Institut Francophone International

Rapport de la partie théorique

TRAVAIL PERSONNEL ENCADRE

Sujet : “Diffusion d’opinions dans les réseaux sociaux : l’évacuation d’une foule”

Encadrement	: Prof. HO Tuong Vinh (IFI)
	: Prof. Dominique LONGIN (IRIT)
Étudiante	: DAO Thuy Hong
Promotion	: 20

Contents

1. Analyse de sujet :

Aujourd'hui, les réseaux sociaux ont un rôle de plus en plus important dans la vie des personnes. Dans ces réseaux, chaque personne peut donner ses points de vue, ses pensées, exprimer ses gouts, se connecter à d'autres personnes, etc.

En général, dans un réseau social, les gens peuvent donner leurs opinions. Ici, le mot "opinion" ne fait pas référence à l'attitude mentale d'une personne, mais plutôt à l'expression son attitude mentale. Et cette expression de chaque personne dépend de son entourage . Ceci peut être étendu à une foule (on parle alors de « l'opinion d'une foule ») comme l'expression d'un groupe de personnes où les opinions individuelles ont été agrégées selon une function donnée. Très souvent, la function choisie est la majorité. C'est la psychologie de masse. À l'égard d'un événement, si des personnes plébiscitent A, alors un individu sera tenté de plébisciter A. À l'opposé, l'opinion d'une personne n'est pas constante. Elle peut changer au fil du temps. Par exemple, à l'instant t1, "Ah mes amis choisissent A, donc je vais choisir A", à l'instant t2 "Humm... beaucoup de gens choisissent B, B semble meilleur, je vais choisir B", etc.

Mon sujet se situe sous le domaine intelligence artificielle (plus précisément : domaine de la diffusion d'opinion). Cette étude s'intéresse au processus selon lequel chaque personne construit et forme ses opinions. On va étudier comment l'opinion d'une personne évolue au fil du temps en raison de l'influence des autres dans les réseaux sociaux. Donc, c'est un sujet de recherche.

Ici, la difficulté est qu'il y a beaucoup de choses avec des formes variées qui concernent ce processus. On doit les chercher et modéliser. En outre, on doit aussi modéliser des opinions pour faciliter des traitements.

Pour démontrer et évaluer des idées que j'obtiens, je vais implémenter un modèle sous Java et puis implémenter ce modèle dans l'architecture GAMA. Sous Java, je vais ajouter des différents pramètres que j'aurai étudiés puis analyser les résultats. Dans l'architecture GAMA, d'abord, chaque agent a un état initial (des opinions initiales). Et puis, on va utiliser le modèle mathématique de mécanisme de diffusion que nous avons déjà mentionné ci-dessus pour déterminer l'état suivant.

2. État de l'art :

Dans mon sujet, il est important de faire un modèle des mécanismes de diffusion et de formation d'opinion de l'individu en raison de l'influence des autres sur un agent donné. Dans des travaux antérieurs, plusieurs modèles sont proposés. Le but de ces modèles est d'étudier le processus selon lequel chaque personne construit et forme ses opinions. Ces modèles sont classés en deux familles (ou types), à savoir les modèles discrets et les modèles continus.

Pour profiter des expériences et des idées des autres, j'ai lu des articles qui concernent les deux types de modèles mentionnés ci-dessus. Afin de présenter les résultat de mes lectures je vais d'abord présenter en détail ces modèles, puis je vais donner quelques points de vue personnels sur ces modèles.

2.1 Modèles discrets :

2.1.1 Introduction :

Ces modèles sont étudiés depuis longtemps. Thosmas C.Schelling [1] et Mark Granovetter [2] en 1978 ont été parmi les premiers à les proposer. Dans ces modèles, chaque agent a seulement une opinion de type oui ou non (décision binaire). Les modèles discrets qui sont cités dans cette partie sont le modèle à seuil et le modèle à seuil linéaire.

2.1.2 Modèle à seuil :

Ce modèle a été proposé par Schelling [1] et Granovetter [2] en 1978. Ils ont observé la façon dont se comportent les gens dans une foule. Ils ont observé que dans une foule, le choix de chaque personne est influencé par celui des autres. À partir de cette observation, ils ont établi le modèle à seuil afin de prédire le choix de chaque individu quand il se retrouve au sein d'une foule.

