

TIPE : Propagation de rumeurs dans un réseau social

Rapport

Hugo LEVY-FALK

2017

Table des matières

1	Préambule	2
2	Introduction	2
3	Propagation de rumeurs	2
3.1	Modalités d'action	2
3.1.1	Étude théorique	2
3.1.2	Génération des graphes	3
3.1.3	Expériences	4
3.2	Résultats	4
3.3	Analyse, exploitation et discussion	6
4	Conclusion	6

1 Préambule

Ce rapport synthétise le travail réalisé sur la propagation de rumeurs dans les réseaux sociaux. On cherche à propager le plus efficacement possible une rumeur à l'ensemble d'un réseau. Les objectifs fixés lors de la mise en cohérence ont été atteints.

2 Introduction

Faisant suite à la recherche documentaire, une première approche théorique de la propagation des rumeurs, s'inspirant du livre de Easley et Kleinberg, a été réalisée. Cela permet de fixer le cadre des simulations ainsi que d'identifier les paramètres des expériences. Dans un deuxième temps trois critères de choix des propagateurs initiaux ont été expérimentés en faisant varier les paramètres mis en évidence précédemment.

3 Propagation de rumeurs

3.1 Modalités d'action

3.1.1 Étude théorique

L'étude théorique a consisté en la formalisation des concepts de rumeur et de propagation. On modélise naturellement un réseau social par un graphe $G = (V, E)$, avec V un ensemble de nœuds et $E \subset V^2$, que l'on supposera connexe par la suite. Chaque membre du réseau est modélisé par un nœud du graphe et ses liens sociaux par des arêtes.

On cherche d'abord à caractériser le comportement d'un nœud. Pour cela on considère qu'un nœud peut se trouver dans deux états distincts : il transmet la rumeur à tous ses voisins (A) ou il ne relaie pas la rumeur (B). Pour chaque nœud $i \in V$, si i est dans l'état A , alors pour chacun de ses voisins également dans l'état A , i réalise un gain a . De même si i est dans l'état B , il réalise un gain b pour chacun de ses voisins dans l'état B . On procède ensuite à une simulation où, à chaque étape, tous les nœuds du graphe tentent de maximiser leur gain en choisissant ou non de passer à l'état A (on n'autorise pas les passages de A vers B).

Lors d'une étape de simulation, si l'on note p la proportion de voisins dans l'état A du nœud i , alors le nœud i maximise son gain en passant à l'état A si $p \times a > (1 - p) \times b$, ou encore

$$p > \frac{b}{a + b} = q$$

Proposition 3.1. *Si $q > 1$, alors la rumeur ne peut pas se propager.*

Proposition 3.2. *Si l'on pose $(V_k)_{k \in \mathbb{N}}$ la suite des nœuds dans l'état A à l'étape k , s'il existe $n \in \mathbb{N}$ tel que $V_n = V_{n+1}$ alors la suite est stationnaire à partir du rang n .*

Démonstration. Si $V_n = V_{n+1}$ alors chaque nœud maximise déjà son gain à l'étape n , la situation étant identique à l'étape $n + 1$, on a $V_{n+1} = V_{n+2}$. \square

Dans la suite on pose $n = |V|$.

Proposition 3.3. *La suite $(V_k)_{k \in \mathbb{N}}$ converge en n étapes au plus.*

Démonstration. La suite $(V_k)_{k \in \mathbb{N}}$ étant croissante pour l'inclusion et les cardinaux des V_k majorés par n , elle est stationnaire. D'autre part, d'après la proposition 3.2, elle est stationnaire à partir du rang $k \in \mathbb{N}$ si $V_k = V_{k+1}$. Or il n'y a que n nœuds, donc la suite est stationnaire à partir du rang $n - 1$ au plus. \square

La proposition 3.3 assure que toutes les simulations pourront être terminées en un temps fini. La proposition 3.2 permet de raccourcir la durée d'une simulation en comptant le nombre d'éléments qui passent à l'état A à chaque étape de simulation. S'il est nul, alors la simulation est terminée.

On cherche ensuite un critère plus complet que la proposition 3.1 pour qualifier la possibilité de propager une rumeur.

Définition 3.1. On appelle p -cluster tout $C \subset V$ tel que pour $i \in C$ il existe $(v_k)_{k \in \llbracket 1, p \rrbracket} \in C^p$ deux à deux distincts et tels que pour tout $k \in \llbracket 1, p \rrbracket$, i et v_k soient voisins.

