

TIPE : Propagation de rumeurs dans un réseau social

Hugo LEVY-FALK

2017

Plan

- 1 Rappel de la problématique
- 2 Modélisation
- 3 Génération de graphes
- 4 Expériences
- 5 Résultats
- 6 Conclusion

Comment propager une rumeur le plus rapidement possible à un maximum de nœuds d'un réseau social ?

Plan

- 1 Rappel de la problématique
- 2 Modélisation
- 3 Génération de graphes
- 4 Expériences
- 5 Résultats
- 6 Conclusion

On modélise un réseau social par un graphe.

On modélise un réseau social par un graphe.

- Personne \rightarrow Nœud

On modélise un réseau social par un graphe.

- Personne \rightarrow Nœud
- Lien social \rightarrow Arrête

On modélise un réseau social par un graphe.

- Personne → Nœud
- Lien social → Arrête

On ne prend pas en compte la "qualité" de la relation.

Rappel de la
problématique

Modélisation

Réseau social

**Caractéristiques des
réseaux simulés**

Simulation de
propagation

Propagation optimale

Génération de
graphes

Expériences

Résultats

Conclusion

- Stanley Milgram : Six degrés de séparation (Facebook 4.57)

- Stanley Milgram : Six degrés de séparation (Facebook 4.57)
- Algorithme de Watts-Strogatz

Rappel de la
problématique

Modélisation

Réseau social

Caractéristiques des
réseaux simulés

**Simulation de
propagation**

Propagation optimale

Génération de
graphes

Expériences

Résultats

Conclusion

- Chaque nœud maximise son gain.

- Chaque nœud maximise son gain.
- Un voisin dans l'état "informé" \rightarrow gain a
- Un voisin dans l'état "non-informé" \rightarrow gain b

- Chaque nœud maximise son gain.
- Un voisin dans l'état "informé" \rightarrow gain a
- Un voisin dans l'état "non-informé" \rightarrow gain b

Si on note p la proportion de voisins informés, le nœud maximise son gain en passant à l'état informé si et seulement si $p \times a > (1 - p) \times b$, ou encore

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

- Chaque nœud maximise son gain.
- Un voisin dans l'état "informé" → gain a
- Un voisin dans l'état "non-informé" → gain b

Si on note p la proportion de voisins informés, le nœud maximise son gain en passant à l'état informé si et seulement si $p \times a > (1 - p) \times b$, ou encore

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

→ On caractérise une rumeur par $q = \frac{b}{a+b}$.

→ On caractérise une rumeur par $q = \frac{b}{a+b}$.

Remarques

Soit un graphe $G = (V, E)$ avec V un ensemble de nœuds et $E \subset V^2$.

- Pas de propagation si $q > 1$;

→ On caractérise une rumeur par $q = \frac{b}{a+b}$.

Remarques

Soit un graphe $G = (V, E)$ avec V un ensemble de nœuds et $E \subset V^2$.

- Pas de propagation si $q > 1$;
- Si l'on pose $(V_k)_{k \in \mathbb{N}}$ une suite des nœuds dans l'état "informé" 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 ;

→ On caractérise une rumeur par $q = \frac{b}{a+b}$.

Remarques

Soit un graphe $G = (V, E)$ avec V un ensemble de nœuds et $E \subset V^2$.

- Pas de propagation si $q > 1$;
- Si l'on pose $(V_k)_{k \in \mathbb{N}}$ une suite des nœuds dans l'état "informé" 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 ;
- La suite étant par ailleurs croissante pour l'inclusion et majorée, la suite converge et on finit une simulation en au plus $|V|$ étapes.

Définition : p -cluster

Soit un graphe $G = (V, E)$ avec V un ensemble de nœuds et $E \subset V^2$. On appelle p -cluster tout sous-ensemble $C \subset V$ tel que pour tout $i \in C$ il existe un p -uplet $(v_k)_{k \in \llbracket 1, p \rrbracket} \in C^p$ deux à deux distincts et tel que pour tout $k \in \llbracket 1, p \rrbracket$, i et v_k soient voisins.

Remarque

Si le graphe est connexe (cas des graphes étudiés), l'ensemble forme un 1-cluster.

Les clusters sont les seuls obstacles aux rumeurs

Propagation
de rumeurs
dans un
réseau social

Hugo
LEVY-FALK

Rappel de la
problématique

Modélisation

Réseau social

Caractéristiques des
réseaux simulés

Simulation de
propagation

Propagation optimale

Génération de
graphes

Expériences

Résultats

Conclusion

Théorème

Les clusters sont les seuls obstacles aux rumeurs.

On pose $n = |V|$, q la note de la rumeur.

Théorème

Les clusters sont les seuls obstacles aux rumeurs.

On pose $n = |V|$, q la note de la rumeur. 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.

Théorème

Les clusters sont les seuls obstacles aux rumeurs.

On pose $n = |V|$, q la note de la rumeur. S'il existe un nœud i 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 $(1 - p) > q \geq 0$. Il existe donc des voisins de i vérifiant cette propriété, on a un z -cluster avec $z > q$.

Comment caractériser une propagation optimale ?

Rappel de la
problématique

Modélisation

Réseau social

Caractéristiques des
réseaux simulés

Simulation de
propagation

Propagation optimale

Génération de
graphes

Expériences

Résultats

Conclusion

- Capacité à atteindre l'ensemble du graphe ;

Comment caractériser une propagation optimale ?

