Rapport de projet de synthèse :
Etude de la diffusion de contenu
sur les réseaux sociaux

Bensabath Léore, Hanine Ilyas.

Sommaire :

Nous remercions le professeur Koen de Turck pour son temps, son aide et ses précieux conseils sans lesquels nous n'aurions pu réaliser le projet.

1. Introduction

Les phénomènes de diffusion sont l'objet d'étude et d'observation par l'Homme depuis très longtemps. Parmi eux nous pouvons citer la propagation de maladies, les déplacements humains et même plus récemment au travers de modèles économiques. Depuis plusieurs années maintenant, la diffusion de l'information –l'action de propager de l'information vers un ou plusieurs publics- suscite un vif intérêt de la part des scientifiques.

En effet avec l'apparition des médias sociaux (réseaux sociaux en particulier) comme Facebook, Twitter etc… notre manière de produire et d'interagir avec l'information a été métamorphosée. Ces médias qui servent de support à la diffusion de l'information font l'objet d'études très poussées.

Le phénomène de diffusion de l'information rentre dans le cadre d'une jeune discipline : l'analyse prédictive. Il s'agit d'un domaine très difficile, où la dynamique humaine et sociale joue un rôle majeur. En conséquence plusieurs modèles parfaitement justifiables mathématiquement peuvent aboutir à des résultats totalement différents.

Ce domaine de l'étude de l'information en général connaît aujourd'hui un irrésistible essor, et trouve de nombreuses applications dans tous les domaines : de l'industrie au marketing en passant par la banque.

2. Approche par des graphes

Une première approche considérée est celle de modéliser le réseau par un graphe orienté dont les nœuds sont les utilisateurs et les arcs sont les liens entre ses utilisateurs, qui sont bien des arcs puisque les liens peuvent être dissymétriques : en effet un utilisateur peut avoir une plus grande influence sur son voisin que voisin n'en est sur lui, c'est notamment le cas de personnes influentes sur les réseaux sociaux.

2.1 Méthode cascade

Le principe de la méthode est le suivant :

- L'utilisateur peut être dans un état actif ou inactif.
- Si l'utilisateur est actif, alors il a une probabilité de diffuser l'information à ses voisins. Ces probabilités sont modélisées par les poids accordés aux arcs partant de l'utilisateur vers ses voisins.
- Dès qu'un utilisateur échoue à diffuser l'information vers un de ses voisins, il n'a plus la possibilité de le faire.

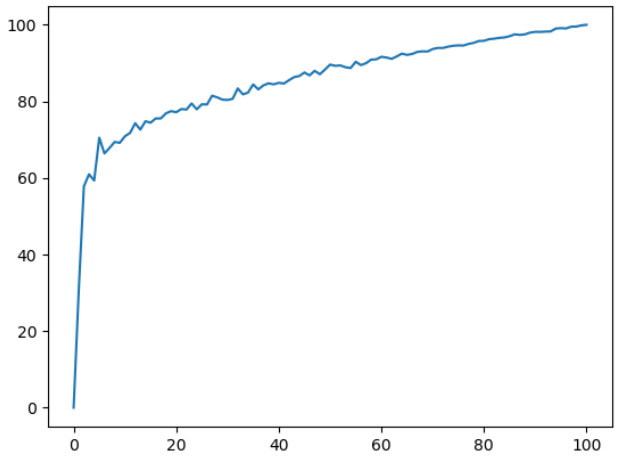
L'algorithme qui vient naturellement est celui d'un parcours en largeur. On se donne un graphe du réseau et une liste des infectés (ceux qui détiennent l'information) au temps 0. Dans notre tableau de sortie on y met initialement sous forme de tuple (le nœud, le temps d'infection ici 0).

Tant que la liste des infectés n'est pas vide on récupère la premier infecté de la liste qui correspond au premier infecté dans le temps. Pour chacun de ses voisins on réalise un tirage en fonction de la probabilité de diffusion –le poids des arcs-, pour savoir s'il y a diffusion ou non.

On ajoute dans notre tableau de sortie (le nœud, le temps d'infection du parent +1). Ici on discrétise le temps sur N.

En sortie on récupère un tableau de tuples contenant les nœuds infectés et leur temps d'infection. On peut réaliser alors une simulation visuelle pour se rendre compte de la manière dont se propage l'information. En moyenne sur 100 individus, le processus se termine au bout de 5 à 6 étapes de diffusion. On observe également que le nombre d'étapes n'est pas une fonction linéaire du nombre d'utilisateurs, ce qui correspond à la réalité ; pour 1000 individus on obtient une moyenne de 10 étapes de diffusion. C'est donc une source de satisfaction.

On réalise ensuite une moyenne sur 50 simulations du nombre final d'infectés en fonction du nombre initial d'infectés. On obtient la courbe suivante :



On observe un saut. Selon nos interprétations ce saut est en fait l'expression du principe de notre modèle selon lequel un individu n'a qu'une seule chance d'infecter ses voisins. C'est, entre autre, une des limites de ce modèle. Dans la réalité sur Facebook par exemple, l'influence d'un utilisateur qui partage du contenu opère toujours même si ses amis n'ont pas partagé l'information tout de suite après… Ceci nous amène à considérer une méthode différente.

