

## EFFETS D'UN FILTRE COGNITIF SUR LA DIFFUSION D'INFORMATION

Huet S., Deffuant G.

Cemagref  
B.P. 50085  
63172 Aubière Cedex, France  
Sylvie.Huet@cemagref.fr, Guillaume.Deffuant@cemagref.fr

**RESUME :** *Les recherches sur la diffusion de l'information montrent que les individus ne la retiennent et ne la propagent pas nécessairement. Nous proposons un modèle simple de ce phénomène. Les individus s'échangent des informations relatives à un objet décrit par des facettes dont l'utilité est positive ou négative. La possession d'une information correspond au fait de disposer de l'utilité d'une facette. Les individus disposent par ailleurs d'une opinion globale a priori sur l'objet. Lors des discussions, les facettes dont l'utilité en valeur absolue est inférieure à un seuil d'intérêt sont ignorées et ne sont pas reçues. Ce seuil est plus élevé lorsque l'utilité et l'opinion globale sont de signes contraires. En outre, un individu ne parle pas des facettes qu'il possède lorsque leur utilité est de signe contraire à son opinion globale. L'étude du modèle à convergence, lorsque tous échangent avec tous, permet le recensement des états finaux et souligne l'importance des caractéristiques des premiers individus qui reçoivent l'information. Nous étudions le modèle pour une population dont les individus échangent avec un nombre constant et faible de voisins, dans le cas «une seule information, de même utilité pour tous, un seul individu source initial » grâce à un calcul approché de la distribution de probabilités du nombre final d'individus possédant l'information. Le nombre de voisins modifie cette distribution car il limite la rencontre entre des individus à même de diffuser l'information.*

**MOTS-CLES :** modèle, individus-centré, agrégé, diffusion, information, réseau

### 1. INTRODUCTION

Nous proposons un modèle individus-centré de filtrage cognitif et comportemental de l'information. Ce modèle, très simple, permet d'étudier des hypothèses sur les mécanismes de filtrage individuel de l'information (Deffuant G., et al. 2003). Nous nous inspirons pour ce faire de résultats de recherche en psychologie sociale et des modèles mathématiques existants.

#### 1.1. Inspiration en psychologie sociale

Allport et Postmann étudient la rumeur en 1947 grâce à une expérience durant laquelle un individu décrit une image à un autre individu qui va lui-même la décrire à un suivant et ainsi de suite. Ils mettent en évidence le fait que la liste détaillée décrivant l'image initiale devient, après une succession de transmission, une liste composée des seuls éléments-clés qui demeurent inchangés par les transmissions suivantes. **Les « détails » ont disparus du message.**

Dans sa théorie de la dissonance cognitive, Festinger (1957) prédit *qu'un individu recherche des informations qui confortent ses attitudes<sup>1</sup> et évitent celles qui les remettent en cause.* Eagly A.H. and

Chaiken S. (1993) dans leur ouvrage de synthèse sur la psychologie des attitudes complète cette affirmation en rappelant le principe selon lequel les attitudes motivent le comportement et induisent des effets sélectifs à différents niveaux du traitement de l'information.

Anzieu et Martin (2000, p. 188-210) , dans leur ouvrage de référence sur la dynamique des groupes, donnent une classification des obstacles à la communication. Ils soulignent l'importance des « *éléments de personnalité* » qui renvoient : chez *le locuteur*, au fait qu'il « *présentera les choses à sa façon avec accentuation de certains éléments, en fonction de ses préjugés et stéréotypes...* »... ; chez *l'allocuté*, au fait que « *la déformation subie par le message est fonction des intérêts propres de l'allocuté ; son interprétation est subordonnée à son propre cadre de référence et aux sentiments qu'il prête au locuteur (possibilité de « projection »)* ». D. Anzieu résume ainsi les facilitateurs de la communication : « *l'aptitude à communiquer avec quelqu'un d'autre est d'autant plus grande que les deux personnes auront pensé dans le même univers symbolique et posséderont les mêmes cadres de référence* ».

Nous nous focaliserons donc sur l'individu et sur la modélisation de mécanismes lui permettant de filtrer l'information, qu'il s'agisse de la recevoir ou de l'émettre. Pour ce faire, nous retiendrons d'une part le fait que toute information n'est pas nécessairement intéressante pour un individu, d'autre part, le fait que sa

<sup>1</sup> Selon Eagly A.H. et Chaiken S.(1993), l'attitude est une tendance psychologique se traduisant par une évaluation plus ou moins favorable d'un objet particulier

façon d'appréhender et d'évoquer l'information dépend d'éléments relativement stables (préjugés, stéréotypes, cadres de références) qui lui sont propres.

Nous proposons un modèle simple des filtres cognitifs et comportementaux d'un individu qui régit l'acquisition et la transmission de l'information. Nous considérons qu'une information n'est pertinente pour un individu que si l'utilité qu'il lui attribue est supérieure à un certain seuil. La valeur de ce seuil varie selon que l'information est congruente ou non à l'opinion globale a priori de l'individu. Un individu n'évoque une information connue que si elle est congruente à son opinion globale.

Nous étudions ce modèle individu-centré par simulation et par des calculs approchés de distribution de probabilités des états finaux dans des cas simples. Les états finaux recensés soulignent l'importance des individus initialement en contact avec l'institution qui diffuse l'information. Ils soulignent également l'importance du nombre d'individus avec qui un individu peut discuter ; la discussion étant le support de la transmission de l'information.

## 1.2. Un sujet relativement peu traité en simulation sociale

L'étude de référence d'Allport et Postmann (1947) sur la diffusion d'information dans une chaîne d'individus échangeant verbalement a débouché sur le premier modèle mathématique de diffusion de la rumeur. L'intensité, l'ampleur de cette dernière, c'est à dire sa capacité à se diffuser, y apparaît comme le produit de l'importance de l'information portée par la rumeur pour un receveur potentiel et de l'ambiguïté ou l'incertitude sur cette même information.

