TIPE : Propagation de rumeurs dans un réseau social

Hugo LEVY-FALK

2017

Plan

Rappel de la problématique

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

Plan

Modélisation

On modélise un réseau social par un graphe.

Modélisation

On modélise un réseau social par un graphe.

- Personne $\rightarrow$ Nœud

Modélisation

On modélise un réseau social par un graphe.

- Personne $\rightarrow$ Nœud
- Lien social $\rightarrow$ Arrête

Modélisation

On modélise un réseau social par un graphe.

- Personne $\rightarrow$ Nœud
- Lien social $\rightarrow$ Arrête

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

Caractéristiques des réseaux simulés

Caractéristiques des réseaux simulés

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

Caractéristiques des réseaux simulés

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

Jeu de coordination

Jeu de coordination

- Chaque nœud maximise son gain.

Jeu de coordination

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

Jeu de coordination

- 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

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

Jeu de coordination

- 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

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

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

Jeu de coordination

$\rightarrow$ 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 $q > 1$;

Jeu de coordination

$\rightarrow$ 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 ;

Jeu de coordination

$\rightarrow$ 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.

Cluster

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 $(v_k)_{k \in [\![1,p]\!]} \in C^p$ deux à deux distincts et tel que pour tout $k \in [\![1, p]\!]$, 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

Théorème

Les clusters sont les seuls obstacles aux rumeurs.

Comment caractériser une propagation optimale ?

- Capacité à atteindre l'ensemble du graphe ;

Comment caractériser une propagation optimale ?

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

Comment caractériser une propagation optimale ?

- 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 ?

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

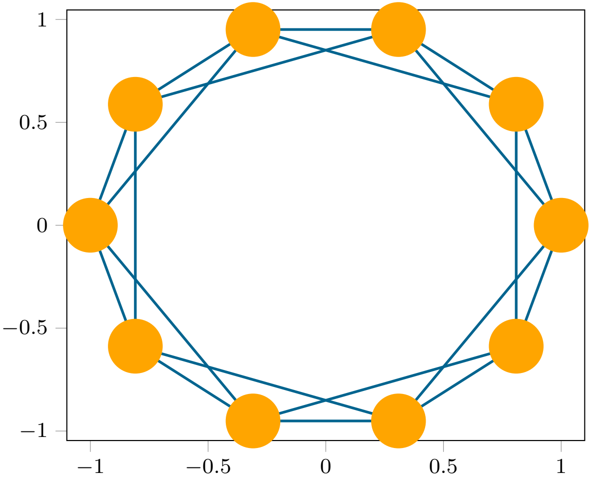
Problème(s) : Unicité de la solution ? identification des propriétés permettant une telle propagation ?
$\rightarrow$ Comparaison de critères arbitraires.

Plan

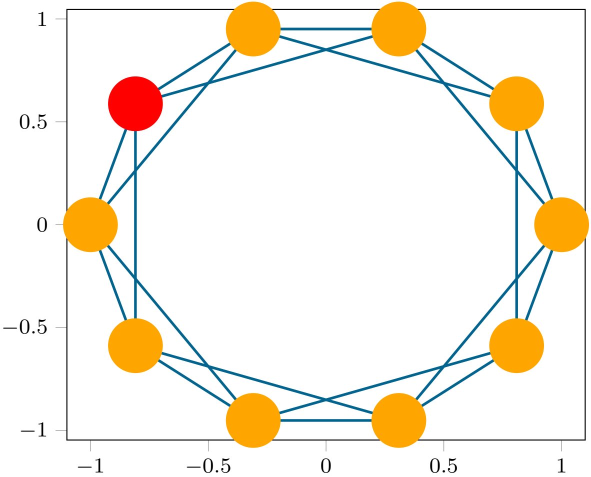
Algorithme de Watts-Strogatz

$$N \in \mathbb{N}, K \in [\![1, \lfloor \frac{N}{2} \rfloor]\!] (N \gg K \gg \ln N), \beta \in [0, 1]$$