Théorème 3.1. Une rumeur de note q ne se propagera pas à l'ensemble du réseau si et seulement si il existe un p -cluster non initialement informé avec $p > q$.

Démonstration. S'il existe un p -cluster C avec $p > q$, alors tout nœud de C possède au moins une proportion p de voisins non informés. Ceci valant pour tous les nœuds de C , aucun nœud de C ne sera informé au bout de n étapes. Les clusters sont des obstacles aux rumeurs.

S'il existe un nœud $i \in V$ tel qu'au bout de n étapes i ne soit pas dans l'état informé, alors la proportion p de voisins de i dans l'état informé vérifie $p \leq q$ ou encore $(1 - p) > q \leq 0$. Il existe donc des voisins de i vérifiant cette propriété, on a un z -cluster avec $z > q$. \square

3.1.2 Génération des graphes

Les graphes sont générés avec l'algorithme de Watts-Strogatz qui permet de générer des graphes connexes vérifiant le phénomène du "petit monde" mis en évidence par Milgram.

Données : $N \in \mathbb{N}, K \in \llbracket 1, \lfloor \frac{N}{2} \rfloor \rrbracket (N \gg K \gg \ln N), \beta \in [0, 1]$
Résultat : Matrice d'adjacence d'un graphe aléatoire.
 $M \leftarrow$ matrice avec pour $i \in \llbracket 0, N - 1 \rrbracket, j \in \llbracket 1, K \rrbracket, M_{i, i+j[N]} = M_{i, i-j[N]} = \text{Vrai}$, Faux pour les autres;
pour $i \in \llbracket 0, N - 1 \rrbracket$ **faire**
 pour $j \in \llbracket 1, K \rrbracket$ **faire**
 $r \leftarrow$ Nombre aléatoire sur $[0, 1]$;
 si $r < \beta$ **alors**
 $M_{i, i+j[N]} \leftarrow \text{Faux}$;
 $M_{i+j[N], i} \leftarrow \text{Faux}$;
 Choisir au hasard k tel que $M_{i, k} = \text{Faux}$;
 $M_{i, k} \leftarrow \text{Vrai}$;
 $M_{k, i} \leftarrow \text{Vrai}$;
 finsi
 finpour
finpour
retourner M

Algorithm 1 : Algorithme de Watts-Strogatz

Pour les expériences on fixera $N = 500$ et $K = 50$. On fera varier β dans $\{0, 0.25, 0.5, 0.75, 1\}$.

3.1.3 Expériences

Il est difficile de définir ce qui est attendu d'une propagation optimale d'une rumeur. On comparera donc trois méthodes de choix des propagateurs initiaux.

- Le tirage au sort, à des fins de comparaison ;
- Les éléments de plus grand degré. En effet une première analyse qualitative laisse penser que ces nœuds seront plus à même de "convaincre" une grande partie du réseau ;
- Les éléments les plus centraux, c'est à dire ceux par lesquels passent la plus grande proportion de plus courts chemins. Pour cela on utilise l'algorithme de Ulrik Brandes. Cela caractérise la capacité d'un nœud à atteindre rapidement les autres nœuds du graphe. Formellement,

$$\forall v \in V, C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

avec $\sigma_{st}(v)$ le nombre de plus courts chemins de s à t passant par v et σ_{st} le nombre de plus courts chemins de s à t .

Pour mesurer la capacité de propagation moyenne d'une méthode de choix, on réalise 100 fois chaque expériences. Étant donné la longueur des calculs, un système de sauvegarde de l'avancement d'une expérience dans une base de données a été implémenté. Ceci permet d'interrompre une expérience et de la reprendre au même endroit lorsque le programme est relancé.

3.2 Résultats

On représente les résultats sous la forme d'une courbe proportion finale = f (proportion initiale). La courbe type obtenue est décrite à la figure 1.

