the trivial partition $\mathcal{P} = \{\{1\}, \ldots, \{n\}\}$ is always a solution. In the following, we show how non-trivial solutions to Problem 1 can be obtained using a model of opinion dynamics with decaying confidence. We evaluate the modularity of the partitions we obtain and compare our results to those obtained using modularity optimization algorithms presented in Newman (2006); Blondel et al. (2008).

3.2 A Solution based on Opinion Dynamics

Let $\alpha \in (0, 1/2)$, we consider the model of opinion dynamics with decaying confidence defined by:

$$x_i(t+1) = \begin{cases} x_i(t) + \dfrac{\alpha \sum_{j \in N_i(t)}(x_j(t) - x_i(t))}{|N_i(t)|} & \text{if } N_i(t) \neq \emptyset \\ x_i(t) & \text{if } N_i(t) = \emptyset \end{cases} \tag{5}$$

where $N_i(t)$ is given by equation (3). It is straightforward to check that this model is a particular case of the model given by equations (1) and (2) and that Assumption 1 holds. Moreover, since $\alpha \in (0, 1/2)$ it follows that for all $i \in V$, $t \in \mathbb{N}$, $p_{ii}(t) > 1/2$. Therefore the matrix $P(t)$ is strictly diagonally dominant and hence it is invertible. Also, $P(t) = P(G(t))$, where for a subgraph G' of G, $P(G') = Id - \alpha Q(G')$ where Id is the identity matrix and

$$Q_{ij}(G') = \begin{cases} 1 & \text{if } i = j \text{ and } d_i(G') \neq 0, \\ \dfrac{-1}{d_i(G')} & \text{if } (i, j) \in E', \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

where $d_i(G')$ denotes the degree of i in the graph G'. Therefore, Assumption 2 holds as well.

Before stating the main result of this section, we need to prove the following lemma:

Lemma 1. Let $\mathcal{P}$ be a partition of V, $I \in \mathcal{P}$ such that $|I| \geq 2$. Then, λ is an eigenvalue of $P_I(G_{\mathcal{P}})$ if and only if $\mu = (1 - \lambda)/\alpha$ is an eigenvalue of $L(G_I)$.

Proof. First, let us remark that $P_I(G_{\mathcal{P}}) = Id - \alpha Q(G_I)$ where $Q(G_I)$ is defined as in equation (6). Then, let us introduce the matrices $R(G_I)$ and $D(G_I)$ defined by

$$R_{ij}(G_I) = \begin{cases} \dfrac{1}{\sqrt{d_i(G_I)}} & \text{if } i = j \text{ and } d_i(G_I) \neq 0, \\ \dfrac{-1}{d_i(G_I)\sqrt{d_j(G_I)}} & \text{if } (i, j) \in E_I, \\ 0 & \text{otherwise.} \end{cases}$$

and

$$D_{ij}(G_I) = \begin{cases} \sqrt{d_i(G_I)} & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

Then, let us remark that $L(G_I) = D(G_I)R(G_I)$ and $Q(G_I) = R(G_I)D(G_I)$. It follows that $L(G_I)$ and $Q(G_I)$ have the same eigenvalues. The stated result is obtained from the fact that $Q(G_I) = (Id - P_I(G_{\mathcal{P}}))/\alpha$. $\square$

We now state the main result of the section which is a direct consequence of Theorem 1 and Lemma 1:

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 (5) is a solution to Problem 1.

4. CASE STUDIES

In this section, we propose to evaluate experimentally the validity of our approach on three benchmarks taken from Newman (2006).

4.1 Case Study 1: Zachary Karate Club

We propose to evaluate our approach on a standard benchmark for community detection: the karate club network initially studied by Zachary (1977). This is a social network with 34 agents shown on the top left part of Figure 1. The original study shows the existence of two communities represented on the figure by squares and triangles.

We propose to use our opinion dynamics model (5) to uncover the community structure of this network. We chose 4 different values for δ. The parameters of the model where chosen as follows: $\alpha = 0.1$, $R = 1$ and $\rho = 1 - \alpha\delta$. For each different value of δ, the model was simulated for 1000 different vectors of initial opinions chosen randomly in $[0, 1]^{34}$. Simulations were performed as long as enabled by floating point arithmetics.