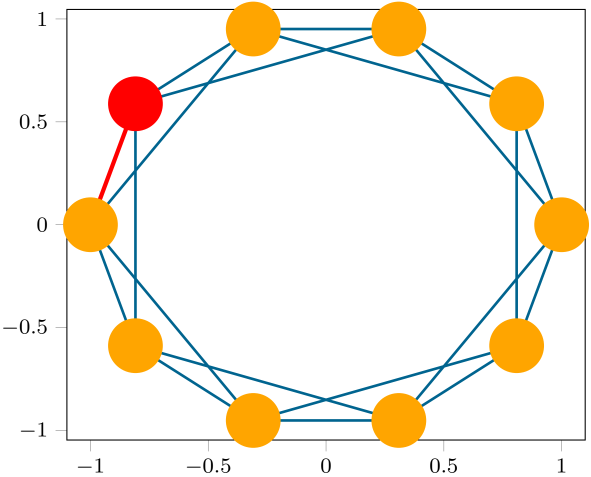
Algorithme de Watts-Strogatz



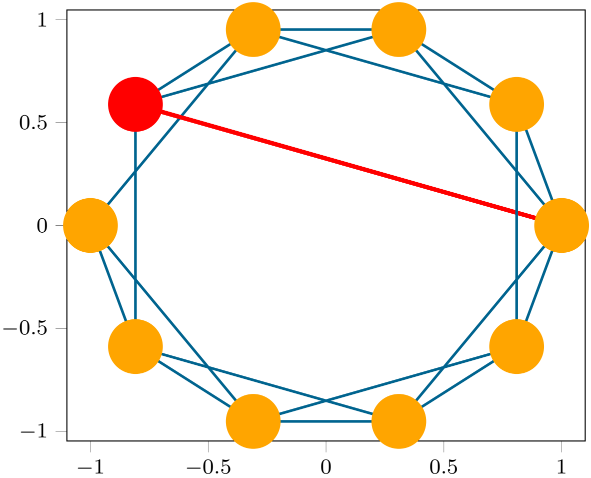
Algorithme de Watts-Strogatz



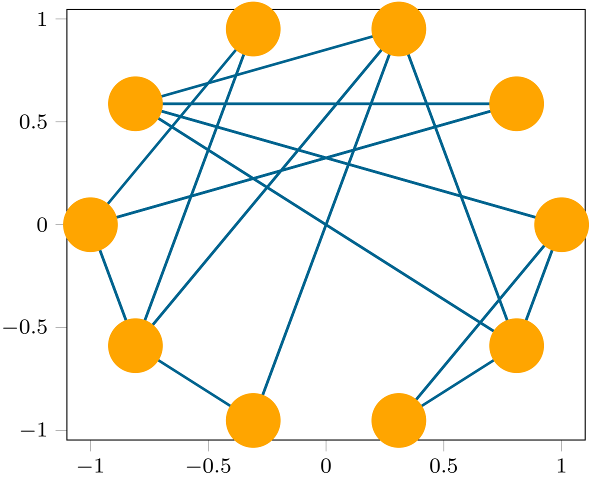
Algorithme de Watts-Strogatz



Algorithme de Watts-Strogatz



Algorithme de Watts-Strogatz



Réseaux simulés

- 500 nœuds ;

Réseaux simulés

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

Réseaux simulés

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

Réseaux simulés

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

Réseaux simulés

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

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

Plan

Protocole expérimental

- On fixe $K = 50$;

Protocole expérimental

- On fixe $K = 50$;
- $\beta \in \{0, \frac{1}{4}, \frac{1}{2}, 1\}$;

Protocole expérimental

- On fixe $K = 50$;
- $\beta \in \{0, \frac{1}{4}, \frac{1}{2}, 1\}$;
- $q \in \{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$;

Protocole expérimental

- On fixe $K = 50$;
- $\beta \in \{0, \frac{1}{4}, \frac{1}{2}, 1\}$;
- $q \in \{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$;
- proportion initiale de 1% à 99% ;

Protocole expérimental

- On fixe $K = 50$;
- $\beta \in \{0, \frac{1}{4}, \frac{1}{2}, 1\}$;
- $q \in \{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$;
- proportion initiale de 1% à 99% ;

$\rightarrow$ 100 expériences de propagation en choisissant les éléments initiaux au hasard et stocker la propagation à chaque étape de la simulation.