Taille du graphe=500 Taille de l'échantillon=100 K=50, q=0.5, Beta =50%

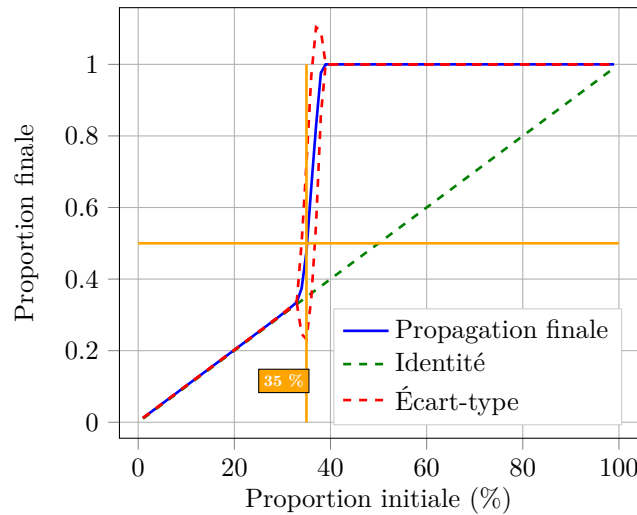


Figure 1 – Courbe type (en pointillés rouge l'écart type)

Le paramètre β a essentiellement une influence dans le choix par centralité. Les résultats sont résumés dans les figures 2, 3 et 4.