Rappel de la
problématique

Modélisation

Réseau social
Caractéristiques des
réseaux simulés

Simulation de
propagation

Propagation optimale

Génération de
graphes

Expériences

Résultats

Conclusion

- Capacité à atteindre l'ensemble du graphe ;
- Nombre d'itérations de simulation le plus faible possible ;

Comment caractériser une propagation optimale ?

Rappel de la
problématique

Modélisation

Réseau social

Caractéristiques des
réseaux simulés

Simulation de
propagation

Propagation optimale

Génération de
graphes

Expériences

Résultats

Conclusion

- Capacité à atteindre l'ensemble du graphe ;
- Nombre d'itérations de simulation le plus faible possible ;

Problème(s) : Unicité de la solution ? identification des propriétés permettant une telle propagation ?

Comment caractériser une propagation optimale ?

Rappel de la
problématique

Modélisation

Réseau social

Caractéristiques des
réseaux simulés

Simulation de
propagation

Propagation optimale

Génération de
graphes

Expériences

Résultats

Conclusion

- Capacité à atteindre l'ensemble du graphe ;
- Nombre d'itérations de simulation le plus faible possible ;

Problème(s) : Unicité de la solution ? identification des propriétés permettant une telle propagation ?

→ Comparaison de critères arbitraires.

Plan

- 1 Rappel de la problématique
- 2 Modélisation
- 3 Génération de graphes
- 4 Expériences
- 5 Résultats
- 6 Conclusion

Algorithme de Watts-Strogatz

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 $M_{i,k} = \text{Faux}$;

$M_{i,k} \leftarrow \text{Vrai}$;

$M_{k,i} \leftarrow \text{Vrai}$;