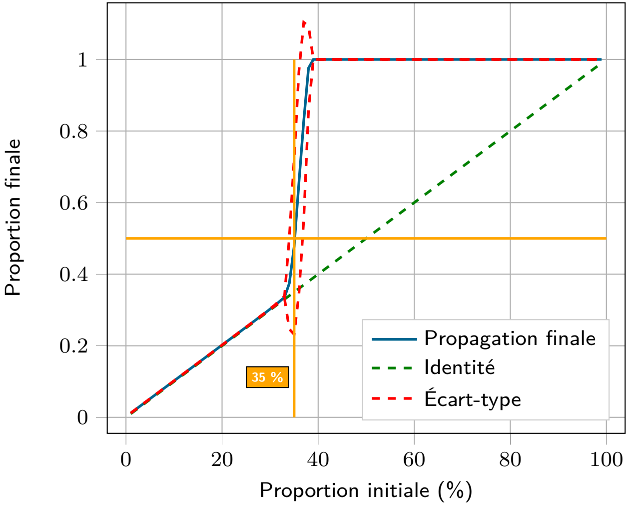
Protocole expérimental

- On fixe $K = 50$;
- $\beta \in \{0, \frac{1}{4}, \frac{1}{2}, 1\}$;
- $q \in \{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$;
- proportion initiale de 1% à 99% ;

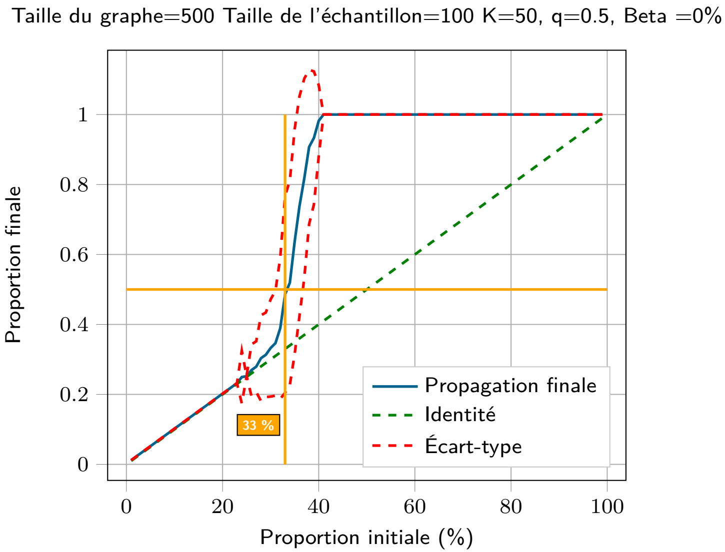
$\rightarrow$ 100 expériences de propagation en choisissant les éléments initiaux au hasard et stocker la propagation à chaque étape de la simulation.
But : pouvoir comparer les résultats des autres expériences, éventuellement fixer certains paramètres qui ont peu d'influence.