2.2 Méthode seuil

Cette méthode comme la méthode cascade, est centré sur l'état de l'utilisateur et non l'état global du réseau.

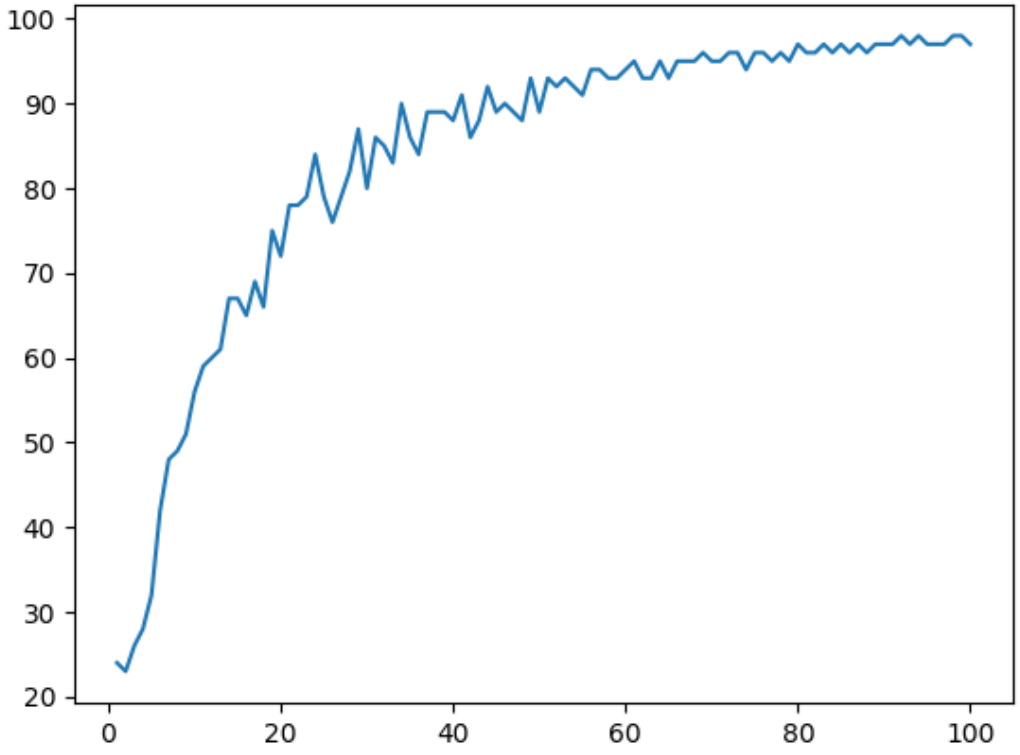
Le principe est le suivant :

- Chaque utilisateur i possède un seuil d'activation ω_i.
- En notant A_i l'ensemble des voisins actifs de l'utilisateur i, obtient l'équation d'activation suivante : $(1): i\ actif \Leftrightarrow \sum_{j \in A_i} p_{i,j} \geq \omega_i$ si on désigne les poids des arcs allant de j vers i par $p_{i,j}$.

La grosse différence qu'il existe avec la méthode cascade est qu'il ne s'agit pas de probabilités, mais de seuil. C'est donc un modèle déterministe : de mêmes conditions initiales vont mener à un même résultat final. L'avantage de cette méthode est que l'influence d'un utilisateur actif est prise en compte pour l'infection de ses voisins jusqu'à l'équilibre du système.

Algorithmiquement on se donne un graphe orienté et une liste d'infectés initiaux. On récupère en sortie un tableau de tuples (booléen, temps d'infection). Tant qu'il n'y a pas de nouveaux infectés, on parcourt l'ensemble des utilisateurs non infecté (dont le booléen est faux), et on vérifie (1). Si (1) est vérifiée, on change le tuple correspondant. On finit la boucle en incrémentant le temps.

De la même manière que pour la méthode cascade on réalise une simulation établissant la moyenne sur 50 simulations du nombre final d'infectés en fonction du nombre initial d'infectés. On obtient la courbe suivante :



Il n’est pas surprenant de noter l’absence de « saut » sur ce modèle par rapport au modèle cascade. La deuxième partie de la courbe est similaire.

Ces modèles présentent l’avantage d’être très naturels, facile interprétables. La donnée du graphe permet une grande liberté sur la modélisation du réseau, en accordant des poids plus ou moins importants, en définissant des clusters d’utilisateurs etc. Cependant comme nous avons pu le vérifier lors de nos simulations, les calculs sont lourds, d’autant plus que pour obtenir des résultats convenables il faudrait se placer sur des graphes de taille plus conséquente.