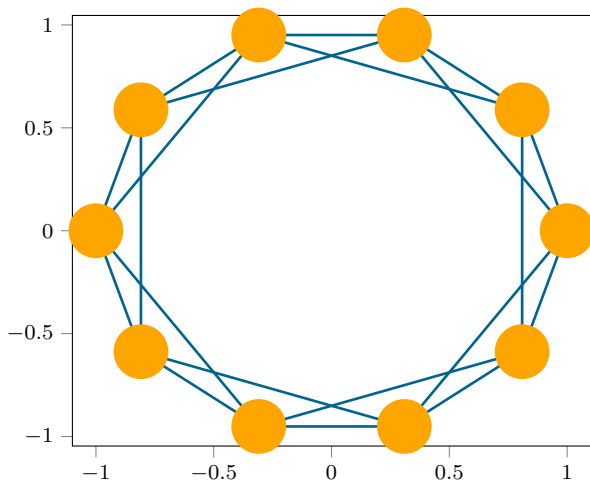
fin

fin

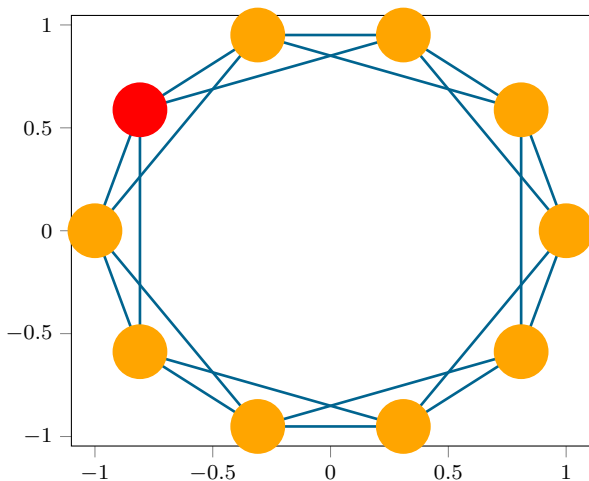
fin

retourner M

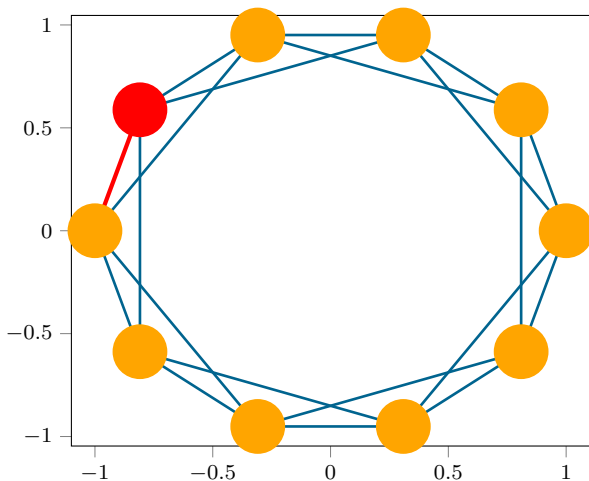
Algorithme de Watts-Strogatz



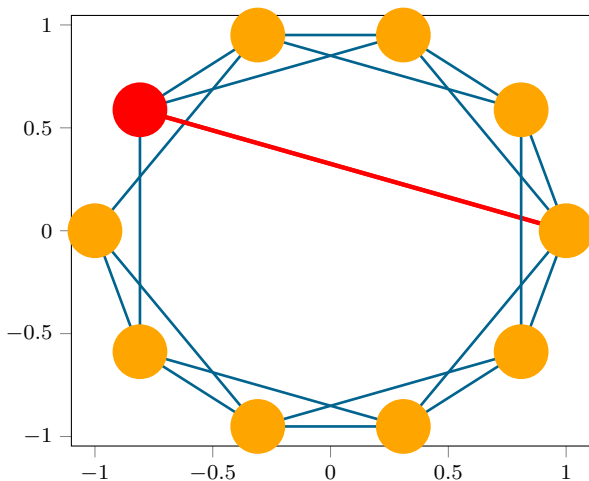
Algorithme de Watts-Strogatz



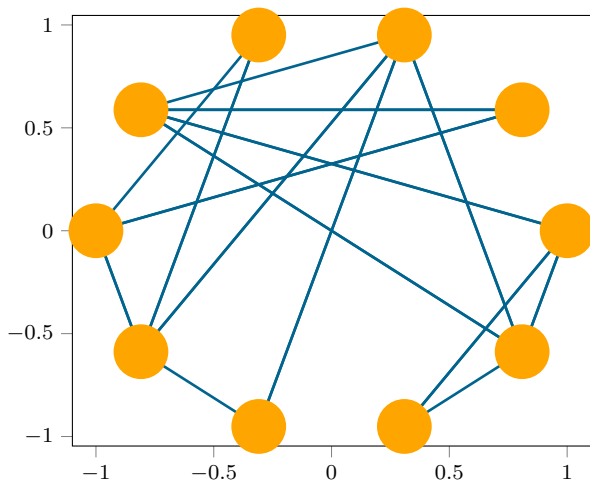
Algorithme de Watts-Strogatz



Algorithme de Watts-Strogatz



Algorithme de Watts-Strogatz



- 500 nœuds ;

- 500 nœuds ;
- Au plus 500 étapes de simulation ;

- 500 nœuds ;
- Au plus 500 étapes de simulation ;
- On lance la simulation 100 fois ;

- 500 nœuds ;
- Au plus 500 étapes de simulation ;
- On lance la simulation 100 fois ;
- 3 paramètres à examiner (β , q , proportion initiale d'informés)

- 500 nœuds ;
- Au plus 500 étapes de simulation ;
- On lance la simulation 100 fois ;
- 3 paramètres à examiner (β , q , proportion initiale d'informés)

→ Stockage des résultats dans une base de donnée des résultats des calculs afin de pouvoir interrompre l'expérience à tout instant.

Plan

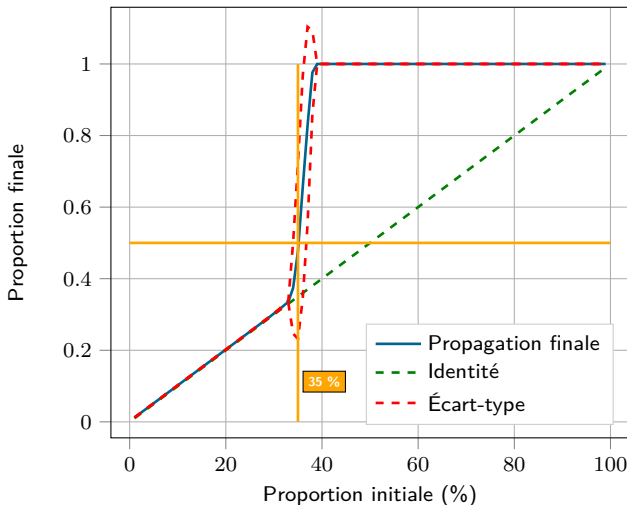
- 1 Rappel de la problématique
- 2 Modélisation
- 3 Génération de graphes
- 4 Expériences
- 5 Résultats
- 6 Conclusion

On fixe $K = 50$. Pour $\beta \in \{0, 0.25, 0.5, 1\}$ et $q \in \{0.25, 0.5, 0.75\}$, pour une proportion initiale de 1% à 99% faire 100 expériences de propagation en choisissant les éléments initiaux au hasard et stocker la propagation à chaque étape de la simulation.

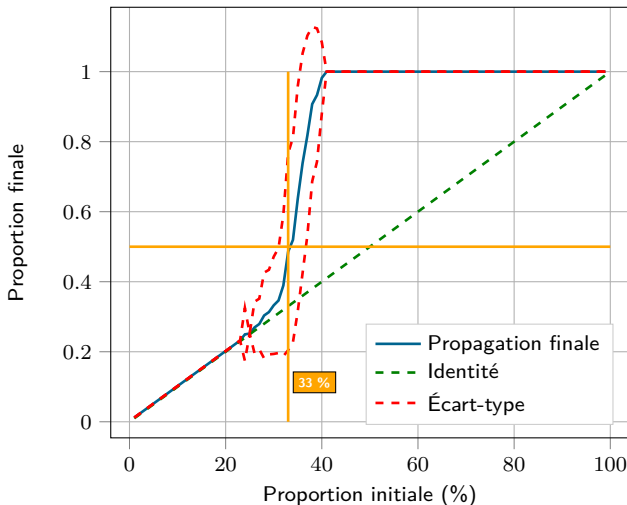
On fixe $K = 50$. Pour $\beta \in \{0, 0.25, 0.5, 1\}$ et $q \in \{0.25, 0.5, 0.75\}$, pour une proportion initiale de 1% à 99% faire 100 expériences de propagation en choisissant les éléments initiaux au hasard et stocker la propagation à chaque étape de la simulation.

But : pouvoir comparer les résultats des autres expériences, éventuellement fixer certains paramètres qui ont peu d'influence.