Lawson et Butts (2004), déplorant le peu de modèles dynamiques de transmission d'information proposent un modèle qui mesure le taux de diffusion de la description à plusieurs dimensions d'un événement critique auquel des témoins ont assisté. La transmission des dimensions descriptives dépend du fait d'avoir été ou non exposé (directement ou indirectement par l'échange) à chacune de ces dimensions et de leur fréquence d'évocation. Cette fréquence d'évocation correspond à un vecteur de paramètres qui, pour chaque dimension descriptive de l'événement, traduit son importance en général.

Plus simplement, Galam (2003) examine le cas où des individus, lorsqu'ils se rencontrent dans des groupes d'échange, se transmettent l'information avec une probabilité qui dépend du taux de disponibilité de l'information sur toute la population.

Ces modèles ne considèrent généralement que des caractéristiques liées au contexte de la diffusion et non aux individus eux-mêmes. L'importance de l'information n'est pas déterminée par un filtrage de

l'information par les individus ; elle est un paramètre du modèle.

D'autres modèles, tel le « cheap talk » (Crawford V.P., Sobel J., 1982) considèrent un récepteur qui décide de l'action qu'il va mener en fonction du message bruité qui lui a été adressé. L'action retenue affecte le bien-être du receveur et de l'émetteur. Ce message a été choisi par un émetteur en fonction de sa connaissance du problème. Le choix du message est réalisé en même temps que le choix de l'action : le message n'est donc pas retenu en fonction du choix du receveur. Ce modèle mêle transmission d'information et décision, sans pour autant s'intéresser réellement aux mécanismes régissant la transmission.

D'autres travaux s'intéressent à la diffusion d'information via des réseaux sociaux en assimilant le processus à un processus d'épidémie. Dans ce cas, la possibilité de transmettre « l'information/virus » est fonction du caractère « infecté » ou non de l'émetteur, d'une probabilité traduisant le caractère plus ou moins virulent du virus et du caractère « immunisé » ou non du récepteur (Tsimring Lev. S., Huerta R., 2003). Ces modèles ne considèrent pas qu'un récepteur doit être intéressé par un virus pour le recevoir. Ils négligent donc les phénomènes de filtrage qui nous intéressent. D'autres travaux s'inspirant toujours des processus d'épidémie font l'hypothèse d'une proximité entre les individus ayant les mêmes intérêts. La transmission de l'information est ici une fonction qui décroît suivant la distance à la source initiale d'émission (Wu F. et al., 2004). Ici, encore, le caractère plus ou moins important d'une information pour un récepteur potentiel n'est pas réellement pris en compte. De plus, l'extinction de la diffusion est pilotée par un paramètre global du modèle et non par un mécanisme de filtrage propre aux individus.

Deffuant G. et al. (2005), s'intéressant à la diffusion de l'adoption de pratiques novatrices par les agriculteurs, modélise un individu qui reçoit, avec une certaine probabilité, l'information sur la pratique s'il est susceptible d'être intéressé par l'innovation et s'il a l'opportunité de discuter avec quelqu'un qui en dispose. Dans ce modèle, l'individu reçoit l'information, quelle que soit sa valeur, à partir du moment où son opinion sur l'innovation l'incite à s'y intéresser.

Se basant sur les travaux empiriques de Wittek R. et Wielers R. (1998) sur les structures sociales favorisant la rumeur, Snijders T. (1998) propose deux modèles de diffusion de la rumeur. La transmission par A à B d'une information portant sur C dépend de l'existence d'un lien et de la nature de ce lien entre A et C d'une part, entre B et C d'autre part. Ce lien peut être un lien d'amitié ou d'inimitié. La réception par B d'une information sur C dépend du fait que B connaît ou non C, quelle que soit la nature de leur relation. La transmission est ici soumise à un filtre complexe prenant en compte le récepteur

potentiel et sa relation à l'information. Le filtre en réception est en revanche plus simple. L'information n'a pas de valeur propre pouvant être comparée à la nature de la relation de l'individu à l'objet sur lequel porte l'information.

Urbig (2003) propose une brève revue des modèles de dynamique d'opinion ou d'attitude pour lesquelles les individus d'une population s'influencent mutuellement en échangeant leur opinion. Il est possible d'assimiler ces modèles à des modèles de transmission de l'information, cette dernière étant l'opinion ou l'attitude. Cependant, les modèles de filtre d'information mentionnés sont fondés sur la distance entre les attitudes, et non la congruence comme dans le modèle proposé ici.

Nous décrivons tout d'abord le modèle en détails. Puis nous l'étudions en connectivité complète, et ensuite en utilisant un graphe de voisinages aléatoires mais de taille constante. Nous proposons une modèle probabiliste approchant les résultats des simulations. Enfin, nous discutons nos résultats et traçons quelques perspectives futures.

## 2. LE MODELE

### 2.1. Hypothèses de modélisation retenues

Pour notre part, nous proposons un modèle qui, à partir d'hypothèses très simples sur le mécanisme expliquant que « *les détails disparaissent* », s'efforce de reproduire la diffusion des informations sur un objet au sein d'une population. Nous supposons que chaque individu dispose d'une opinion globale qui forme une prédisposition favorable ou défavorable au regard de l'objet ou problème sur lequel vont circuler des informations.

Globalement, l'objet ou problème peut être décrit par un ensemble de caractéristiques que nous appellerons « facettes ». L'objet est publicisé par une institution qui diffuse à un ensemble de récepteurs des éléments d'information leur permettant de déterminer l'utilité des facettes. Ces récepteurs sont des interlocuteurs privilégiés de l'institution, lecteurs d'une revue ou membres d'un réseau spécifique et sont contactés par elle selon une fréquence plus ou moins élevée. Ces récepteurs filtrent les informations ainsi délivrées selon leur attitude. Ils vont ensuite discuter avec leurs voisins qui, à leur tour, discuteront avec leurs autres voisins...