2.3 Modèle SIR

2.3.1 Modèle SIR déterministe

Le modèle SIR considère une population fixe coupées en 3 parties : les ”susceptibles” d'être infectés, QS(t); les infectés ou ”infected”, QI (t); et les enlevés ou ”removed”, QR(t). QS(t) représente le nombre de personnew susceptibles d'être infectées par la maladie. QI (t) est le nombre d'individus capables de transmettre la maladie aux individus de la catégorie ”susceptibles” précédente. QR(t) est le nombre d'individus qui ont été infectés puis guéris de la maladie (donc plus capables de la transmettre ni de la recevoir d'où ”removed”). On définit β et γ tels que les taux de nouvelles infections et le taux de nouvelles rémissions soient respectivement à l'instant t: $\beta QS(t)QI(t)/N$ et $\gamma QI(t)/N$. Ce qui donne les équations différentielles suivantes pour les fonctions QS, QI et QR :

$$\frac{dQS}{dt} = -\frac{\beta}{N} * QS * QI$$

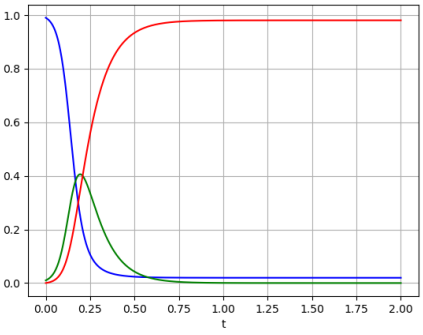
$$\frac{dQI}{dt} = \frac{\beta}{N} * QS * QI - \gamma * QI$$

$$\frac{dQR}{dt} = \gamma * QI$$

Où N est la taille de la population initiale, soit N=QS(0)+QI(0).

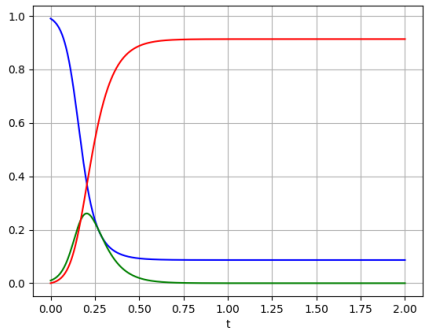
Ce modèle épidémiologique s'adapte assez intuitivement au cas de la diffusion de contenu sur un réseau social.

On considère ainsi les trois populations S, I et R comme trois groupes d'utilisateurs du réseau suite à la diffusion d'un document, d'un message ou de n'importe quel contenu. Le groupe S correspond aux utilisateurs qui n'ont pas encore accepté le document (et donc pas encore diffusé), l'état I correspond aux utilisateurs qui ont accepté le document mais qui ne l'ont pas encore diffusé, le groupe R correspond aux utilisateurs qui ont accepté et diffusé le document : ils n'interviennent donc plus dans la diffusion.



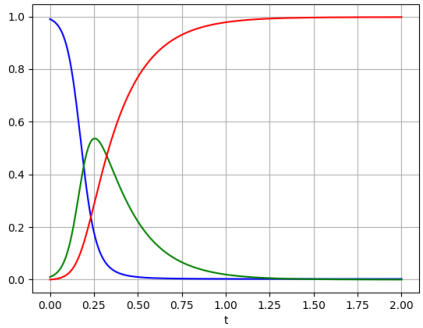
S(0)=99 ; I(0)=1 ; R(0)=0

N=100 ; beta=40 ; gamma=15



S(0)=99 ; I(0)=1 ; R(0)=0

N=100 ; beta=40 ; gamma=1



S(0)=99 ; I(0)=1 ; R(0)=0

N=100 ; beta=30 ; gamma=5

Bleu : Susceptibles ; Vert : infectés ; Rouge : Retirés

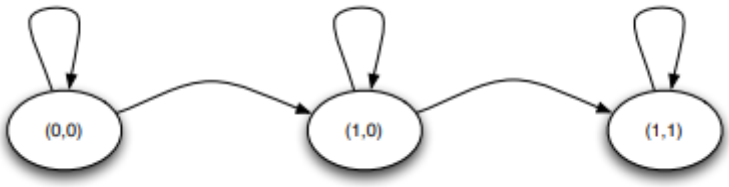
Les grandeurs S, I et R ont été normalisées par le nombre d'utilisateurs N

Cette légende est valable pour le reste de la partie.

2.3.2 Modèle SIR stochastique

Nous avons étudié un modèle SIR déterministe, où la population est considérée comme continue et uniforme. Nous allons maintenant étudier un modèle stochastique du schéma SIR, c'est-à-dire discrétiser la population en N individus distincts caractérisés aléatoirement, modéliser les liens entre eux et parler de propagation en termes de probabilités.

On renomme les trois états S, I et R respectivement (0, 0), (1, 0), (1, 1), pour plus de lisibilité : en effet, le premier paramètre du tuple correspond à l'acceptation du document (0 = document non accepté ; 1 = document accepté) et le second paramètre du tuple correspond à la diffusion du document (0 = document non transmis ; 1 = document transmis).



Un utilisateur a deux actions possible lorsqu'au moins un de ses voisins lui a diffusé le document : l'accepter (le lire, ne pas le jeter) et le diffuser auprès de ses propres voisins. Pour pouvoir diffuser le document, un utilisateur doit obligatoirement l'avoir accepté auparavant. A tout moment, un utilisateur se trouve dans chacun des trois états avec une certaine probabilité.

Soit $\rho(0,0)$ (i, t) la probabilité que l'utilisateur i se trouve dans l'état (0,0) au temps t. Il en est de même pour les états (1,0) et (1,1). On a alors la contrainte suivante : $\rho(0,0)(i, t) + \rho(1,0)(i, t) + \rho(1,1)(i, t) = 1$.