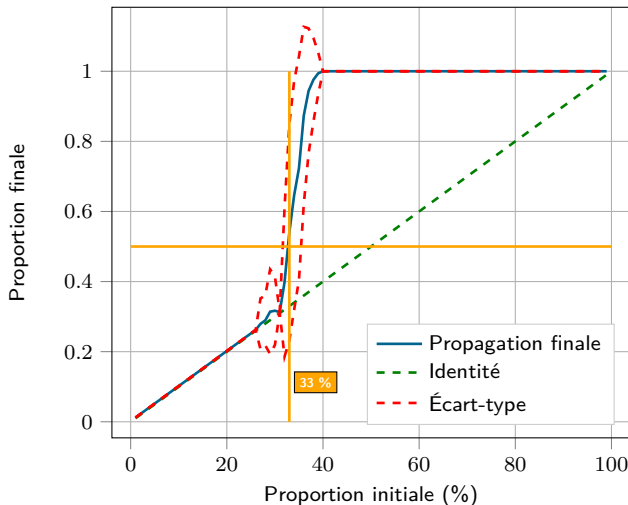
Taille du graphe=500 Nb gen=100 $K=50$, $q=0.5$, Beta =50%



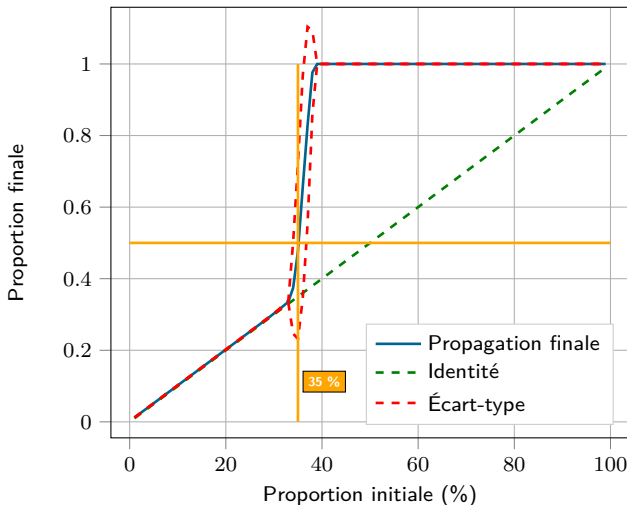
Taille du graphe=500 Nb gen=100 K=50, q=0.5, Beta =0%



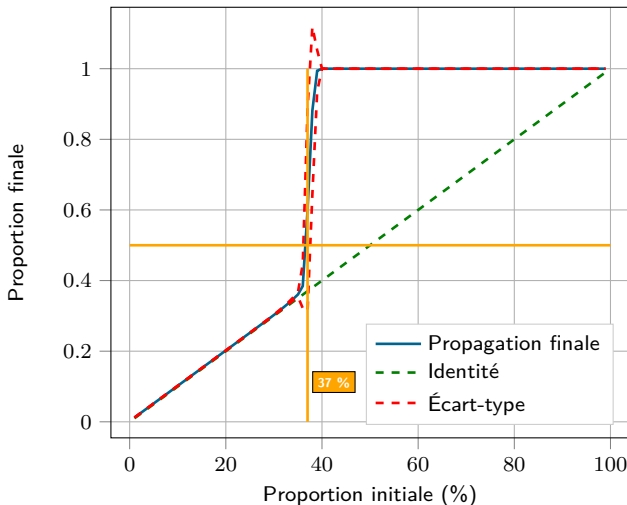
Taille du graphe=500 Nb gen=100 K=50, q=0.5, Beta =25%



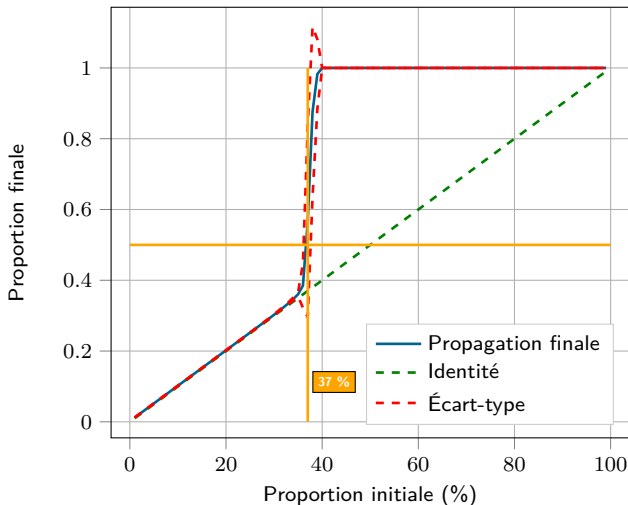
Taille du graphe=500 Nb gen=100 K=50, q=0.5, Beta =50%



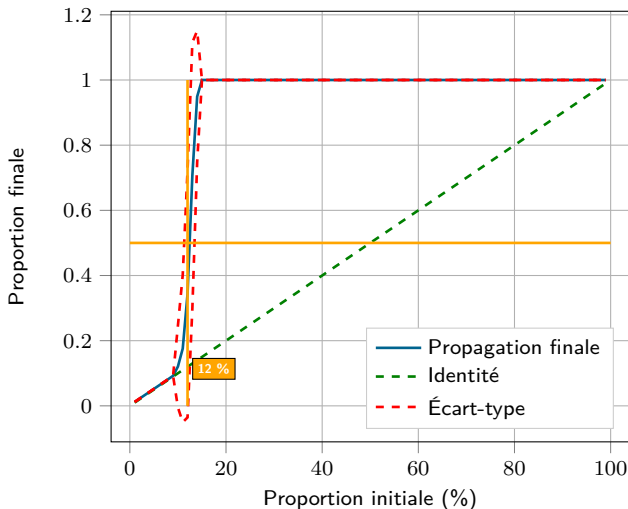
Taille du graphe=500 Nb gen=100 K=50, q=0.5, Beta =75%