In Table 1, we indicate the partitions in communities that are the most frequently obtained after running the opinion dynamics model. For each partition $\mathscr{C}$, we give the number of communities in the partition, the measure $\mu_2(\mathscr{C})$, this value being greater than δ indicates that Problem 1 has been solved. We computed the modularity $\mathsf{Q}(\mathscr{C})$ in order to evaluate the quality of the obtained partition. We also indicate the number of times that each partition occurred over the 1000 simulations of the opinion dynamics model.

We can check in Table 1 that all the partitions are solutions of Problem 1. Let us remark that in general the computed partition depends on the initial vector of opinions, this is the case for $\delta = 0.3$ and $\delta = 0.4$.

However, it is interesting to note that the partitions that are obtained for the same value of parameter δ have modularities of the same order of magnitude which seems to show that these are of comparable quality. The partition with maximal modularity is obtained for $\delta = 0.4$, it is a partition in 4 communities with modularity 0.417. As a comparison, algorithms in Newman (2006); Blondel et al. (2008) obtain a partition in 4 communities with modularity 0.419. This shows that our approach not only allows to solve Problem 1 but also furnishes partitions with a good modularity which might seem surprising given the fact that our approach, contrarily to Newman (2006); Blondel et al. (2008) does not try to maximize modularity.

In Figure 1, we represented the graphs of communities $G_{\mathscr{C}}$ that are the most frequently obtained for the different values of δ. It is interesting to remark that for $\delta = 0.2$

δ	$|\mathscr{C}|$	$\mu_2(\mathscr{C})$	$\mathsf{Q}(\mathscr{C})$	Occurrences
0.1	1	0.132	0	1000
0.2	2	0.250	0.360	1000
0.3	3	0.334	0.399	691
0.3	3	0.363	0.374	283
0.4	4	0.566	0.4174	924

Table 1. Properties of the partitions of the karate club network (1000 different vectors of initial opinions for each value of δ)
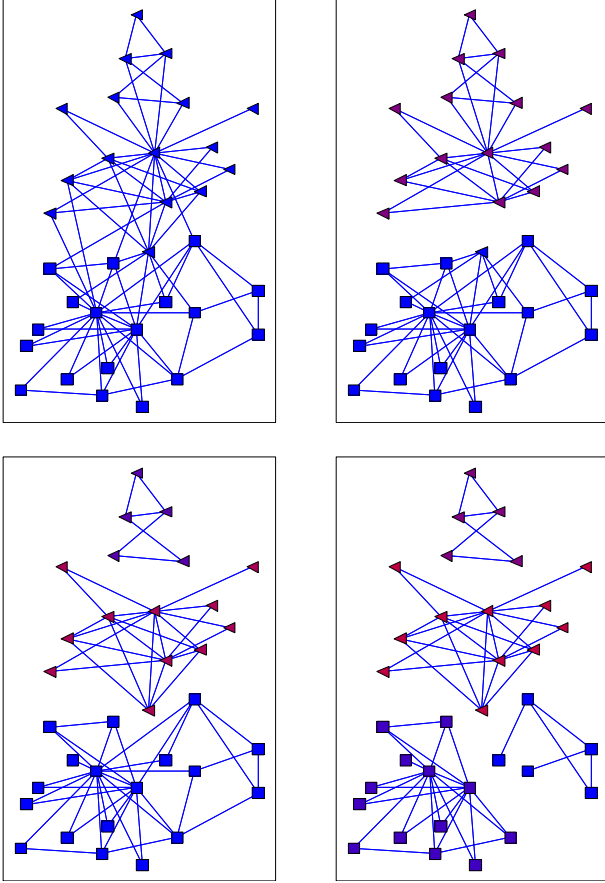
Fig. 1. Graphs $G_{\mathscr{C}}$ for the most frequently obtained partition of the karate club network for $\delta = 0.1$ (top left), $\delta = 0.2$ (top right), $\delta = 0.3$ (bottom left), $\delta = 0.4$ (bottom right).