Il y a deux probabilités à définir qui sont la probabilité d'acceptation Fi(t) (pour passer de l'état (0,0) à l'état (1,0)) et la probabilité de diffusion Gi(t) (pour passer de l'état (1,0) à l'état (1,1)). La probabilité de rester dans l'état (0,0) se définit alors par $1 - Fi(t)$. Il en va de même pour celle de rester dans l'état (1,0). Chacune de ces probabilités dépend des paramètres de l'utilisateur i :

$Fi(t) = F(\nu(i,d), s(i,d), a(i(t)))$ (probabilité d'acceptation) où :

– $\nu(i,d)$ est le taux de propagation externe du document d pour un nœud/utilisateur i. Il représente toutes les informations qui sont diffusées hors du réseau et qui influencent quand même l'utilisateur. Si l'on se place dans un réseau social de type Twitter, la télévision fera partie des moyens de communication externes au réseau qui contribueront quand même à la diffusion de l'information au sein de celui-ci ;

– $s(i,d) = sim(\varphi i , d)$ est la similarité entre le profil (centres d'intérêt) de l'utilisateur i et le document d. Ce paramètre pénalise la probabilité d'acceptation si le document ne fait pas partie des centres d'intérêt de l'utilisateur et l'augmente dans le cas inverse ;

– a(i(t)) correspond au nombre de voisins (entrants) qui ont déjà diffusé le document d au temps t. L'intuition qui sous-tend notre modèle est que plus un utilisateur a de contacts qui lui diffusent une information, plus la probabilité qu'il l'accepte est élevée. On voit bien ici les deux clés de la diffusion dans ce modèle : si le contenu est très proche du profil d'un utilisateur, celui-ci va être prompt à le rediffuser. Si par contre le contenu ne correspond que peu, voire pas du tout, aux centres d'intérêt de l'utilisateur, il faudra qu'un grand nombre de ses contacts lui aient diffusé ce même contenu pour qu'il l'accepte.

De même, nous définissons :

$Gi(t) = G(Fi(t), wi)$ (Probabilité de diffusion) où wi est un paramètre de volonté (willingness) de i. En effet, certains utilisateurs ont une plus grande tendance à diffuser de l'information que d'autres. Par l'intermédiaire de la probabilité d'acceptation, la probabilité de diffusion dépend elle aussi du nombre de voisins diffuseurs et de la proximité de l'information avec le profil de l'utilisateur. Pour diffuser un contenu un utilisateur doit l'avoir accepté auparavant, et il diffusera ce contenu avec une probabilité d'autant plus grande si sa probabilité d'acceptation était élevée.

Nous avons pris pour nos fonctions de transition celles du modèle de Cédric Lagnier et Eric Gaussier [1] qui les utilisent car elles permettent d'activer l'utilisateur si les paramètres atteignent un certain niveau dit "de seuil".

$$F_i(t) = \frac{1}{1 + \exp\left(-\lambda_1(s(i,d) - 0.5) - \lambda_2\left(a\big(i(t)\big)\right) - \lambda_3\big(v(i,d)\big)\right)} \; si \; a\big(i(t)\big) \geq 1$$

$$F_i(t) = 0 \; sinon$$

$$G_i(t) = \frac{1}{1 + \exp(-\beta_1(F_i(t) - 0.5) - \beta_2 * w_i)}$$

Certains paramètres (la similarité pour F et F pour G) pénalisent la probabilité lorsqu'ils sont en dessous d'une certaine valeur, alors que les autres ne font que contribuer à l'augmentation de la probabilité.

On peut maintenant exprimer l'évolution des utilisateurs au cours du temps en fonction des probabilités de transition :

$$\rho(1,1)(i,t+1) = \rho(1,1)(i,t) + \rho(1,0)(i,t) * G_i(t)$$

$$\rho(1,0)(i,t+1) = \rho(1,0)(i,t) * \big(1 - G_i(t)\big) + \rho(0,0)(i,t) * F_i(t)$$

$$\rho(0,0)(i,t+1) = \rho(0,0)(i,t) * \big(1 - F_i(t)\big)$$

Pour chacun des 3 états, la probabilité que l'utilisateur s'y trouve à l'étape t + 1 est égale à la probabilité qu'il n'en soit pas parti à laquelle s'ajoute la probabilité qu'il y soit entré. On s'aperçoit avec ces équations que la convergence va se faire lorsque tous les utilisateurs reliés à l'initiateur de la diffusion seront dans l'état (1,1), c'est-à-dire qu'ils auront tous accepté puis diffusé le document.

Illustration :

Nous définissons la densité d'utilisateurs dans un état comme étant la moyenne des probabilités de présence dans l'état sur tous les utilisateurs.

$$\rho(0,0)(t) = \frac{1}{N} * \sum_{1 \leq i \leq N} \rho(0,0)(i,t)$$

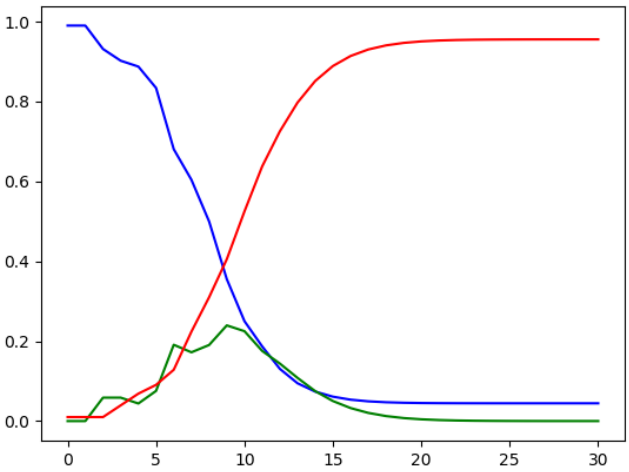
$$\rho(1,0)(t) = \frac{1}{N} * \sum_{1 \leq i \leq N} \rho(1,0)(i,t)$$

$$\rho(1,1)(t) = \frac{1}{N} * \sum_{1 \leq i \leq N} \rho(1,1)(i,t)$$

Pour une population indifférenciée : tous les paramètres sont à 0.