Courbe type

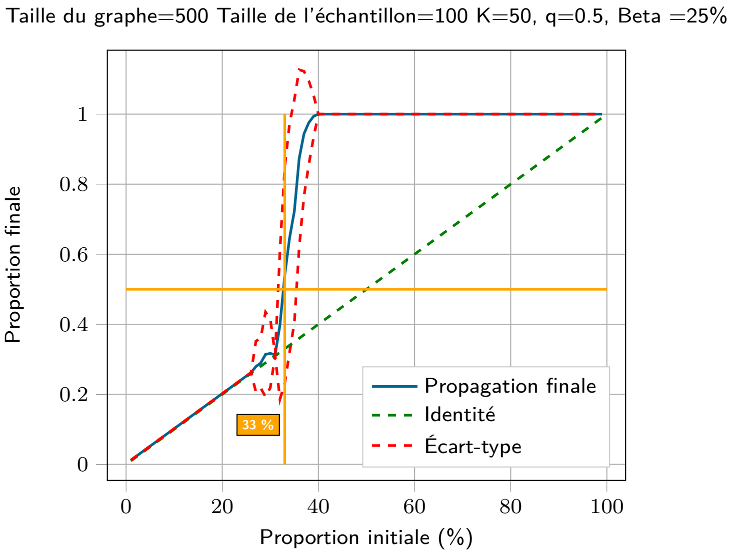
Taille du graphe=500 Taille de l'échantillon=100 K=50, q=0.5, Beta =50%