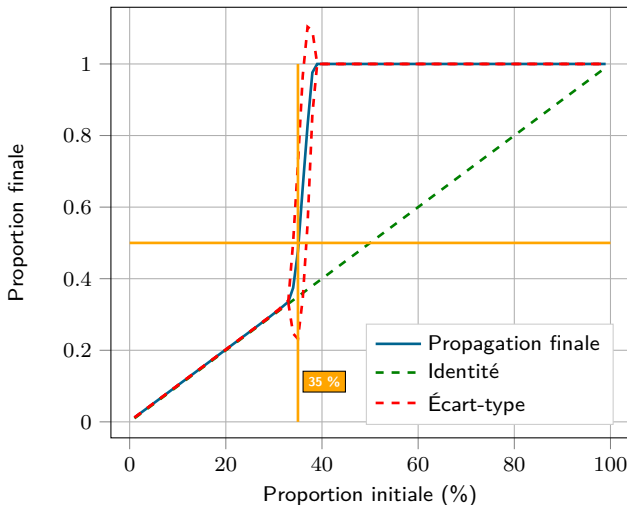
Taille du graphe=500 Nb gen=100 K=50, q=0.5, Beta =100%



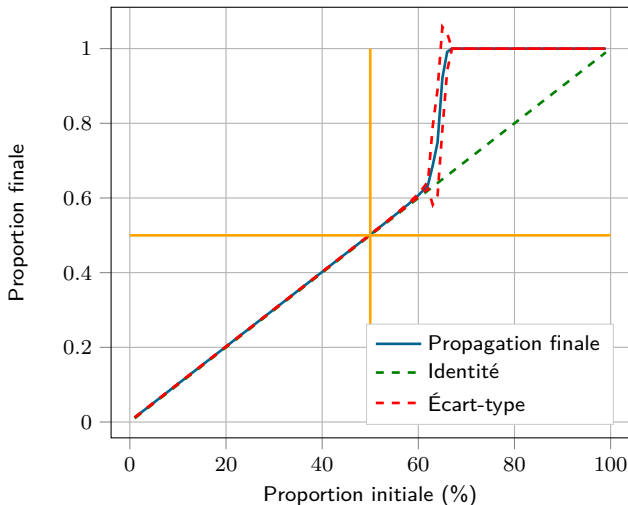
Taille du graphe=500 Nb gen=100 $K=50$, $q=0.25$, Beta =50%



Taille du graphe=500 Nb gen=100 $K=50$, $q=0.5$, Beta =50%



Taille du graphe=500 Nb gen=100 $K=50$, $q=0.75$, Beta =50%



On fixe $K = 50$. Pour $\beta \in \{0, 0.25, 0.5, 1\}$ et $q \in \{0.25, 0.5, 0.75\}$, pour une proportion initiale de 1% à 99% faire 100 expériences de propagation en choisissant les éléments initiaux possédant les plus grands degrés et stocker la propagation à chaque étape de la simulation.

- Le résultat pour $\beta = 0$ est inexploitable ;

- Le résultat pour $\beta = 0$ est inexploitable ;
- On retrouve les mêmes effets qualitatifs de β et q .

On fixe $K = 50$. Pour $\beta \in \{0, 0.25, 0.5, 1\}$ et $q \in \{0.25, 0.5, 0.75\}$, pour une proportion initiale de 1% à 99% faire 100 expériences de propagation en choisissant les éléments initiaux possédant les plus grandes centralités (proportion de plus courts chemins passants par un nœud) et stocker la propagation à chaque étape de la simulation.

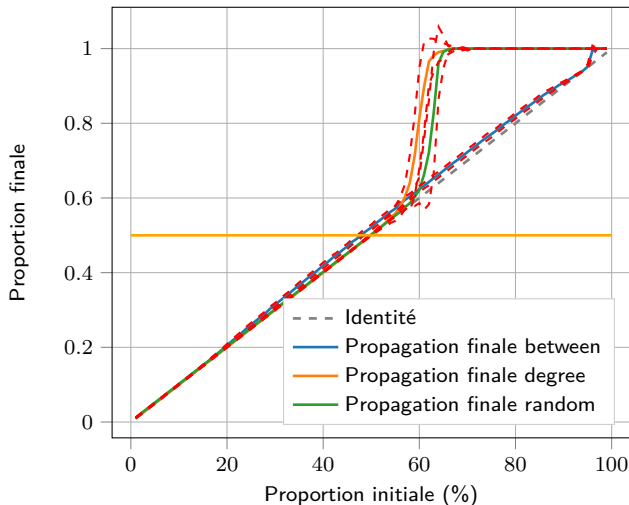
- Le résultat pour $\beta = 0$ est inexploitable ;

- Le résultat pour $\beta = 0$ est inexploitable ;
- On retrouve les mêmes effets qualitatifs de q .