On effectue d'abord une première simulation où tous les paramètres sont à 0 pour se mettre dans le même cas que le modèle SIR déterministe : les utilisateurs sont indifférenciés. On considère de plus que chaque utilisateur peut envoyer le document à chaque autre, pour la

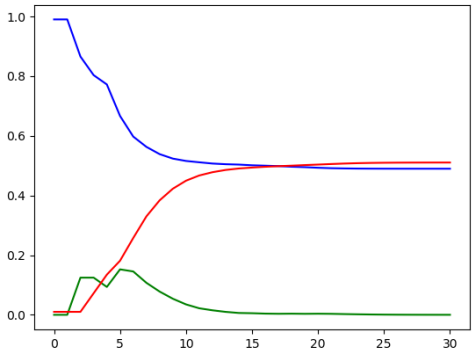
même raison.
On obtient le graphe suivant pour un réseau de 100 utilisateurs, avec un seul diffuseur initial :



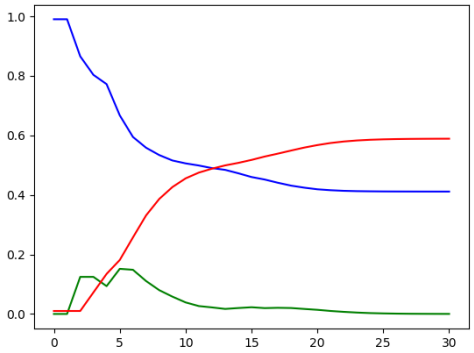
On retrouve l'évolution du modèle SIR déterministe, avec certes moins de régularité, dû au caractère probabiliste du modèle.

Considérons maintenant un graphe composé de deux groupes d'utilisateurs très reliés entre eux, disons des groupes d'amis. On considère également que quelques utilisateurs du premier groupe et quelques utilisateurs du second groupe sont en contact, les gens ayant naturellement des connaissances hors de leur « noyau dur » de relations. Nous allons étudier avec notre modèle comment un message né dans un des groupes se transmet au sein des utilisateurs du second groupe, en fonction du nombre de lien intergroupe.

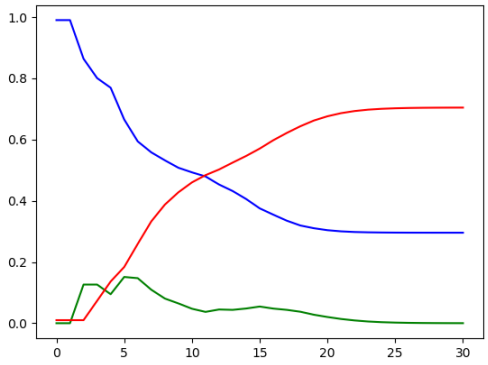
Nous avons générer des simulations pour deux groupes de 50 personnes ayant chacun en moyenne 25 liens au sein de leur groupe. Les courbes qui suivent sont les courbes moyennes pour 30 simulations sous les mêmes paramètres. Le nombre de liens intergroupe est indiqué en légendes des graphes.