C'est durant ces discussions que les informations pourront être évoquées par un émetteur si elles sont suffisamment intéressantes pour lui et en accord avec son opinion globale. Il s'agit de prendre en compte les notions « *d'importance* » de l'information, d'« *accentuation* », la fréquence d'évocation des éléments d'information les plus intéressants est plus élevée, et de filtre cognitif lié aux « *préjugés et au croyances* » du locuteur. Un individu ne discutera donc que des informations relatives aux facettes dont l'utilité

est cohérente avec son opinion. Ainsi, s'il est a priori contre l'objet, il n'informerait son interlocuteur que des éléments négatifs dont il dispose, dans la mesure où il estime ceux-ci suffisamment intéressants, et tairait les éventuels éléments positifs en sa possession.

De même, lors de la discussion, l'information ne pourra être reçue par un allocuté qu'après avoir été confrontée « *à ses intérêts propres* ». Chaque individu est doté d'un filtre cognitif qui l'amène à ne considérer que les informations qui sont cohérentes avec son opinion et qui lui paraissent importantes. Ainsi, par exemple, lorsque l'opinion globale d'un individu est positive, il aura tendance à ne retenir que les informations positives sur l'objet et à ignorer les informations négatives, sauf si elles sont très fortement négatives. Globalement, il ne s'intéressera qu'aux informations saillantes, dont l'utilité est forte quel qu'en soit le signe.

Notre objectif est d'explorer la dynamique de ce modèle simple, notamment lorsque la structure d'interactions varie. Il s'agit d'identifier les phénomènes collectifs qui peuvent être observés.

### 2.2. Les variables du modèle

Nous considérons une population comprenant  $N$  individus. Les individus discutent d'un problème sur lequel ils échangent éventuellement de l'information sur ses facettes. Le problème est spécifié par l'ensemble  $F = (1, 2, \dots, d)$ , qui représente l'ensemble de ses facettes.

Chaque individu d'indice  $i$  est décrit par les paramètres suivants :

- $G_i \in \{-1; 1\}$ : opinion globale sur le problème. Elle est positive ou négative. Elle correspond aux *a priori* de l'individu. Dans le présent modèle, elle est supposée fixe, ce qui est évidemment peu réaliste. Nous comptons la rendre dynamique dans des versions ultérieures du modèle.
- $L_i$ : un sous-ensemble de  $F$  représentant la liste des facettes reçues par l'individu  $i$ . Lors d'une rencontre avec un voisin, c'est au sein de cette liste que l'individu choisit la facette dont il souhaite parler.
- $U_i = (u_1^i, \dots, u_j^i, \dots, u_d^i)$  avec  $u_j^i \in \mathcal{R}$ , sont les utilités, pour l'individu  $i$ , de chacune des facettes du problème. Nous supposons que l'utilité d'une facette apparaît immédiatement à l'individu qui en entend parler. Cette utilité lui permet de décider de retenir ou non la facette dans sa liste  $L_i$ .
- $\theta_i \in \mathcal{R}^+$ : seuil d'intérêt d'une facette dont l'utilité est cohérente avec l'opinion globale de l'individu. Ce seuil est augmenté de la valeur absolue de  $G_i$  lorsque l'utilité de la facette est contraire à l'opinion globale de l'individu.
- un voisinage qui est un sous-ensemble d'individus avec lequel l'individu entretient des relations

régulières. En fonction des valeurs des paramètres du modèle, ce voisinage peut être la totalité de la population ou un sous-ensemble de taille fixée.

Un individu peut faire partie des *abonnés* à l'institution et recevoir des informations à une fréquence paramétrable.

Nous supposons qu'à chaque itération du modèle, chacun des individus discute dans un ordre aléatoire avec les individus de son voisinage. Une discussion porte sur une seule facette du problème. Chaque facette retenue par un individu a une probabilité d'être transmise à d'autres, et retenue ensuite par un autre individu qui en entend parler et n'en dispose pas encore.

Voyons à présent plus en détail chacun des éléments de la dynamique du modèle en commençant par la mise à disposition des informations par l'institution « média ».

### 2.3. Une institution diffuse les informations

Une institution et son réseau de diffusion permettent de lancer la dynamique du modèle en délivrant, à l'initialisation, les informations sur les différentes facettes du problème. De nouvelles diffusions peuvent intervenir par la suite selon une fréquence qui est un paramètre du modèle.

Les individus membres du réseau de diffusion sont tirés aléatoirement dans la population à l'initialisation du modèle. Leur nombre est un paramètre du modèle. Un autre paramètre du modèle définit si ce réseau est composé d'individus ayant une attitude quelconque ou d'individus dotés des seules opinions globales positives ou négatives. Les interlocuteurs directs de l'institution sont les mêmes sur toute la durée de la simulation.

Décrivons à présent les dynamiques d'échange d'informations sur le problème.

### 2.4. Des discussions durant lesquelles les individus peuvent échanger les informations dont ils disposent

Lors d'une discussion et alors qu'une information sur une facette  $j$  est évoquée en sa présence, un individu a trois états possibles : il ignore l'information ; il se contente de la retenir ; il la retient et la diffusera lors des prochaines discussions.

La fonction déterminant l'état de l'individu est la suivante :

- état "retient l'information sur  $j$  (sans diffusion ultérieure)" si l'information est incongruente à l'attitude :  $G_i u_i^j < 0$  et  $|u_i^j| \geq \theta_i + |G_i|$

- état "ignore l'information sur  $j$ " lorsqu'elle l'information est jugée sans intérêt : si  $G_i u_i^j \geq 0$  et  $|u_i^j| < \theta_i$  ou si  $G_i u_i^j < 0$  et  $|u_i^j| < \theta_i + |G_i|$
- état "retient l'information sur  $j$  et la diffusera" si l'information est congruente à l'attitude :  $G_i u_i^j \geq 0$  et  $|u_i^j| \geq \theta_i$

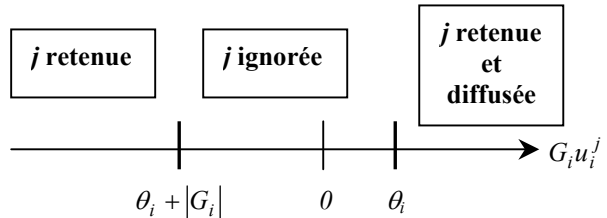


Figure 1. Etat de l'individu  $i$  face à une information  $j$  sur l'axe  $G_i u_i^j$