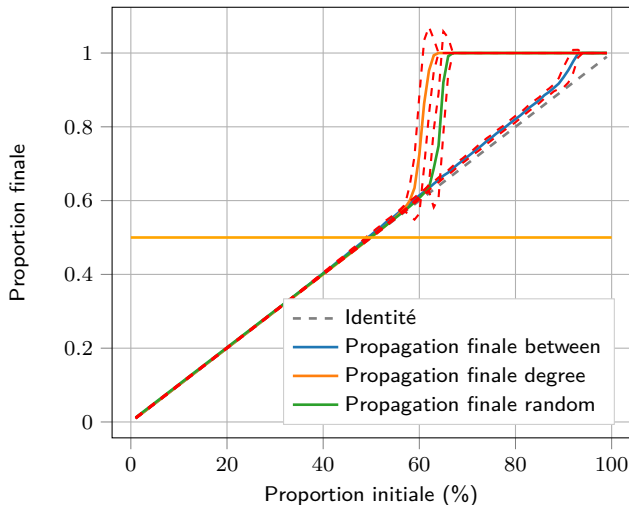
Plan

- 1 Rappel de la problématique
- 2 Modélisation
- 3 Génération de graphes
- 4 Expériences
- 5 Résultats
- 6 Conclusion

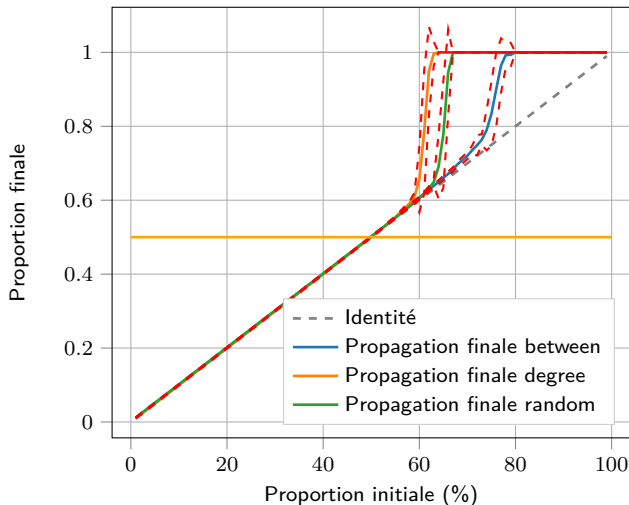
Taille du graphe=500 Nb gen=100 K=50, $q=0.75$, Beta =25%



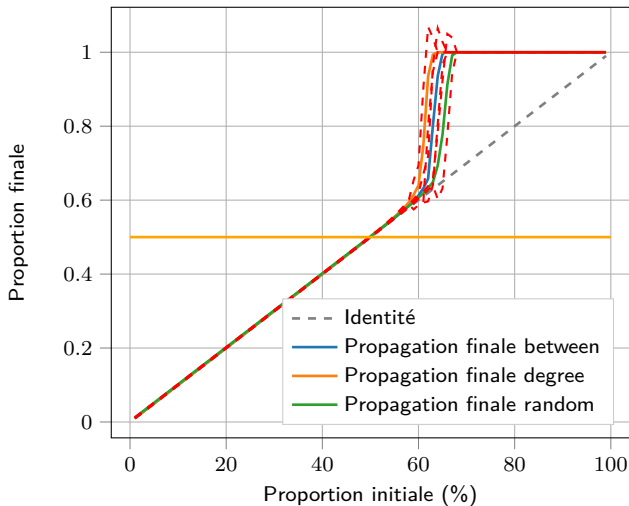
Taille du graphe=500 Nb gen=100 K=50, $q=0.75$, Beta =50%



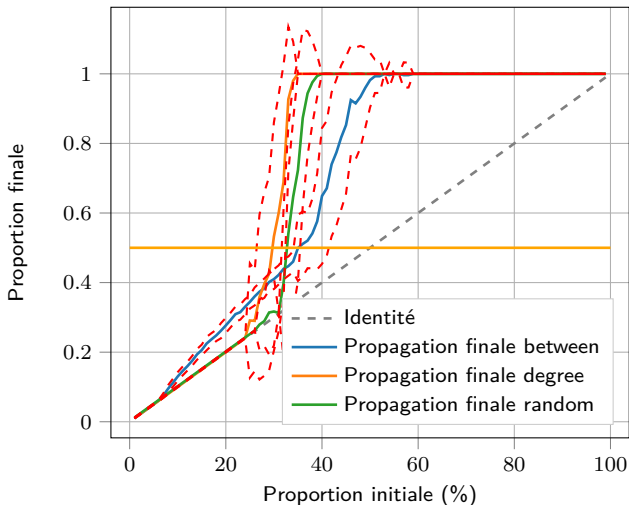
Taille du graphe=500 Nb gen=100 K=50, $q=0.75$, Beta =75%



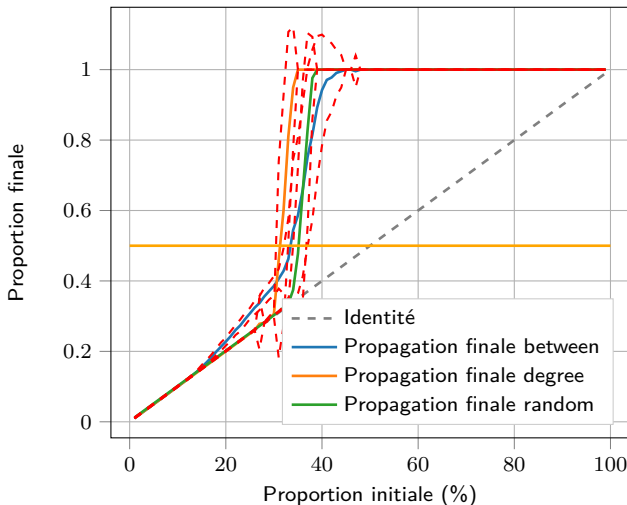
Taille du graphe=500 Nb gen=100 K=50, $q=0.75$, Beta =100%