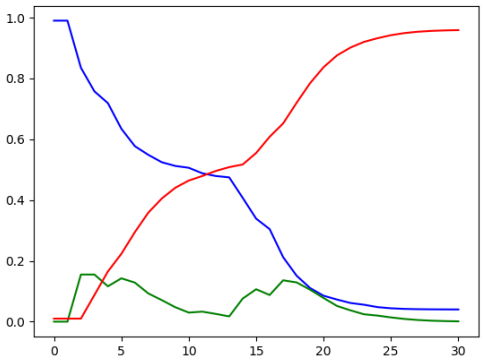
2



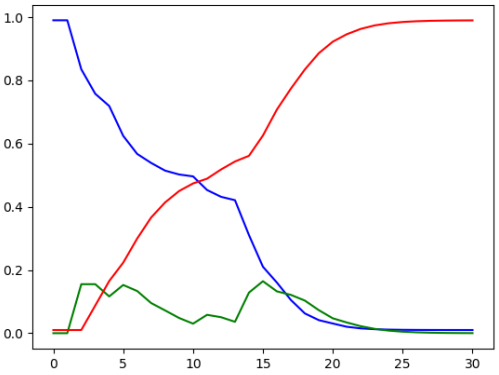
4



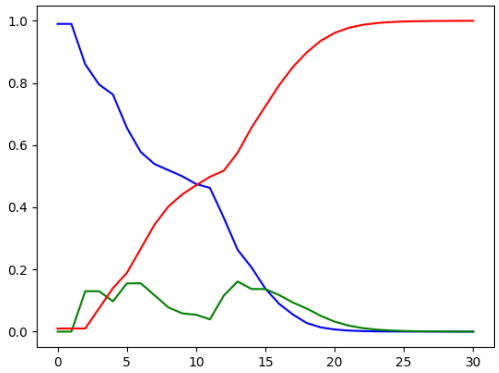
6



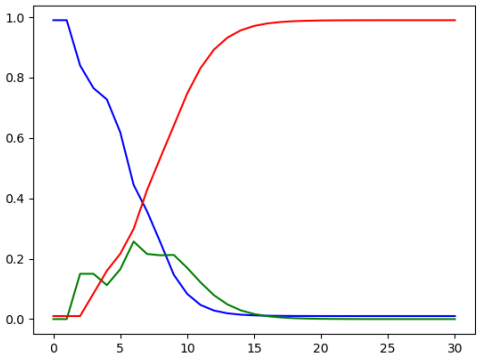
8



10



15



25

Quand les liens sont très peu nombreux, on constate que le message ne parvient pas à se répandre jusqu'au deuxième groupe, du moins pas entièrement. Ainsi, pour 2 liens intergroupes seulement

On observe bien que la moitié seulement des utilisateurs ont pris connaissance du message à l'équilibre. Ce nombre croit au fur et à mesure que les liens augmentent. A partir de huit liens, le message finit par atteindre tous les utilisateurs des deux groupes. A partir de là, on observe bien que le message se transmet en deux étapes, d'abord dans un groupe puis dans

l’autres, comme si deux modèles SIR étaient lancés l’un après l’autre. Pour 25 liaisons intergroupes, on retrouve comme attendue une situation à un seul graphe uniforme.

Enfin, la présence de différents paramètres permet d’affiner le modèle par rapport à un modèle SIR déterministe où seuls les paramètres β et γ permette de jouer sur la singularité des réseaux. Le modèle stochastique, par font entrer en jeux davantage de paramètres qui permette de spécifier la singularité des utilisateurs et du réseau global.

On étudie par exemple le temps de stabilisation selon différentes valeurs des différents paramètres. On présente des moyennes sur 50 graphes. La première ligne du tableau correspond aux paramètres de F et G donnés dans cet ordre : lambda1, lambda2, lambda3 ; beta1, beta2

0.5, 0.5, 0 ; 1, 0	0, 1, 0 ; 1, 0	1, 0, 0 ; 1, 0	0.5, 0.5, 0 ; 0, 1	0.5,0.5,0 ; 0.5,0.5
24.5	22.9	27.7	23.8	23.2

Par exemple, on remarque que si l’on ne tient compte que du paramètre de similarité et pas du nombre de voisins diffuseurs, la diffusion est plus longue que lorsque l’on ne tient compte que du nombre de voisins diffuseurs. Ceci est dû au fait que le paramètre de similarité avec le contenu peut pénaliser la probabilité de transition si elle est inférieure à 0.5, alors que le nombre de voisins ne fait que l’améliorer.

3. Approche stochastique

3.1 Cadre théorique

Processus de comptage : Un processus de comptage est un processus stochastique N(t) avec t≥0 tel que : N(0)=0 ; N prend des valeurs entières et s'incrémente de 1 en 1 avec la réalisation de différents évènement.

Processus de point : un processus de point simple est donné par un vecteur aléatoire T = {T1, T2, …}, prenant des valeurs dans [0, ∞), tel que P(0 ≤ T1 ≤ T2 ≤ …) = 1, et tel que le nombre de points dans une région bornée est fini presque surement.

Fonction d'intensité conditionnelle :

On considère un processus de comptage N(.) et on note H(.) son historique. Si une fonction positive λ*(t) existe telle que :

$$\lambda^*(t) = \lim_{h \to 0} \frac{E[N(t+h) - N(t)|H(t)]}{h}$$

Alors λ*(t) est appelée la fonction intensité conditionnelle de N(.).

Voici une autre définition équivalente à la définition ci-dessus:

$$\lambda^*(t) = \frac{f(t)}{1 - F(t)}$$

3.2 Processus de Hawkes (PH)

Processus auto-excitant : Processus dont une nouvelle arrivée cause l'augmentation de l'intensité conditionnelle.

Processus de Hawkes : Considérons un processus de comptage avec un historique associé H(t), qui satisfait :

$$P\big(N(t+h) - N(t) = m \big| H(t)\big) = \begin{cases} \lambda^*(t)h + o(h) & m = 1 \\ o(h) & m > 1 \\ 1 - \lambda*(t)h + o(h) & m = 0 \end{cases}$$