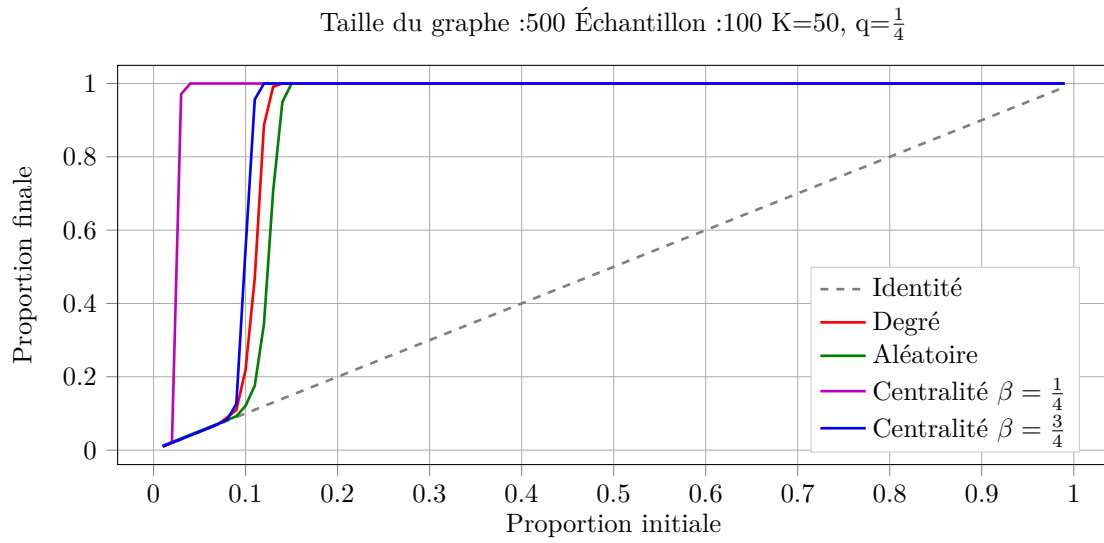


Figure 2 – Résultats

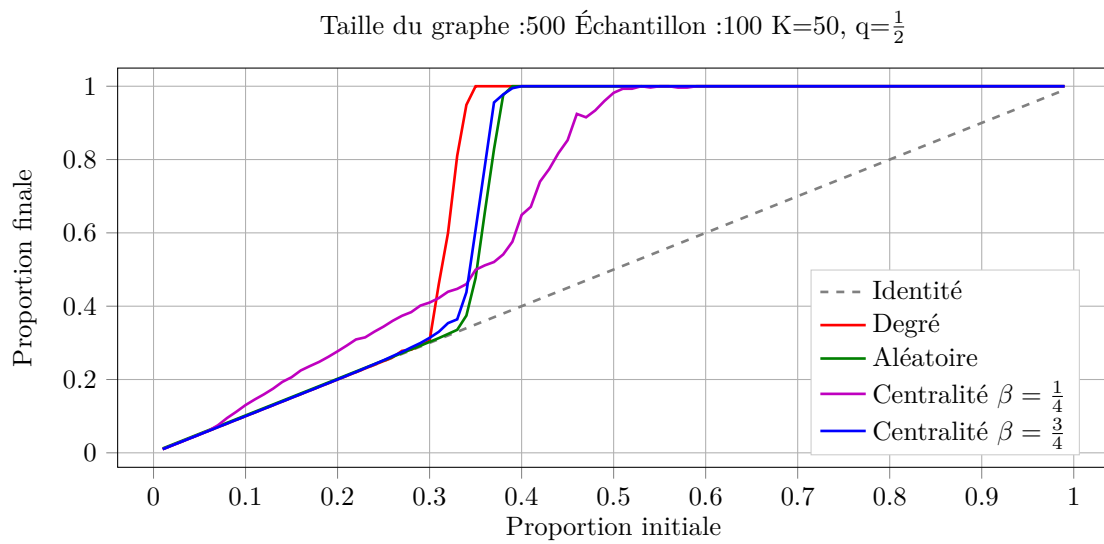


Figure 3 – Résultats

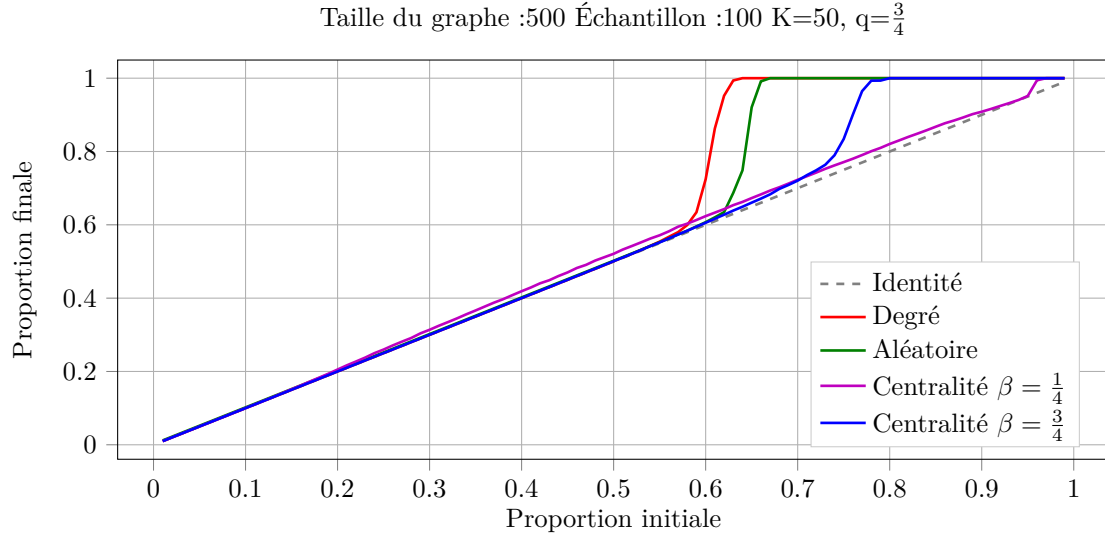


Figure 4 – Résultats

3.3 Analyse, exploitation et discussion

Les résultats montrent que pour des rumeurs relativement convaincantes, c'est-à-dire q faible (ici $< 25\%$), le choix par centralité est plus intéressant. En effet pour $\beta = \frac{1}{4}$, une proportion initiale de 5% des plus centraux permet d'atteindre tout le graphe là où il faut 10% des éléments de plus hauts degrés.

En revanche, pour des rumeurs avec q élevé ($> 50\%$), la méthode des plus hauts degrés est à privilégier.

Les résultats devraient être confirmés par une étude sur un échantillon plus important de graphes. De plus, les graphes produits par l'algorithme sont irréalistes de par le degré des nœuds et le paramètre β dont il est difficile de mesurer l'influence.

Une étude plus précise prenant en compte le caractère cinétique de la propagation (nombre d'étapes pour atteindre tous les nœuds) serait nécessaire pour la tranche $\frac{1}{4} < q < \frac{1}{2}$.

4 Conclusion

Cette étude met en évidence une stratégie de choix des éléments propagateurs initiaux, bien que les résultats obtenus demandent une étude complémentaire qui n'a pas été réalisée du fait de la longueur des calculs (l'étude actuelle a déjà nécessité plusieurs centaines d'heures de simulation). Lorsque la rumeur est convaincante, il est préférable de choisir ces éléments pour leur centralité élevée. Sinon, lorsque la rumeur est moins convaincante, il est plus efficace de choisir les éléments de plus haut degré.