δ	$|\mathscr{C}|$	$\mu_2(\mathscr{C})$	$Q(\mathscr{C})$	Occurrences
0.1	2	0.134	0.457	980
0.15	3	0.182	0.499	898
0.15	3	0.187	0.494	102
0.2	4	0.269	0.523	678
0.2	4	0.266	0.512	218

Table 2. Properties of the partitions of the books network (1000 different vectors of initial opinions for each value of parameter δ)

we almost obtained the communities that were reported in the original study Zachary (1977). Only one agent has been classified differently. One may argue that this agent has originally 4 neighbors in each community so it could be classified in one or the other. It is also interesting to see that our approach allows us to search for communities at different scales of the graph. When δ increases, the communities become smaller but more densely connected.

4.2 Case Study 2: Books on American Politics

We propose to use our approach on a larger example consisting of a network of 105 books on politics initially compiled by V. Krebs (unpublished, see `www.orgnet.com`). In this network, 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. The network is presented on the top left part of Figure 2 where the shape of the vertices represent the political alignment of the book (liberal, conservative, centrist).

We used our opinion dynamics model (5) to uncover the community structure of this network. We chose 3 different values for δ. The parameters of the model are the same than in the previous example: $\alpha = 0.1$, $R = 1$ and $\rho = 1 - \alpha\delta$. For each different value of δ, the model was simulated for 1000 different vectors of initial opinions

chosen randomly in $[0, 1]^{105}$. Simulations were performed as long as enabled by floating point arithmetics. The experimental results are reported in Table 2.

Let us remark that the computed partitions are solutions to the Problem 1. Also, for the same value of parameter δ, the modularity is very similar for all partitions. Actually, all the partitions obtained for the same value of δ are almost the same. The partition with maximal modularity is obtained for $\delta = 0.2$, it is a partition in 4 communities with modularity 0.523. As a comparison, algorithms in Newman (2006) and Blondel et al. (2008) obtain partitions in 4 communities with modularity 0.526 and 0.527, respectively. As we can see, our partition has a modularity that is quite close from those obtained by these algorithms.

In Figure 2, we represented the graphs of communities $G_{\mathscr{C}}$ that are the most frequently obtained for the different values of δ. 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.

4.3 Case Study 3: Political blogs

The last example we consider consists of a significantly larger network of 1222 political blogs (Adamic and Glance (2005)). In this network, an edge between two vertices means that one of the corresponding blogs contained a hyperlink to the other on its front page. We also have the information about the political alignment of each blog based on content: 636 are conservative, 586 are liberal.

We used 17 values of δ between 0.05 and 0.75. The parameters of the model are the same than in the previous examples: $\alpha = 0.1$, $R = 1$ and $\rho = 1 - \alpha\delta$. For each value of δ, the model was simulated only once for a vector of initial opinion chosen randomly in $[0, 1]^{1222}$. Simulations were performed as long as enabled by floating point arithmetics.

The partition with maximal modularity was obtained for $\delta = 0.4$. It is a partition in 12 communities with modularity 0.426. There are 2 main communities: one with 653 blogs, from which 94% are conservative, and one with 541 blogs, from which 98% are liberal. The 28 remaining blogs are distributed in 10 tiny communities. When we progressively increase δ, we can see that the size of the two large communities reduces moderately but progressively until $\delta = 0.65$ where the conservative community splits into several smaller communities, the largest one containing 40 blogs. The liberal community
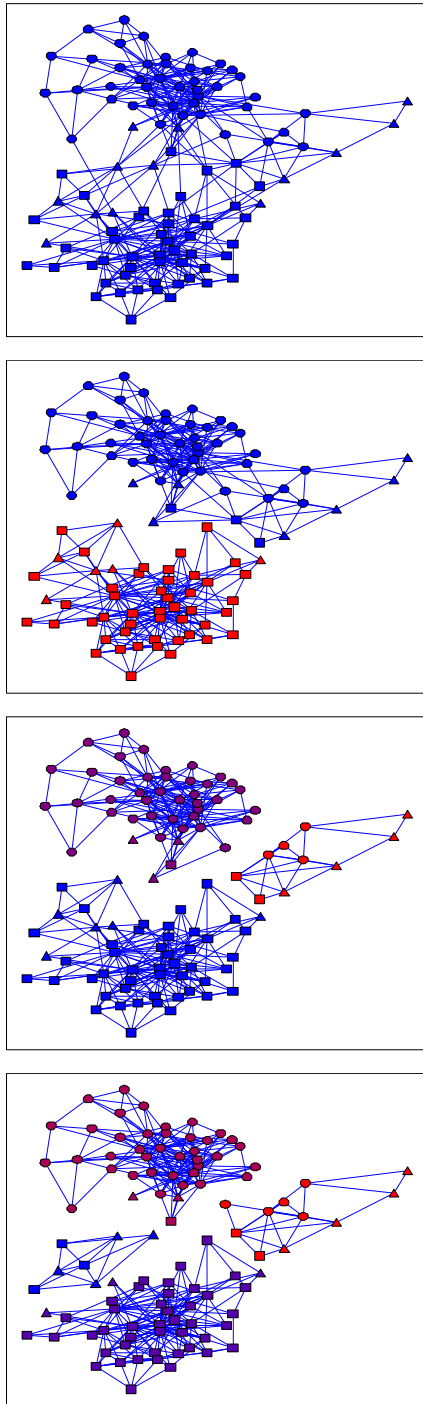
Fig. 2. Graphs $G_\mathscr{C}$ for the most frequently obtained partition of the books network from top to bottom: initial graph, $\delta = 0.1$, $\delta = 0.15$, $\delta = 0.2$. Shapes represent political alignment of the books: circles are liberal, squares are conservative, triangles are centrist.