De plus la fonction d'intensité conditionnelle est de la forme :

$$\lambda*(t) = \lambda + \int_0^t \mu(t-u)dN(u)$$

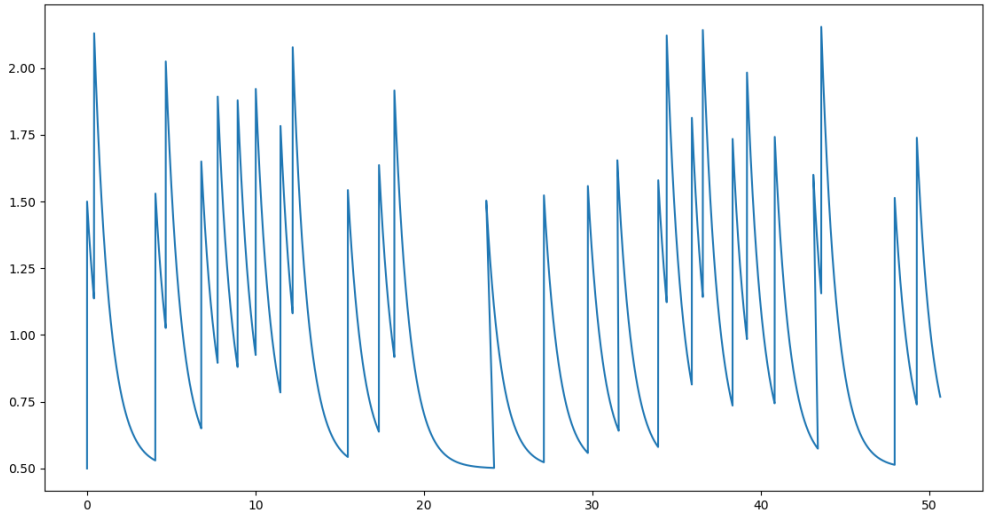
Pour λ>0 et μ :[0, ∞) →[0, ∞) qui sont respectivement appelés l'intensité de seuil et la fonction d'excitation. Supposons que μ(.)≠0 pour éviter le cas trivial qui est un processus de poisson homogène.

Un tel processus est appelé un processus de Hawkes.

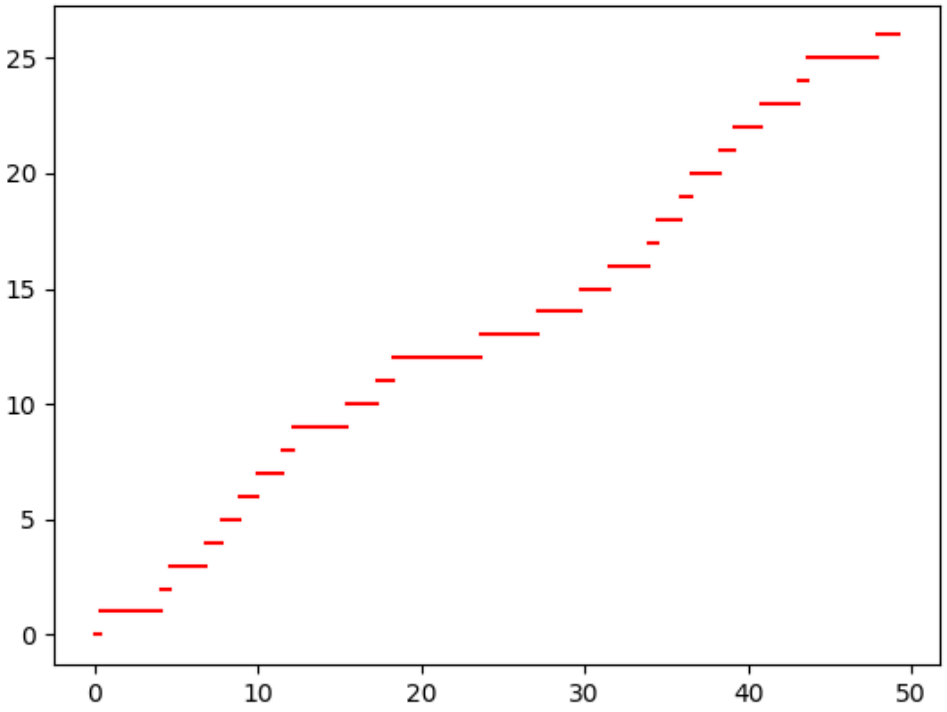
Un choix commun pour la fonction d'excitation est une exponentielle de sorte que :

$$\lambda^*(t) = \lambda + \sum_{t_i<t} \alpha e^{-\beta(t-t_i)}$$

On génère un processus de Hawkes en considérant cette fonction pour cette fonction intensité. En utilisant l'algorithme « by thinning » de l'article [2]



Fonction intensité : λ=0.5 ; α=1 ; β=1.1



3.3 PH appliqué au modèle SIR

Dans cette partie nous nous intéresserons à combiner les équations du modèle SIR avec un processus de Hawkes. Pour cela nous redéfissions la fonction d'excitation :

$$\widetilde{\lambda^*}(t) = \left(1 - \frac{N_t}{N}\right) * \lambda(t) : (2)$$

Où N_t désigne le processus de comptage associé et N la population totale. On désigne par I_t, S_t, R_t, les processus de comptage respectivement associés aux infectés, susceptibles et retirés. Notons H_I, H_R les historiques de comptage (qui suivent un processus de Hawkes). D'après l'article [3] on obtient les équations suivantes :

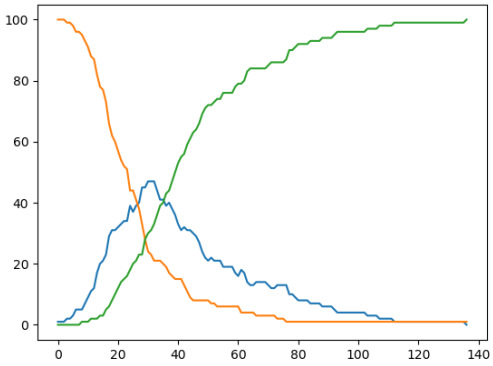
$$\begin{cases} I_t = \sum_{j\geq 1} 1_{t_j^I < t, t_j^R > t} \\ R_t = \sum_{j\geq 1} 1_{t_j^R < t} \\ S_t = N - R_t - I_t \end{cases}$$