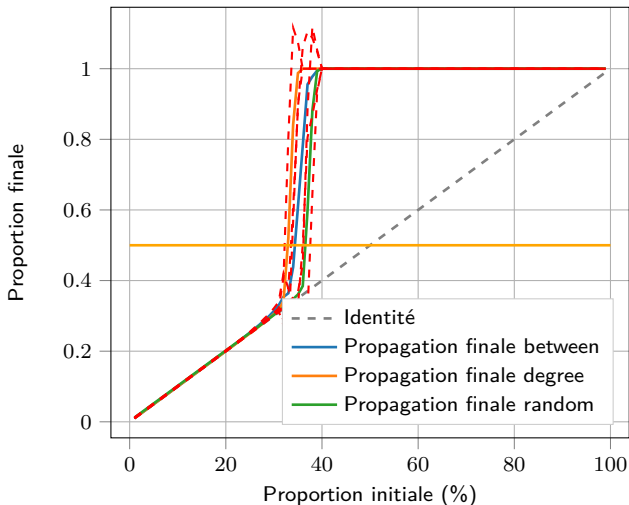
Taille du graphe=500 Nb gen=100 K=50, $q=0.5$, Beta =25%



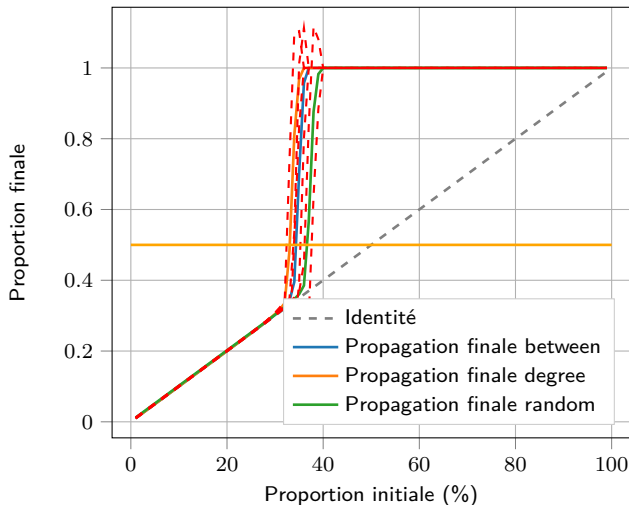
Taille du graphe=500 Nb gen=100 K=50, $q=0.5$, Beta =50%



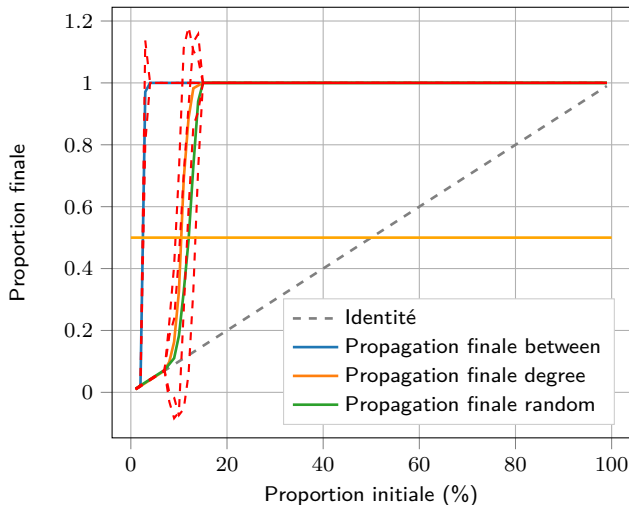
Taille du graphe=500 Nb gen=100 K=50, $q=0.5$, Beta =75%



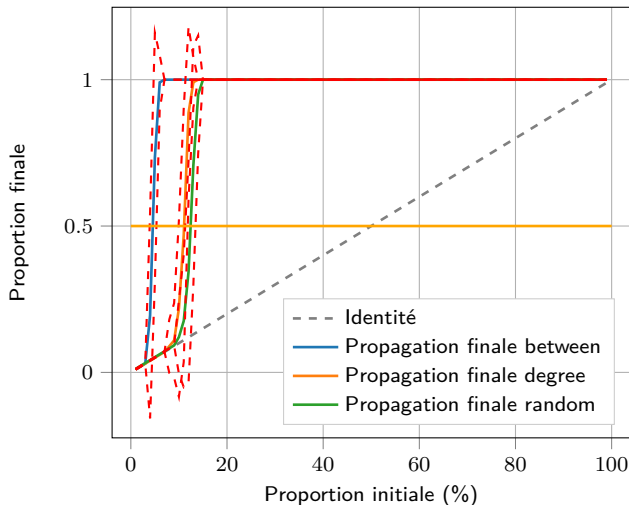
Taille du graphe=500 Nb gen=100 K=50, $q=0.5$, Beta =100%



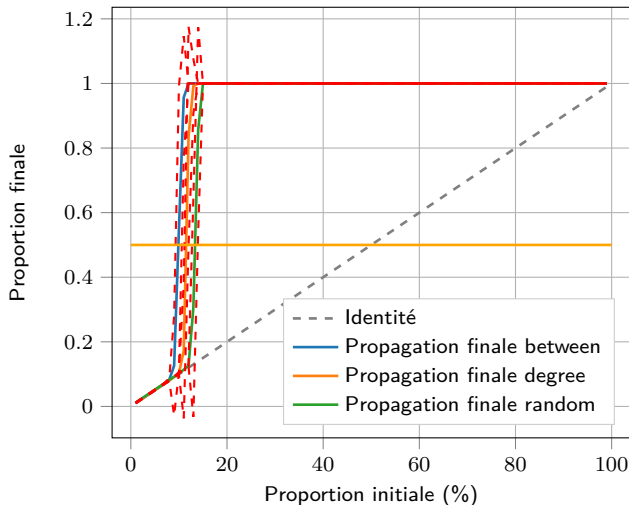
Taille du graphe=500 Nb gen=100 K=50, $q=0.25$, Beta =25%