Dans ce modèle, chaque individu est associé à des caractéristiques personnelles (ses buts ou motivations, ses préférences, ses connaissances, etc.). Parmi ses préférences, tout individu dispose d'un « seuil d'influence ». En pratique, un individu adopte une opinion à condition que le nombre d'individus dans son entourage ayant choisi cette opinion atteint ou dépasse ce seuil d'influence (qui peut varier d'un individu à l'autre).

2.1.3 Modèle à seuil linéaire :

Ce modèle (voir par exemple [3] et [4]) est une généralisation de modèle à seuil.

Dans ce modèle, chaque nœud est un individu. Un nœud sera actif si ce nœud adopte l'opinion, et inactif sinon. On suppose qu'une fois qu'un nœud devient actif, il le reste. Un nœud v est influencé par chaque voisin w en fonction d'un poids $b_{v,w}$ et le total des $b_{v,w}$ (w est un voisin de v) ≤ 1. Ici, le poids est la force de la relation d'influence. Chaque nœud v choisit un seuil θ_v au hasard dans l'intervalle [0; 1], cela représente la fraction de poids des voisins de v qui doivent devenir actifs pour v. Étant donné le choix aléatoire des seuils, et un ensemble initial de nœuds actifs A_0 (avec tous les autres nœuds inactifs), le processus de diffusion se déroule par étapes discrètes: à l'étape t, tous les nœuds qui étaient actifs à l'étape $t - 1$ restent actifs, et on active le nœud v dont le total de poids de ses voisins actifs est au moins θ_v.

2.2 Modèles continus :

2.2.1 Introduction :

Contrairement aux modèles discrets, dans des modèles continus, chaque individu peut-être avoir des opinions continues sur le problème donné (l'opinion n'est pas seulement oui ou non). Chaque opinion va être représentée par une valeur numérique. Beaucoup de modèles continus ont été proposés comme par exemple le modèle de DeGroot [5], le modèle de Chatterjee and Seneta [7], le modèle de confiance bornée [6], le modèle en relation [8]. Ici, je vais présenter deux modèles : celui de DeGroot [5] et celui de la confiance bornée [6].

2.2.2 Modèle de DeGroot [5] :

Comme le modèle de seuil linéaire, le modèle de DeGroot utilise aussi le concept de "poids" pour représenter la force de la relation d'influence. Ce modèle considère un groupe de k individus. Chaque individu peut spécifier sa distribution de probabilité subjective pour la valeur inconnue d'un paramètre θ. F_i est la distribution de probabilité subjective de l'individu i $(i = 1,\ldots,k)$ que celui-ci assigne au paramètre θ. Maintenant, on va considérer comment un individu peut changer sa distribution subjective quand il connait les distributions subjectives des autres dans le groupe. Quand l'individu i sait F_j (pour tout $j \neq i$), il va réviser sa distribution subjective F_i pour tenir compte des avis du reste du groupe.

On suppose que lorsque l'individu i révise sa distribution de cette façon. Alors sa distribution révisée sera une combinaison linéaire des distributions $F_1,\ldots, F_k$ des membres du groupe. Pour $i,j = 1,\ldots,k;$ $p_{ij} > 0$ ($\sum_{j=1}^{k} pij = 1$ pour chaque i) représente le poids que l'individu i assigne à la distribution de l'individu j quand il effectue une révision de ses probabilités subjectives à l'instant t+1, alors

$$F_{i(t+1)} = \sum_{j=1}^{k} pij \times F_j .$$

2.2.3 Modèle de confiance bornée [6] :

Dans l'article de Rainer Hegselmann et Ulrich Krause [6], la formule suivante sert de base de calcul :

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

Avec :
- $x_i(t+1)$: l'opinion x de l'agent i à l'instant $t+1$.
- a_{ij} : le niveau d'influence de l'agent j sur l'agent i.
- n : le nombre d'agent dans le groupe étudié.