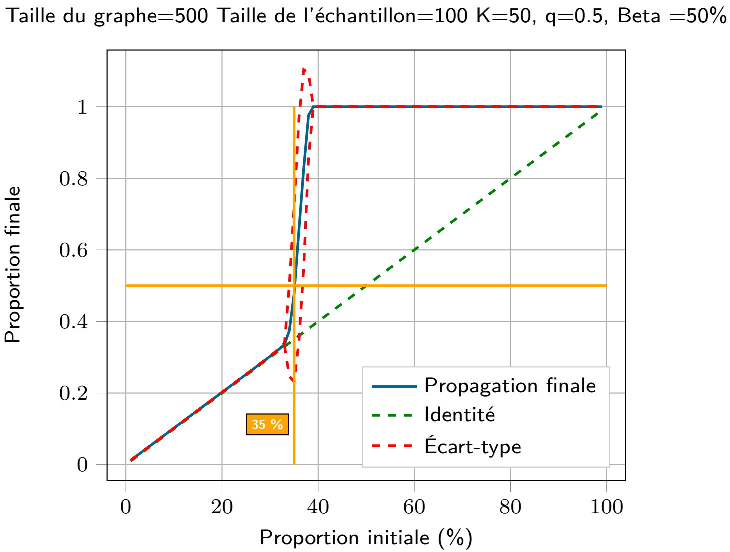
Faible influence du paramètre β



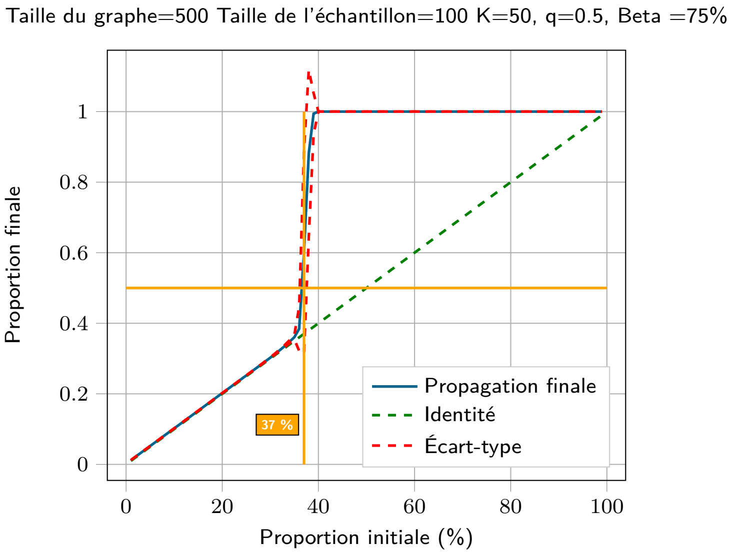
Faible influence du paramètre β



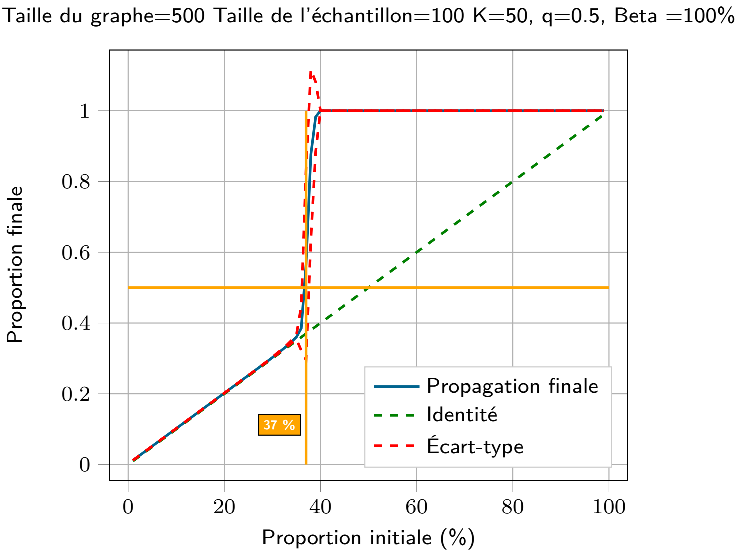
Faible influence du paramètre β



Faible influence du paramètre β

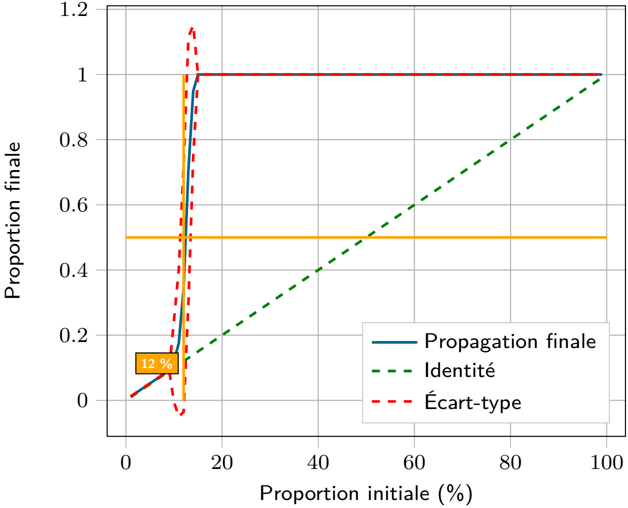


Faible influence du paramètre β



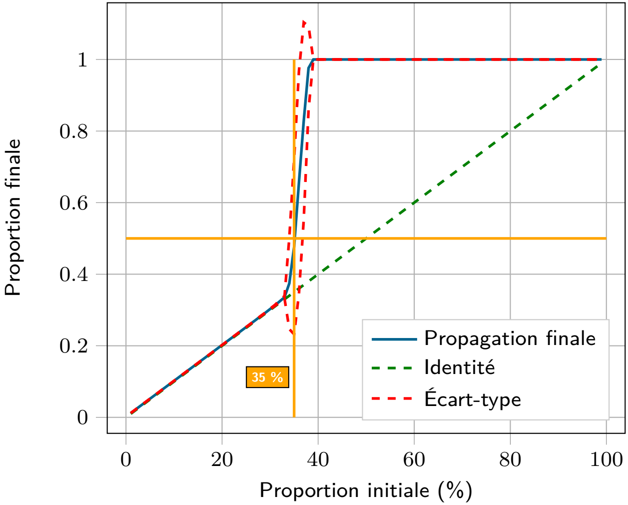
Influence de q

Taille du graphe=500 Taille de l'échantillon=100 K=50, q=0.25, Beta =50%



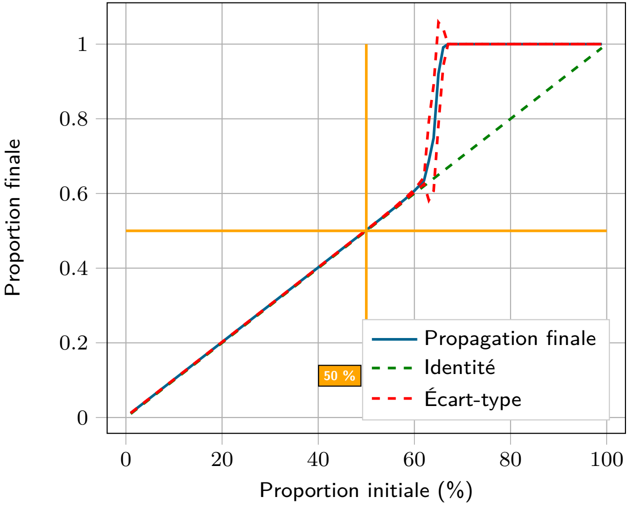
Influence de q

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



Influence de q

Taille du graphe=500 Taille de l'échantillon=100 K=50, q=0.75, Beta =50%



Protocole expérimental

- On fixe $K = 50$;
- $\beta \in \{0, \frac{1}{4}, \frac{1}{2}, 1\}$;
- $q \in \{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$;
- proportion initiale de 1% à 99% ;

$\rightarrow$ 100 expériences de propagation en choisissant les éléments initiaux possédant les plus grands degrés et stocker la propagation à chaque étape de la simulation.