remains until $\delta = 0.725$ where it splits into smaller communities, the largest one containing 54 blogs.

As a comparison, algorithm in Newman (2006) obtains a partition in 2 communities with modularity 0.426 whereas algorithm in Blondel et al. (2008) obtains a partition in 9 communities with modularity 0.427. As we can see, the partition we obtain is very acceptable in terms of modularity.

5. CONCLUSION AND FUTURE WORK

In this paper, we introduced a model of opinion dynamics with decaying confidence where agents may only reach local agreements organizing themselves in communities. We have provided an algebraic characterization of communities in terms of eigenvalues of the matrix defining the collective dynamics. The detailed analysis of the model can be found in Morărescu and Girard (2009). To complete the analysis of our model, future work should focus on proving (or disproving) that the set of initial opinions $\mathbb{R}^n \setminus X^0$ is a set of zero measure.

We have applied our opinion dynamics model to address the problem of community detection in graphs. We believe that this new approach offers an appealing interpretation of community detection: communities are sets of agents that succeed to reach an agreement under some convergence rate constraint. We have shown on three examples that this approach is not only appealing but is also effective.

REFERENCES

L. Adamic and N. Glance, The political blogosphere and the 2004 U.S. election: divided they blog, *Conf. on Knowledge Discovery in Data: Proc. of the 3rd int. workshop on Link discovery*, 2005.

V.D. Blondel and J.-L. Guillaume and R. Lambiotte and E. Lefebvre: Fast unfolding of communites in large networks, *Journal of Statistical Mechanics: Theory and Experiment*, vol. 1742-5468(08), pp. 10008+12, 2008.

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.

U. Brandes, D. Delling, M. Gaertler, R. Gorke, M. Hoefer, Z. Nikoloski, and D. Wagner. On modularity clustering. *IEEE Trans. on Knowledge and Data Engineering*, vol. 20(2), pp. 172-188, 2008.

F. Chung: *Spectral Graph Theory*, AMS, 1997.

S. Fortunato: Community detection in graphs. *Physics Reports*, vol. 486, pp. 75-174, 2010.

U. Krause: Soziale dynamiken mit vielen interakteuren. Eine problemskizze. In *Modellierung und Simulation von Dynamiken mit vielen interagierenden Akteuren*, pp. 37-51, 1997.

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.

I.-C. Morărescu and A. Girard: Opinion dynamics with decaying confidence: application to community detection in graphs. ArXiv, 0911.5239, 2009.

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.

M.E.J. Newman, Modularity and community structure in networks, *Proc. Natl. Acad. Sci. USA*, vol. 103, pp. 8577-8582, 2006.

W.W. Zachary: An information flow model for conflict and fission in small groups. *Journal of anthropological research*, vol. 33(4), pp. 452-473, 1977.