Ils considère le concept « de niveau de confiance » ε_i . $I(i,x(t))$ est l'ensemble des agents dont l'opinion ne diffère qu'au plus de ε_i par rapport à l'opinion de i. Soit formellement : $I(i, x) = \{j : 1 \leq j \leq n \mid |x_i - x_j| \leq \varepsilon_i \}$, $\varepsilon_i > 0$ est le niveau de confiance accordé aux autres agents par l'agent i. Autrement dit, plus le niveau de de confiance d'un agent i est élevé, plus cet agent va prendre en compte une différence importante de l'opinion des autres par rapport à la sienne, et donc plus le nombre d'agents l'influençant est important Les auteurs considèrent alors que $a_{ij}(x) = |I(i, x)|^{-1}$. Alors, la formation de l'opinion x de l'agent i à l'instant $t+1$ peut-être décrite de la façon suivante :

$$x_i(t+1) = |I(i, x(t))|^{-1} \sum_{j \in I(i,x(t))} x_j(t)$$

Dans cette formule, $|I(i,x(t))|$ est le nombre d'agents influençant i et plus ce nombre est important, plus $|I(i, x(t))|^{-1}$ est petit. Cela signifie que plus il y a d'agents qui influencent i, et plus l'influence de ces agents sur i sera (uniformément) petite.

$|I(i, x(t))|^{-1}$ est donc un facteur de normalisation uniforme : il assure que $x_i (t+1)$ restera dans l'intervalle $[0,1]$ en prenant l'opinion de chaque agent j pris en compte au même niveau que celui des autres agents j' (alors que dans la formule initiale, les a_{ij} permettaient de rendre compte qu'un agent j et un agent j' pouvaient ne pas avoir la même influence). De plus, ici, $|x_i - x_j| \leq \varepsilon_i \Rightarrow -\varepsilon_i \leq x_i - x_j \leq \varepsilon_i \Rightarrow -\varepsilon_{l(left)} \leq x_i - x_j \leq \varepsilon_{r(right)}$ est l'intervalle de confiance symétrique. On non seulement considère l'intervalle de confiance symétrique ($\varepsilon_l = \varepsilon_r$), pas encore l'intervalle de confiance asymétrique ($\varepsilon_l \neq \varepsilon_r$).

2.3 Conclusion :

Il y a plusieurs modèles intéressants proposés. Dans les articles de Schelling et Granovetter, les auteurs ont décrit en détail la psychologie de l'individu dans la foule et donner un mécanisme pour prévenir le choix de l'individu dans une foule. Mais leurs modèles peuvent seulement s'appliquer aux opinions binaires. Dans les réseaux sociaux, les opinions sont en général plus compliquées. Les modèle peuvent utiliser pour des opinions variées et ils ont modélisé mathématiquement la façon dont l'opinion d'une personne évolue au fil du temps en raison de l'influence des autres.

Je pense que les modèles continus conviennent mieux pour mon sujet parce que les opinions dans les réseaux sociaux sont très compliquées et changent au fil de temps. Tandis que des modèles discrets s'appliquent uniquement à des opinions binaires. Mais je trouve aussi que l'analyse de psychologie de masse dans l'article de Schelling peut être utile dans le processus que je modélise (le mécanisme de la diffusion d'opinion).

3. Solution proposé :

Dans la partie précédente (l'état de l'art), j'ai déjà présenté quatre modèles de diffusion d'opinion. Tous ces modèles ont abordé une notion de "seuil". Un individu va changer son opinion si le résultat de la fonction estimative atteint son seuil. Dans le premier modèle à seuil, le paramètre qui participe à la fonction estimative est le nombre d'individus dans l'entourage de l'agent considéré. Dans les trois modèles restants (modèle à seuil linéaire, modèles continus et modèle de confiance bornée), le paramètre principal est le poids qui représente la force de la relation d'influence entre deux individus (Dans ce dernier modèle, le poids représente le niveau de confiance). Mon sujet concerne le processus d'évolution de l'opinion d'un agent au fil du temps en fonction de l'opinion des autres agents considérés au sein d'un réseau social. Pour proposer une solution adéquate, je vais modéliser à la fois un modèle d'influence d'une foule basé sur le modèle de seuil et modéliser une modèle de confiance (c'est-à-dire l'influence des personnes qui sont connectées à l'individu considéré) en améliorant les trois derniers modèles.