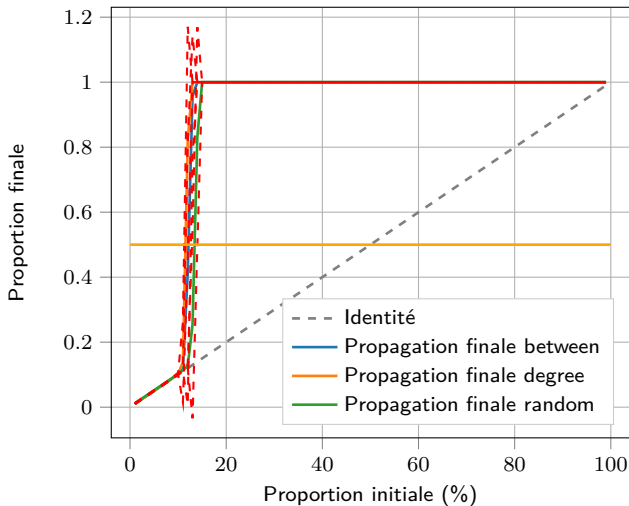
Taille du graphe=500 Nb gen=100 K=50, $q=0.25$, Beta =50%



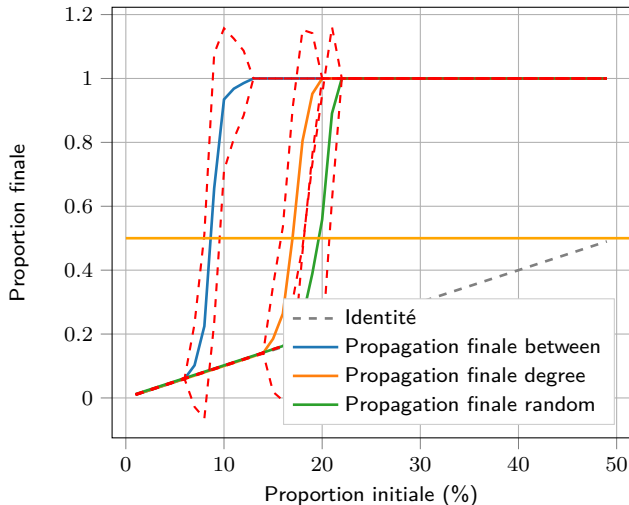
Taille du graphe=500 Nb gen=100 K=50, $q=0.25$, Beta =75%



Taille du graphe=500 Nb gen=100 K=50, $q=0.25$, Beta =100%



Taille du graphe=500 Nb gen=50 K=50, $q=0.333$, Beta =50%



On peut choisir les éléments initialement propagateurs en connaissant q .

Pour $\beta = 50\%$,

On peut choisir les éléments initialement propagateurs en connaissant q .

Pour $\beta = 50\%$,

- Si $q \geq \frac{1}{2}$ On choisit une proportion $\approx q$ des éléments de plus haut degré ;

On peut choisir les éléments initialement propagateurs en connaissant q .

Pour $\beta = 50\%$,

- Si $q \geq \frac{1}{2}$ On choisit une proportion $\approx q$ des éléments de plus haut degré ;
- Si $q \leq \frac{1}{3}$ On choisit une proportion $\approx \frac{q}{2}$ des éléments de plus haut degré ;

On peut choisir les éléments initialement propagateurs en connaissant q .

Pour $\beta = 50\%$,

- Si $q \geq \frac{1}{2}$ On choisit une proportion $\approx q$ des éléments de plus haut degré ;
- Si $q \leq \frac{1}{3}$ On choisit une proportion $\approx \frac{q}{2}$ des éléments de plus haut degré ;
- Une étude plus quantitative serait nécessaire pour $\frac{1}{3} < q < \frac{1}{2}$.

Plan

- 1 Rappel de la problématique
- 2 Modélisation
- 3 Génération de graphes
- 4 Expériences
- 5 Résultats
- 6 Conclusion

- On a un premier critère de choix des éléments initiaux

- On a un premier critère de choix des éléments initiaux
 - Nécessite d'être affiné

- On a un premier critère de choix des éléments initiaux
 - Nécessite d'être affiné
 - Problème : longueur des calculs

- On a un premier critère de choix des éléments initiaux
 - Nécessite d'être affiné
 - Problème : longueur des calculs
- Est-ce le meilleur critère ? (vérification difficile à cause de la longueur des calculs)

- On a un premier critère de choix des éléments initiaux
 - Nécessite d'être affiné
 - Problème : longueur des calculs
- Est-ce le meilleur critère ? (vérification difficile à cause de la longueur des calculs)
- Certains choix de modélisation sont discutables (Non "retour en arrière" de la rumeur)

- On a un premier critère de choix des éléments initiaux
 - Nécessite d'être affiné
 - Problème : longueur des calculs
- Est-ce le meilleur critère ? (vérification difficile à cause de la longueur des calculs)
- Certains choix de modélisation sont discutables (Non "retour en arrière" de la rumeur)
- La méthode de génération des graphes est également problématique : degré des nœuds, choix de β ?