Remarques sur les résultats

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

Remarques sur les résultats

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

Protocole expérimental

- On fixe $K = 50$;
- $\beta \in \{0, \frac{1}{4}, \frac{1}{2}, 1\}$;
- $q \in \{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$;
- proportion initiale de 1% à 99% ;

$\rightarrow$ 100 expériences de propagation en choisissant les éléments initiaux possédant possédant les plus grandes centralités (proportion de plus courts chemins passants par un nœud, algorithme de Ulrik Brandes) et stocker la propagation à chaque étape de la simulation.

Remarques sur les résultats

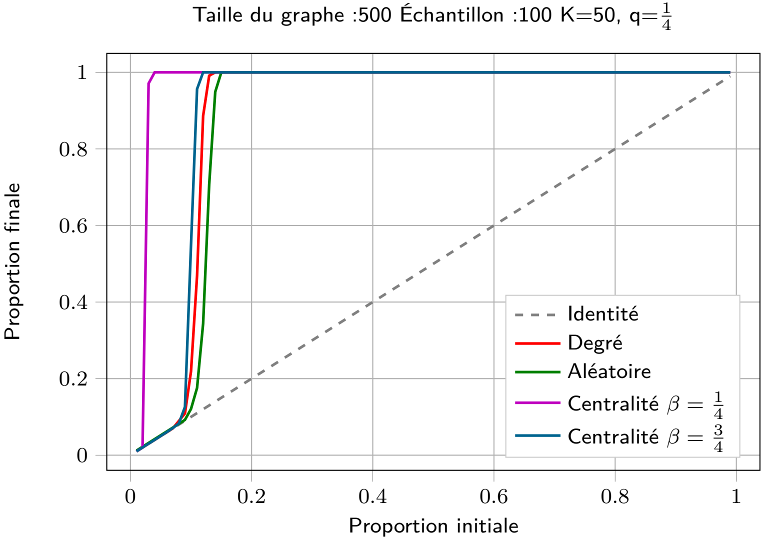
- Le résultat pour $\beta = 0$ est inexploitable ;

Remarques sur les résultats

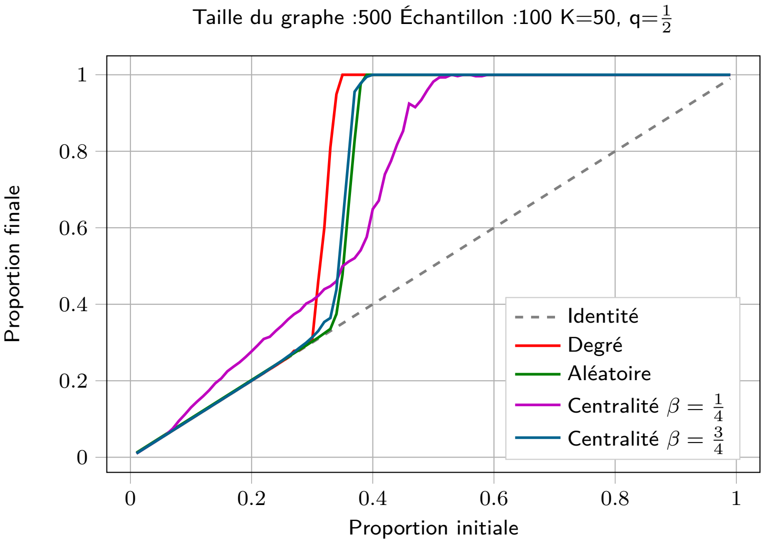
- Le résultat pour $\beta = 0$ est inexploitable ;
- On retrouve les mêmes effets qualitatifs de q.