La figure 1 représente le filtre déterminant l'état au regard de l'information sur une facette  $j$ . On y observe que :

- lorsque l'utilité de  $j$  est cohérente avec l'opinion globale de l'individu (de même signe) et jugée suffisamment intéressante (valeur supérieure au seuil de saillance  $\theta_i$ ), la facette est retenue et pourra être diffusée par l'individu détenteur ;
- lorsque l'utilité de  $j$  n'est pas cohérente avec l'opinion globale de l'individu mais qu'elle est jugée suffisamment importante pour affecter son point de vue (valeur supérieure à  $\theta_i +$  valeur d'opinion globale de l'individu), la facette est retenue mais ne sera pas diffusée par l'individu ;
- dans les autres cas, la facette  $j$  est ignorée.

Nous étudions maintenant ce modèle, en supposant d'abord que chaque individu discute avec tous les autres, ensuite que les individus discutent avec un sous-ensemble restreint de la population.

## 3. ETUDE DU MODELE EN CONNECTIVITE COMPLETE

Pour notre étude, nous considérerons le cas simple où les utilités des facettes ne varient pas d'un individu à l'autre. Nous examinons tout d'abord le cas «connectivité complète» pour lequel chaque individu est connecté à tous les autres. Dans ce cas, tous les individus de la population sont voisins de chaque individu. L'institution est initialement seule détentrice de l'information qu'elle diffuse à ses abonnés qui pourront la diffuser à leurs tours à leur voisins. Dans ce qui suit, le terme « information » équivaut à « l'information sur la facette » et est ainsi employé pour alléger le discours.

Nous pouvons dans ce cas recenser les trois états finaux atteignables du modèle : nul n'a l'information ; seule une partie des individus « abonnés » à l'institution dispose de

l'information; une partie de la population « non abonnée » et une partie des « abonnés » à l'institution disposent de l'information.

### 3.1. Nul n'a l'information

Cet état intervient lorsqu'aucun des abonnés à l'institution ne trouve l'information intéressante ( $|u_i^j| \leq \theta_i$  si  $G_i u_i^j \geq 0$  ou  $|u_i^j| \leq \theta_i + |G_i|$  si  $G_i u_i^j < 0$  pour les abonnés). Dans ce cas, nul ne dispose de l'information, même si celle-ci apparaît cohérente et des plus utiles pour la population n'ayant pas de contact direct avec à l'institution.

Cet état souligne le caractère déterminant, pour la diffusion, des individus en position d'obtenir l'information. L'institution doit, pour diffuser une ou plusieurs informations, choisir de façon pertinente ses « contacts » au sein de la population.

### 3.2. Seule une partie des individus « abonnés » à l'institution dispose de l'information

Ce second état peut être atteint dans deux situations très différentes.

1. Une partie des individus en contact avec l'institution retient cette information qui lui semble importante, l'autre partie l'ignore. Cependant cette information a une utilité de signe inverse à leur opinion globale. Bien qu'en disposant, ils n'en parlent donc pas. Ainsi, l'information n'est pas diffusée, même si elle paraît utile et importante au reste de la population ( $G_i u_i^j < 0$  et  $|u_i^j| > \theta_i + |G_i|$  pour les abonnés recevant l'information).
2. Une partie des individus en contact avec l'institution retient l'information et en parle. Le reste de la population juge toutefois cette information sans intérêt et l'ignore. Cette situation peut intervenir si les opinions globales du reste de la population vérifient  $|u_i^j| \leq \theta_i + |G_i|$  ou  $|u_i^j| \leq \theta_i$  pour  $G_i u_i^j \geq 0$  pour le reste de la population.

Cet état souligne, comme l'état précédent, l'importance du choix, par l'institution, des individus destinataires de l'information. Si dans la première situation, comme dans les deux situations liées au premier état, les individus « abonnés » sont bloquants pour la diffusion sans toutefois apparaître comme des marginaux, ce n'est pas le cas de la seconde situation où la cible médiatique de l'institution juge important ce que les autres estiment négligeable.

### 3.3. Une partie de la population non « abonnée » et une partie des « abonnés » à l'institution disposent de l'information

La situation est ici telle qu'une partie au moins des abonnés de l'institution peut recevoir et transmettre l'information. Tout ou partie de la population peut alors in fine retenir l'information. Seuls les individus ne trouvant pas l'information intéressante la filtreront.

Globalement, nous retrouvons le caractère déterminant de l'intérêt ou non que porte l'individu à l'information.

Voyons à présent en quoi le fait qu'un individu n'échange pas avec tous les autres affecte les états finaux.

## 4. ETUDE DU MODELE A CONNECTIVITE PARTIELLE ET CONSTANTE

Le cas de la connectivité partielle induit une contrainte supplémentaire pour la diffusion de l'information. En plus des conditions sur l'information elle-même, il faut que deux individus soient voisins.

Pour étudier l'effet d'une connectivité partielle faible des individus, nous nous intéressons au cas où les abonnés ne sont pas limitants pour la diffusion et où la population est intéressée. Tous les individus ont même  $\theta_i$ . Nous considérons un problème doté d'une seule facette dont l'utilité est constante pour tous les individus et supérieure à  $\theta_i$ . Cette facette apparaît donc intéressante pour tous. Nous avons une population dont chaque individu a un même nombre de voisins. Un seul individu est contact de l'institution. Il reçoit et retransmet l'information unique. La population a pour moitié une opinion globale de même signe que l'utilité de l'information, pour moitié, une opinion globale de signe contraire.

Les simulations du modèle individus-centré montrent l'importance de la variation du nombre de voisins d'un individu. Pour mieux étudier cet effet « taille du voisinage », nous proposons un calcul approché de la distribution de probabilité du nombre final d'individus qui ont retenu l'information.