3.1 Modèle de réseau d'influence :

Dans un réseau social, entre utilisateurs, il y a une multitude de relations et interactions. Un utilisateur connaît difficilement l'opinion de tous les autres. À un instant donné, il peut recevoir juste l'opinion de certains agents et pas d'autres. Dans le cadre d'une application de ce travail à l'évacuation des personnes, on va considérer que les agents dont l'opinion est connue par l'agent *i* sont ses *voisins*, c'est-à-dire les agents qui sont géographiquement/spatialement proches de *i*. Et comme c'est le foule au sein d'une foule, ses voisins peuvent changer au fil du temps. À partir de cette remarque, on a la définition suivante.

Définition 1 **:** *On appelle A l'ensemble de tous les agents considérés et T l'ensemble fini des indices temporels (T = {1, 2, … N}). Dans ces conditions, $V_i(t)$ est l'ensemble des voisins de l'utilisateur i à l'instant t. C'est donc une fonction qui à chaque agent i de A et à chaque instant t indique l'ensemble des voisins de i et elle est définie de la manière suivante :*

- $V_i(0) \subseteq A$ *est un ensemble d'agents de A tiré aléatoirement pour chaque agent i*

- *$V_i(t) = P_i(t-1) \cup Q_i(t-1)$ pour tout $t > 0$, où $Pi(t) \subseteq Vi(t)$ et $Qi(t) \subseteq A \setminus Vi(t)$.*

  En d'autres termes, les voisins de i à l'instant t>0 sont composés en partie d'agents qui étaient déjà ses voisins l'instant d'avant, et en partie d'agents qui ne l'étaient pas. (Cette hypothèse est tout à fait intuitive dans le cadre d'une foule où les voisins de quelqu'un changent partiellement : certains marchent à la même vitesse que soi, mais d'autres arrivent, partent, sont bousculé et poussé ailleurs, etc.)

$V(t) = \{V_i(t) \mid 1 \leq i \leq n\}$ est l'ensemble de toutes les relations de voisinage des agents de A à l'instant t.

Définition 2 ***:*** *On appelle $I_i(t)$ l'ensemble des **influenceurs** de l'utilisateur i à l'instant t. (Par influenceur, on entend les agents dont l'opinion va influencer celle de l'agent i à l'instant t.) $I = \{I_i \mid 1 \leq i \leq n\}$ est l'ensemble de toutes les relations d'influence entre agents.*

Hypothèse : *Dans ce qui suit, par soucis de simplification du modèle, on suppose que $I_i(t) = V_i(t)$.*

Autrement dit, on assimile les personnes qui influencent l'opinion de *i* à ses voisins proches. Cette hypothèse est plausible dans le cadre de l'évacuation d'une foule où une personne ne peut voir et interagir qu'avec les personnes directement à côté d'elle. Elle n'est pas restrictive, car techniquement il est facile de définir $I_i(t)$ comme une restriction (au sens mathématique) de $V_i(t)$.

Les relations dans $I(t)$ créent un réseau d'influence à l'instant *t*. Pour modéliser ce réseau, on utilise un graphe partiel orienté irréflexif $G(t) = \langle N, B(t) \rangle$ où l'ensemble *N* des nœuds du graphe est isomorphe à l'ensemble des agents A, et où $B(t) \subseteq \{p_{ij}(t) \in [0,1] : (i,j) \in N \times N\}$ est l'ensemble des arcs reliant les nœuds *i* et *j* avec une probabilité p_{ij}. Les arcs traduisent une relation d'influence pondérée et $p_{ij}(t)$ se lit : l'agent *i* influence l'agent *j* à l'instant *t* avec une probabilité *p*. (Il est important de noter que $B(t)$ est un sous-ensemble de $N \times N$: le graphe n'est pas total, et l'absence d'un arc entre deux agents indique qu'il n'y a pas d'influence entre eux pour le sens considéré.)

Nous sommes donc en mesure de formaliser les influenceurs de l'agent i au temps t à l'aide du graphe d'influence $G(t) = \langle N, B(t) \rangle$:

$$I_i(t) = \{j \in A : p_{ji}(t) > 0\} \; pour \; i, j \in N \; et \; p_{ij}(t) \in B(t)$$

En d'autres termes, les influenceurs de l'agent i à l'instant t sont tous les agents j qui influencent i à l'instant t avec une probabilité non nulle.