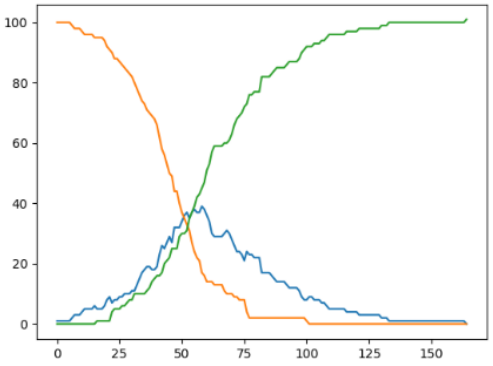
Notons qu'au départ de notre réflexion nous comptions générer les processus de Hawkes puis, fixer une condition pour le choix des temps d'incrémentation de notre système global. Les simulations ont suffi à nous montrer que ce choix n'est pas pertinent. La nouvelle fonction d'excitation ainsi défini en (2) induit une construction récursive de nos processus de comptage.

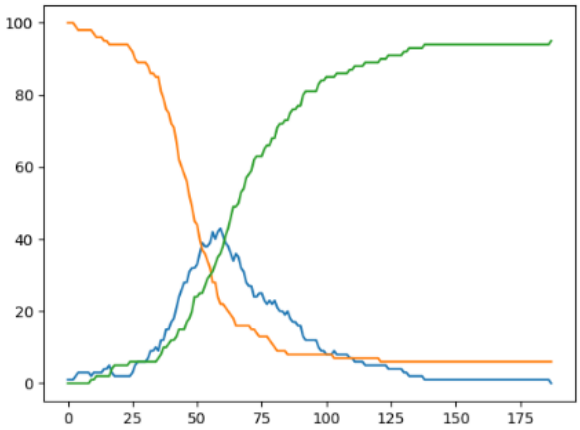
Algorithmiquement on se donne I_0, R_0, S_0, T_{max}. On récupère en sortie un tableau de temps T, et des tableaux contenant les valeurs des populations S, I, R au temps T correspondants.

Initialement t=0, T=[] Q_R=[R_0] NR_0=R_0 etc. On génère également H_I et H_R en utilisant T.

$Tant\ que\ t < T_{max}$
$t = \{t \in H_I \cup H_R | t > T[-1]\}$
$t \to T$
$H_I = HP\_bythinning(H_I, T)$
$H_R = HP\_bythinning(H_R, T)$
$Si\ t \in H_I : NI_t = NI_t + 1\ , Si\ t \in H_R : NR_t = NR_t + 1$
$Q_R \leftarrow Q_{R_0} + NR_t$
$Q_I \leftarrow Q_{I_0} + NI_t - NR_t$
$Q_S \leftarrow Q_{S_0} - NI_t - NR_t$On obtient les courbes suivantes :







Les résultats correspondent aux courbes obtenus avec un modèle SIR déterministe : c'est donc très satisfaisant. Il serait maintenant intéressant d'utiliser un processus de Hawkes multivarié pour rendre un peu plus compte de ce qu'il se passe réellement, en particulier entre les infectés et les retirés, ou dans une autre mesure entre deux groupes de population différentes.

4. Limites et perspectives

Nous avons testé différentes méthodes prenant en compte différents paramètres et différents schémas de diffusion, mais dans l'ensemble nos modèles sont restés assez simplifiés et uniformes. De plus, nous n'avons pu générer, avec les outils à notre disposition, que des graphes de faible dimension, pas du tout représentative d'un véritable réseau. Il faudrait également étudier plus en détail la topologie réelle d'un réseau, plus complexe que ce que nous avons proposé.

Il serait de plus intéressant d'étudier l'influence du diffuseur initial sur la diffusion du message. On pourrait par exemple définir un seuil de contamination à partir duquel un individu est appelé un influenceur et considérer les différents paramètres d'un individu pour qu'il soit considéré comme tel.

Nous avons par ailleurs travaillé avec différents paramètres choisis à peu près aléatoirement, tant que le résultat restait réaliste.

Il faudrait en fait étudier pour un réseau donné les paramètres les plus adéquats pour correspondre au modèle, à partir de données statistiques de diffusions de messages et de méthodes d'estimation. On pourrait alors prévoir par la suite la façon dont un nouveau message au sein d'un réseau dont nous avons déterminé les paramètres peut se transmettre.

Bibliographie :

[1] **Un Modèle de Diffusion de l'Information dans les Réseaux Sociaux** de Cédric Lagnier et Eric Gaussier.

[2] **Hawkes Processes** de Patrick J. Laub, Thomas Taimre, Philip K. Pollett.

[3] **SIR-Hawkes: Linking Epidemic Models and Hawkes Processes to Model Diffusions in Finite Populations** de Marian-Andrei, Swapnil Mishra, Quyu Kong, Mark Carman, Lexing Xie.