### 4.1. Les simulations du modèle individus-centré montrent l'importance de l'impact de la taille du voisinage d'un individu

Nous avons implémenté un modèle individus-centré dont chaque individu a un même nombre de voisins  $v$ , paramètre du modèle. Un plan d'expérience visant à appréhender l'effet du réseau de voisinage a été réalisé. Nous faisons varier le nombre d'individus de la population (50, 100, 200, 500, 5000) et le nombre de voisins d'un individu (de 1 à 10). Une simulation est à convergence lorsqu'aucun individu nouveau n'a reçu l'information durant une itération.

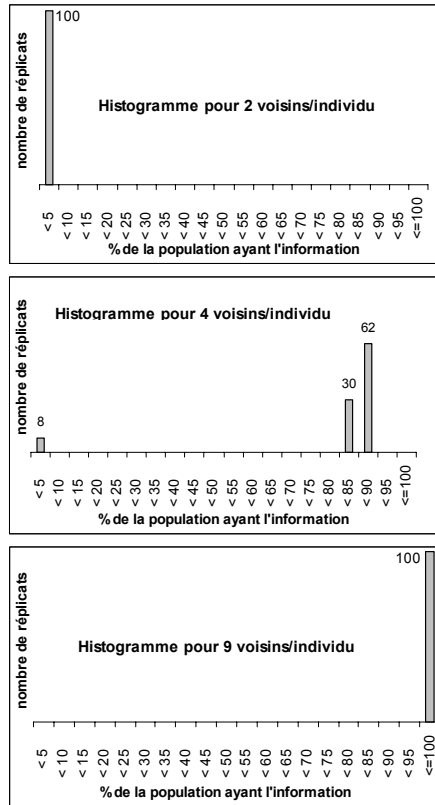


Figure 2. – Histogramme des résultats finaux de pourcentage de population ayant l'information pour 100 répliqués du modèle individus-centré et pour une population de 5000 individus dont 50 % ont une opinion globale positive dans les cas 2, 4 et 9 voisins/individu

La figure 2 illustre, pour une population de 5000 individus, les résultats de l'expérimentation en présentant l'importance, pour la détermination de l'état final, de la taille du voisinage d'un individu. En effet, si pour deux voisins par individu et sur les 100 répliqués, la diffusion ne dépasse pas 10 individus, pour 9 voisins, elle varie de 4978 à 4998 individus. Lorsque le voisinage est de taille 4, la diffusion peut s'éteindre rapidement (5 à 14 individus touchés) ou bien s'étendre et atteindre la majorité des individus ayant une opinion globale positive (4120 à 4364 individus). Pour les mêmes valeurs de paramètre en réseau complet pour lequel les voisins d'un individu sont tous les autres individus de la population, l'information est délivrée à toute la population.

Ce résultat s'explique par le fait que l'information ne se diffuse que lorsque les individus liés par une relation de voisinage retiennent l'information et la diffusent.

Pour mieux étudier ce phénomène et disposer de plus d'informations sur la distribution résultat, nous avons construit un algorithme approximant la distribution de probabilité du nombre d'individus finaux ayant retenu l'information.

## 4.2. Calcul approché de la distribution de probabilités du nombre d'individus finaux ayant retenu l'information

### 4.2.1 Le principe

Nous considérons un problème avec une seule facette dont l'utilité est la même pour tous les individus, supposée positive et supérieure à  $\theta$ . Nous avons une population dont chaque individu a même nombre de voisins.

$$U_i = (u_i) \text{ et } u_i > \theta, \quad \forall i$$

Un individu, contact de l'institution, reçoit et retransmet l'information unique.  $N_+$  individus ont une opinion globale positive (de même signe que l'utilité de la facette), les autres, une opinion globale de signe contraire.

Nous définissons l'ensemble des individus atteignables par l'information comme  $P_+$ , ensemble des individus pour lesquels  $G_i = 1$  alors que  $u_i < 1$ . Nous avons  $\text{card}(P_+) = N_+$  et  $(N - N_+)$  individus pour lesquels  $G_i = -1$ .

Nous cherchons à prévoir la diffusion de l'information sur la facette, c'est à dire la part des  $N_+$  individus qui retient la facette.

Le principe consiste à recenser toutes les diffusions possibles et à en calculer les probabilités d'occurrences. Nous considérons donc l'ensemble des déploiements possibles de la diffusion dans le temps. Pour cela, nous définissons un état de diffusion à un instant donné  $t$  par :

- $T^t$ , le nombre total d'individus atteint à un instant  $t$ ,
- $n^t$ , le nombre d'individus parmi les  $T^t$  qui viennent de retenir l'information,
- $p^t$ , la probabilité d'occurrence de cet état de diffusion.

La donnée de ces caractéristiques à l'instant  $t$  nous permet de calculer tous les états de diffusion à l'instant  $t+1$ , et leur probabilité associée. Pour cela, nous calculons  $m$  le nombre maximum de nouveaux individus pouvant retenir l'information à l'instant  $t+1$ . Ce nombre est inférieur ou égal à la somme des nombres de voisins des individus ayant reçu l'information à l'instant  $t$ , sans compter le voisin qui avait transmis l'information à l'instant  $t-1$ . Mais  $m$  est également inférieur au nombre d'individus non encore touchés par la diffusion dont l'opinion globale est positive. Le nombre  $m$  est donc :

$$m = \min(n^t \cdot (v - 1), N_+ - T^t)$$

Nous évaluons la probabilité que  $i$  voisins sur  $m$  aient une opinion globale positive en utilisant la loi binômiale :

$$p^i(T^t, n^t) = \binom{m^t}{i} p_+^{t,i} (1 - p_+^t)^{m^t - i}$$

Où  $p_+^t$  est la proportion au temps  $t$  d'individus de  $P_+$  n'ayant pas encore été atteints par la diffusion.

$p_+^t$  est recalculée à chaque itération et pour chaque quantité totale d'individus ayant déjà reçu l'information de façon à garantir le fait que les individus considérés sont bien des « nouveaux » individus, i.e. non encore pris en compte par l'algorithme. La correction de la valeur de  $p_+^t$  est réalisée de façon simplifiée et entraîne une surestimation de la probabilité d'atteindre un nouvel individu. On formule l'hypothèse que cette surestimation est négligeable.