Il est intéressant de remarquer que $p_{ij}(t)$ peut être vu comme la confiance accordée par j à l'opinion de l'agent i *à* l'instant t. Un corollaire à cela est que parmi les agents qui influencent i, seuls sont pris en compte ceux ayant un degré de confiance supérieur à 0. Par ailleurs, l'opinion des différents influenceurs ne sera donc pas pris au même degré par i.

La figure 1 montre un exemple de réseau d'interaction simple. Dans ce réseau, il y a trois utilisateurs et on considère qu'aà l'instant t, l'utilisateur 1 influence l'utilisateur 2 avec $p_{12} = 0.45$; l'utilisateur 3 influence l'utilisateur 2 avec $p_{32} = 0.7$ et l'utilisateur 3 influence l'utilisateur 1 avec $p_{31} = 0.63$.

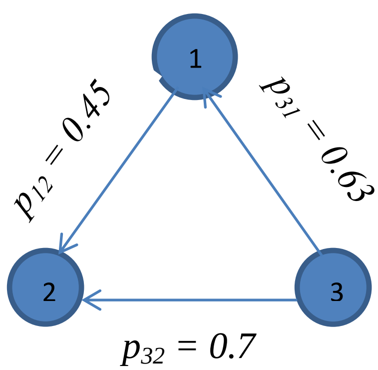

Figure 3.1 : Un réseau d'influence simple

Ainsi, l'agent 1 n'est par exemple pas influence par l'agent 2 (ce qui revient à dire que $p_{21}(t) = 0$). En revanche, 1 est influencé par 3.

3.2 Modèle de problème:

***Définition 5* :** *On définit l'opinion* $o_i(t)$ *de l'agent* i *à l'instant t comme un nombre réel appartenant à [0,1]. Alors,* $o(t) = \{o_i(t): i \in A\}$ *est appelé « profil d'opinion à l'instant t ».*

Il est essentiel de noter qu'à l'instant *t,* un agent i va changer son opinion (ou non) en fonction de deux types de facteurs : les *facteurs internes*, et les *facteurs externes*.

Les facteurs internes sont relatifs aux croyances et aux buts, désirs, normes, *etc*. de l'agent considéré. Dans ce qui suit, on considère que ces facteurs sont synthétisés dans l'opinion de l'agent considéré à l'instant t=0 puisque c'est le seul instant où il n'a pas encore été influencé par les (opinions des) autres. Pour l'agent i, les facteurs internes sont donc représentés par $o_i(0)$.

Les facteurs externes sont des facteurs d'influence de deux types : l'influence des agents en qui i a confiance d'une part, et l'influence de l'opinion de masse, c'est-à-dire de l'opinion de l'ensemble des agents susceptibles de l'influencer, qu'il ait ou non confiance en eux. Par exemple, si on est entouré de personnes en qui nous n'avons pas confiance, mais que toutes vont dans une direction, on peut se laisser quand même influencer par elles en se disant que vu leur nombre, elles ont peut-être raison d'aller dans cette direction.

Par exemple, dans la Figure 1, l'agent 1 a confiance en l'agent 3 mais est susceptible de se laisser influencer par 2 et 3 (la foule dans son ensemble).

On appelle $x_i(t)$ l'opinion de i issue de l'influence des voisins en qui il a confiance. Alors dans ce cas :

$$x_i(t+1) = \frac{\sum_{j \in I_i(t)} p_{ji}(t) o_j(t)}{\sum_{j \in I_i(t)} p_{ji}(t)}$$

Il est important de noter que si $p_{ji}(t) = 0$ alors l'agent j considéré n'entre pas en ligne de compte dans le calcul de $x_i(t)$. Autrement dit, seuls sont pris en compte les agents en qui i a confiance. En plus, le fait de normaliser par la somme des coefficients de confiance nous assure que la valeur de $x_i(t)$ est bien une valeur comprise entre 0 et 1.