#### 4.2.2 L'algorithme

La figure 3 nous présente, sur deux itérations, le principe de l'algorithme sur un exemple portant sur un réseau de taille « deux voisins par individu ».

L'algorithme se caractérise donc à l'instant  $t$  par une liste de triplets décrivant les  $z$  états de diffusion possibles à cet instant :

$$L_t = \left( (T^t, n^t, p^t)_1, \dots, (T^t, n^t, p^t)_d, \dots, (T^t, n^t, p^t)_z \right)$$

La liste  $L_{t+1}$  est calculée en deux étapes :

1. On étend chaque état de diffusion  $(T^t, n^t, p^t)_d$  de la liste  $L_t$  selon le calcul décrit au paragraphe précédent :

$$\begin{aligned} (T^t, n^t, p^t)_d &\rightarrow \\ \left( (T^{t+1}, n^{t+1}, p^{t+1})_1 = (T^t + 1, 1, p^t p^1(T^t, n^t))_d, \dots, \right. \\ \left. (T^{t+1}, n^{t+1}, p^{t+1})_i = (T^t + i, i, p^t p^i(T^t, n^t))_d, \dots, \right. \\ \left. (T^{t+1}, n^{t+1}, p^{t+1})_m = (T^t + m, m, p^t p^m(T^t, n^t))_d \right) \end{aligned}$$

2. On regroupe dans  $L_{t+1}$  les nouvelles diffusions dont les valeurs  $T^t$  et  $n^t$  sont similaires en sommant leur probabilité d'occurrence.

Comme on le voit sur la figure 3, le développement d'une branche est arrêté lorsqu'on tire 0 nouveaux voisins positifs. L'algorithme se termine lorsque toutes les branches ont été développées.

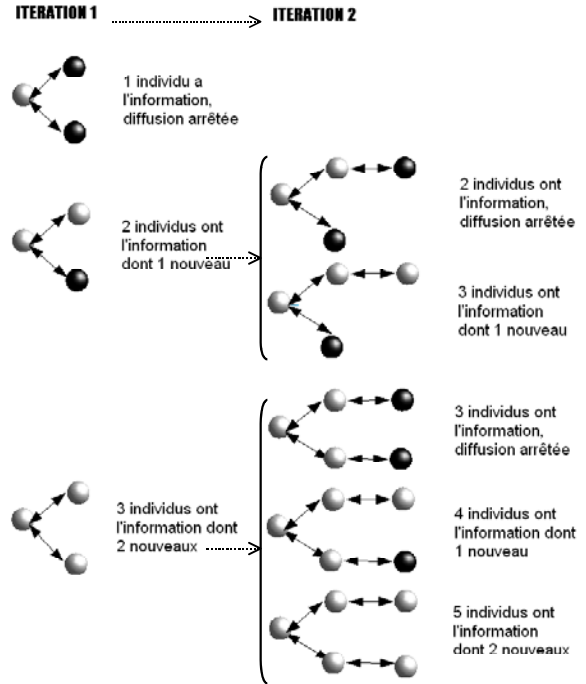


Figure 3. Principe de l'algorithme d'approximation de la diffusion d'information (exemple pour un réseau de voisinage de taille 2). Les points blancs sont les individus ayant une opinion globale de même signe que la valeur de l'information ; la flèche en tireté indique que la diffusion se poursuit à partir des nouveaux informés

L'algorithme est coûteux en occupation mémoire car la taille des tableaux stockant les diffusions possibles croît en fonction du plus fort accroissement du nombre d'individus informés à l'instant  $t-1$  à la puissance du nombre d'itération. Ainsi donc, plus l'information se diffuse, plus la taille des tableaux augmente. Des « astuces », permettent de minimiser l'occupation mémoire des données. Par exemple :

- les tableaux de stockages sont basés sur des primitives ;
- des branches de l'arbre sont rendues caduques : d'une part parce que jugées insignifiantes, d'autre part parce que faisant double emploi. On considère, qu'un événement ne doit être pris en compte que si la valeur de sa probabilité d'occurrence est supérieure à  $10^{-9}$ . Ce seuil semble pertinent pour des populations de petites tailles bien que l'on puisse penser que tel ne serait pas le cas pour de grandes populations.

L'algorithme peut en fine traiter sans problème une population de 500 individus. Pour information, pour certaines simulations, la taille cumulée des tableaux est de l'ordre de  $5 \cdot 10^7$ .

#### 4.2.3. Comparaison entre le calcul et la simulation du modèle individus-centrés

Pour évaluer la qualité de l'algorithme d'approximation, nous confrontons ses résultats aux résultats du modèle individus-centré. Nous considérons une population de 500 individus dont un unique abonné à l'institution. Nous faisons varier la taille du voisinage (de 2 à 10 voisins). Nous mesurons, à convergence des modèles, le nombre d'individus dont l'opinion globale est de même signe que l'information diffusée et qui sont détenteurs de l'information.

Pour le modèle individus-centré, 3000 répliqués ont été réalisés pour chaque jeu de valeurs de paramètres de façon à obtenir une approximation de la distribution des résultats.

Les figures 4 présentent partiellement cette comparaison pour 2, 3, 4, 5 et 9 voisins pour une population de 500 individus. La distribution est globalement bien reproduite. Nous observons que les deux modèles sont relativement proches bien que l'algorithme d'approximation sous-estime légèrement le résultat pour 4 voisins/individu et surestime légèrement le résultat pour les voisinages de taille 5 et 6 (non présenté ici).

La comparaison a été également pratiquée pour une population de 50 et de 100 individus (toujours dotées d'un unique abonné à l'institution). L'approximation garde la même qualité. L'algorithme ne peut pour l'instant pas s'exécuter pour une population de taille supérieure à 500 individus.

Nous avons pratiqué une série d'expériences pour apprécier la qualité de l'approximation lorsque la part d'individus ayant une opinion globale de même signe que l'utilité de l'information varie. Nous avons comparé les deux modèles pour une population de 500 individus, pour un nombre de voisins par individu variant de 2 à 7 par pas de un, et pour des taux d'individus ayant même signe d'opinion globale que l'utilité de l'information diffusée variant de 5 % à 95 % par pas de 5%. 3000 répliqués sont réalisées pour chaque expérience sur le modèle individus-centré.

La figure 5 présente les résultats des expériences menées dans le cas 3 voisins par individu. On y observe que les deux modèles renvoient globalement des résultats similaires. Tous deux indiquent le fait que pour un taux d'individus ayant une opinion globale positive variant de 5 à environ 50 %, le nombre d'individus touchés par la diffusion en fin de simulation varie entre 1 individu et 25 % de la population. Pour des tailles de  $P_+$  allant d'environ 50 % à 75 % de la population totale, la diffusion peut être faible (de 0 à 20 % de  $P_+$  ou forte (40-45 % à 100 % de  $P_+$  dans les cas les plus favorables). Pour des tailles de  $P_+$  supérieures à 75-80 % de la population totale, tous les individus ont l'information en fin de simulation.

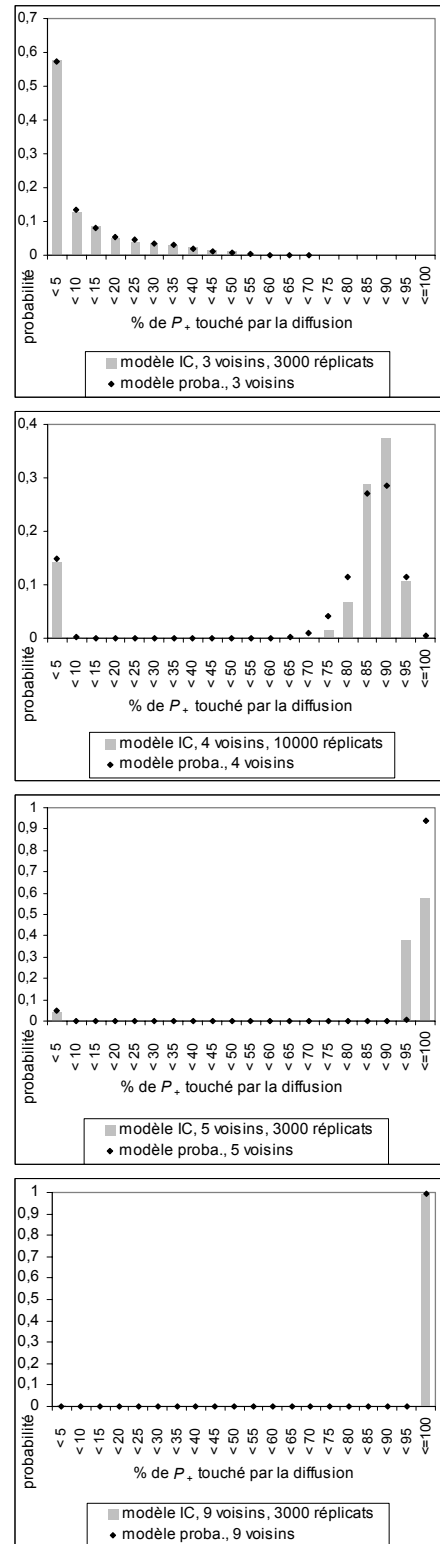


Figure 4. Comparaison de l'algorithme d'approximation au modèle individu-centré pour 2, 3, 4, 5, 9 voisins/individus pour une population de 500 individus

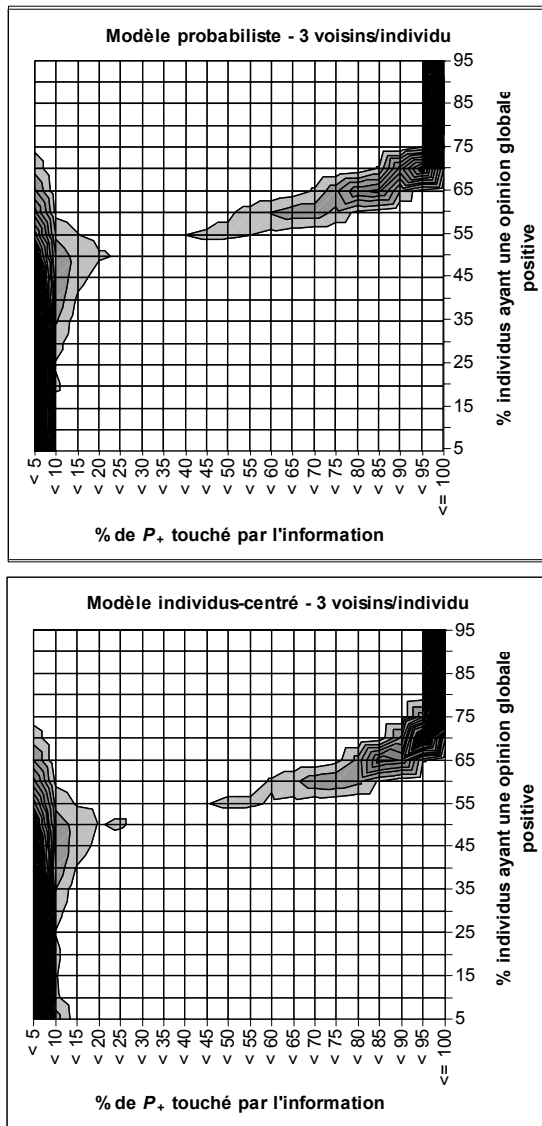


Figure 5. Comparaison des distributions des % de  $P_+$  ayant retenu l'information pour les deux modèles pour différentes tailles de  $P_+$ , exprimées en pourcentage de la population totale. 500 individus ayant 3 voisins par individu (3000 répliques du modèle individus-centré)

Des différences entre les deux modèles sont observables. Elles sont identiques pour toutes les tailles de voisinages testées (de 2 à 7 voisins par individu). Pour des tailles de  $P_+$  induisant une faible diffusion de l'information, le calcul approché tend à légèrement sous-estimer la diffusion sur le modèle IC (observable en bas à gauche de la figure n°5). En revanche, pour des tailles de  $P_+$  induisant une diffusion forte mais pas totale, c'est l'inverse, le modèle approché tend à sur-estimer légèrement les observations (observable en haut à droite de la figure n°5).