Pour le deuxième facteur, je m'inspire de l'observation du processus d'évacuation d'un lieu en proie aux flammes, la direction qu'on choisit de prendre dépend juste des personnes qui se trouvent pas loin de lui-même (ses entourages). Dans le réseau social, cela correspond aux utilisateurs avec qui on est directement relié par la relation influence. Cette influence est donc plutôt reliée à la proximité des individus par rapport à l'individu considéré, indépendamment de la confiance que ce dernier a en eux. Dans le graphe G, pour un nœud i, c'est tous les nœuds qui sont directement liés à i. Soit, formellement :

$$y_i(t+1) = \frac{\sum_{j \in I_i(t)} o_j(t)}{|I_i(t)|}$$

Ici, on prend l'opinion de tous les influenceurs de i, y compris ceux en qui i n'a pas confiance.

3.3 ***Mécanisme d'agrégation des opinions:***

Maintenant, comment choisir une valeur pour l'opinion de l'utilisateur i à l'instant $t + 1$ (c'est-à-dire $o_i(t + 1)$) ? Pour répondre à cette question, je vais calculer la distance entre l'opinion initiale (facteurs internes) et les opinions que je viens de calculer ci-dessus (facteurs externes). On a les deux distances suivantes :

$$\Delta_{x_i}(t + 1) = |o_i(0) - x_i(t+1)|$$
$$\Delta_{y_i}(t + 1) = |o_i(0) - y_i(t+1)|$$

Si $\Delta_{x_i}(t + 1) \leq \Delta_{y_i}(t + 1)$ cela veut dire que l'opinion des influenceurs en qui i a confiance est plus proche des opinions initiales de l'agent i, que ce que l'est l'opinion de la foule des influenceurs. le but de l'utilisateur i que l'opinion de la foule alors $o_i(t+1) = x_i(t+1)$. Dans le cas contraire, on a $o_i(t + 1) = y_i(t + 1)$.

4. Plan de travail pratique :

Afin de faire bien la partie pratique, j'ai fait un plan des travaux pratiques. Dans cette partie, je vais implémenter le modèle que j'ai proposé dans la partie théorique. Et puis j'évaluerai des résultats que j'aurai obtenu.

4.1 Implémentation :

L'implémentation mettre en place tout d'abord en langage Java et puis en langage GAML. Le plan d'implémentation est décris dans le tableau suivant :

Table 4.1 : Plan de travail

Travaux	Temps
Faire la conception	1 semaine
Programmer en langage Java	2 semaines
Implémenter et tester le programme	1 semaine
Progammer en langage GAML	2 semaines
Lancer la simulation et évaluer le résultat obtenu	1 semaines

4.2 Expérimentation :

Par rapport à mon sujet, le résultat attendu est l'agrégation d'opinion après le temps concret. Dans le modèle mathématique, il y a quelques paramètres qui affecte le processus d'agrégation d'opinion comme : valeur d'opinion de chaque personne au début, ensemble des voisins chois à chaque instant, confiance accordée par un agent à l'opinion d'autre agent qui est dans l'ensemble des voisins. Donc je change ces paramètres chaque fois que je lance le programme afin d'observer le processus de changement d'opinion. Pour évaluer plus facilement, des résultats à chaque étape exprimeront par le graphique.

Références :

[1] T. Schelling. *Micromotives and macrobehavior*. Norton, 1978.

[2] M.Granovetter. Threshold models of collective behavior. *American Journal of Sociology, 83(6):1420–1443, 1978.*

[3] D. Kempe, J. M. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. *In Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2003.*

[4] D. Kempe, J. M. Kleinberg, and E. Tardos. Influential nodes in a diffusion model for social networks. *In Proceedings of the 32nd International Colloquium on Automata, Languages and Programming (ICALP-2005), 2005.*

[5] M. H. de Groot. Reaching a consensus. *Journal of the American Statistical Association, 69(345):118–121, 1974.*

[6] R. Hegselmann and U. Krause. Opinion dynamics and bounded confidence models, analysis, and simulations. *Journal of Artificial Societies and Social Simulation, 5(3), 2002.*

[7] S. Chatterjee and E. Seneta. Toward consensus: some convergence theorems on repeated averaging. *Journal of Applied Probability, 14:89–97, 1977.*

[8] G. Deffuant, F. Amblard, G. Weisbuch, and T. Faure. How can extremism prevail? a study based on the relative agreement interaction model. *Journal of Artificial Societies and Social Simulation, 4, 2002.*