## 5. BILAN - PERSPECTIVES

Nous avons proposé un modèle de filtrage de l'information en fonction d'une opinion globale a priori

et d'un niveau d'intérêt accordé à l'information, et nous en avons étudié certaines versions simples.

L'étude du modèle pour une population d'individus échangeant avec tous, a rappelé, de façon extrêmement simple, que pour que les individus d'une population disposent d'une information, il ne suffit pas qu'ils se rencontrent, il faut encore que l'information passe le filtre. Par ailleurs, les individus relais entre l'institution publicisant l'objet et le reste de la population, peuvent, s'ils sont mal choisis par l'institution, bloquer la diffusion des informations.

L'étude du modèle lorsque les individus ne parlent qu'avec leurs voisins (leur nombre étant supposé identique pour tous), met en lumière une contrainte supplémentaire pour l'obtention de l'information. Elle peut s'exprimer par l'existence ou non d'un « chemin » permettant à l'information de circuler jusqu'à un individu susceptible de la recevoir. Ce chemin est en fait un ensemble d'individus liés au moins, deux à deux, et étant tous à même de retenir et transmettre l'information (i.e. intéressés par l'information et d'opinions globales similaires).

Nous avons observé que pour des populations composés d'individus ayant des opinions globales différentes, les « détails disparaissent », notamment lorsque la population diffère en opinion des individus abonnés pour recevoir initialement l'information. Il n'est par ailleurs pas nécessaire qu'il existe une structure d'interactions limitant la diffusion pour observer des diffusions nulles ou partielles.

Nous proposons par ailleurs un calcul approché de la distribution de probabilité du nombre d'individus atteints par la diffusion, dans le cas de réseaux à nombre de voisins constants, dans des conditions simples (une seule source, une seule facette d'utilité identique pour tous). Il approxime avec une qualité intéressante les résultats du modèle individus-centré.

Ce travail est, à bien des égards, préliminaire. Il serait nécessaire de le confronter à ceux réalisés dans le cadre de la théorie de la percolation, et notamment à l'étude du modèle d'épidémiologie SIR (Susceptible, Infective, Removed) (Newmann, 2002). Il faut par ailleurs faire évoluer l'algorithme de façon à pouvoir étudier l'impact d'une structure de voisinage quelconque. On peut également envisager de prendre en compte des mesures plus sophistiquées de la structure des réseaux, telle que la mesure du « small world ».

Notre étude du modèle individus-centrés est également assez restreinte. Il serait intéressant d'étudier l'impact du nombre de caractéristiques descriptives du problème sur la vitesse de diffusion. De même, différentes distributions des utilités dans la population pourraient engendrer des effets globaux qui restent à mettre en évidence.

Mais la perspective la plus intéressante est encore celle qui consistera à étudier l'impact d'une opinion globale dynamique, en considérant éventuellement des types différents de dynamique. En effet, si l'attitude, que nous appelons ici opinion globale, conditionne le traitement de l'information, de nombreux chercheurs en psychologie sociale supposent également qu'elle se forme et se modifie au gré des informations recueillies (Eagly A.H., Chaiken S., 1993). Le caractère stable de l'opinion globale peut donc utilement remis en cause.

Wu F., Huberman Bernardo A., Adamic Lada A., Tyler Joshua R., 2004 – Information flow in social groups. *Physica A* 337, 327-335.

## REFERENCES

- Allport G.W., Postman L., 1947. *The psychology of Rumor*, New York: Henry Holt.
- Anzieu D., Martin J.Y., 2000. *Les communications dans les groupes restreints*. In : La dynamique des groupes restreints, PUF, Le Psychologue, p. 188-210.
- Crawford V.P., Sobel J., 1982. Strategic Information Transmission. *Econometrica* 50:1431-51.
- Deffuant G., Weisbuch G., Amblard F., Faure T., 2003. "Simple is beautiful... and necessary." *Journal of Artificial Societies and Social Simulation* 6(1).
- Deffuant, G., Huet S., Amblard, F., 2005. "An individual-based model of innovation diffusion mixing social influence and rational anticipation". *American Journal of Sociology* , 110-4, January 2005, pp.1041-1069.
- Eagly Alice H., Chaiken S., 1993. *The psychology of attitudes*. Thomson/Wadsworth; 1st edition (January, 1993), 800 pages
- Galam S., 2003. Modeling Rumors: The No Plane Pentagon French Hoax Case. *Physica A: Statistical Mechanics and its Applications*, Volume 320, 15 March 2003, Pages 571-580
- Lawson Gabriel C., Butts Carter T., 2004. Information Transmission Through Human Informants: Simulation. *North American Association for Computational Social and Organizational Science Meeting*, [pascalfroissart.online.fr/3-cache/2004-lawson-butts.pdf](http://pascalfroissart.online.fr/3-cache/2004-lawson-butts.pdf)
- Newmann M.E.J., 2002. The spread of epidemic disease on networks. *Phys. Rev.* E66, 016128.
- Snijders Tom A.B., 1998. Methodological Issues in Studying Effects of Networks in Organizations. *Computational & Mathematical Organization Theory*, 4, 205-215.
- Tsimring Lev. S., Huerta R., 2003. Modeling of contact tracing in social networks. *Physica A* 325, 33-39.
- Urbig D., 2003. Attitude Dynamics with Limited Verbalisation Capabilities. *Journal of Artificial Societies and Social Simulation*, vol. 6, n° 1, [jasss.soc.surrey.ac.uk/6/1/2.html](http://jasss.soc.surrey.ac.uk/6/1/2.html).
- Wittek R., Wielers R., 1998. Gossip in Organizations. *Computational & Mathematical Organization Theory* 4:2, 189-204.