Plan

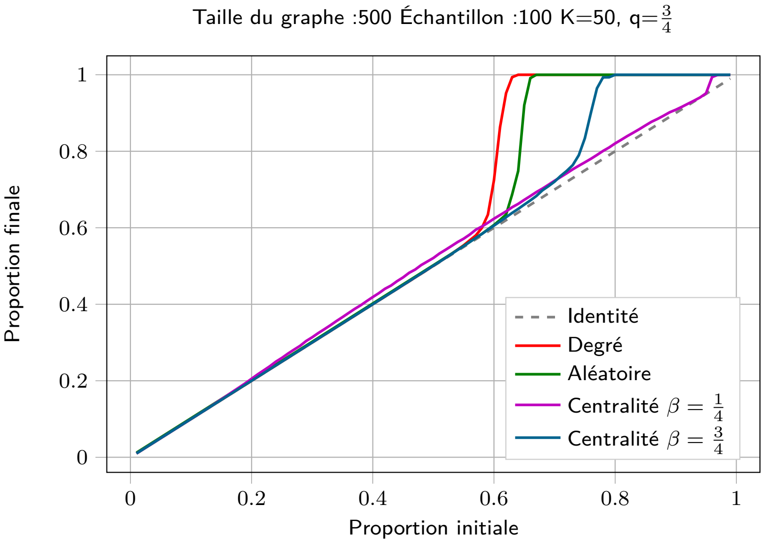
$q = \frac{1}{4}$


Taille du graphe :500 Échantillon :100 K=50, q=$\frac{1}{4}$
Proportion finale
Proportion initiale
0
0.2
0.4
0.6
0.8
1
Identité
Degré
Aléatoire
Centralité $\beta = \frac{1}{4}$
Centralité $\beta = \frac{3}{4}$

$q = \frac{1}{2}$


Taille du graphe :500 Échantillon :100 K=50, q=$\frac{1}{2}$
Proportion finale
Proportion initiale
0
0.2
0.4
0.6
0.8
1
Identité
Degré
Aléatoire
Centralité $\beta = \frac{1}{4}$
Centralité $\beta = \frac{3}{4}$

$q = \frac{3}{4}$


Taille du graphe :500 Échantillon :100 K=50, q=$\frac{3}{4}$
Proportion finale
Proportion initiale
0
0.2
0.4
0.6
0.8
1
Identité
Degré
Aléatoire
Centralité $\beta = \frac{1}{4}$
Centralité $\beta = \frac{3}{4}$

Conséquences

On peut choisir les éléments initialement propagateurs en connaissant q.

Pour $\beta = \frac{1}{4}$,

Conséquences

On peut choisir les éléments initialement propagateurs en connaissant q.

Pour $\beta = \frac{1}{4}$,

- Si $q \geq \frac{1}{2} \rightarrow \approx q$ des éléments de plus haut degré ;

Conséquences

On peut choisir les éléments initialement propagateurs en connaissant q.

Pour $\beta = \frac{1}{4}$,

- Si $q \geq \frac{1}{2} \rightarrow \approx q$ des éléments de plus haut degré ;
- Si $q \leq \frac{1}{4} \rightarrow \approx 5\%$ des éléments de plus haut degré ;

Conséquences

On peut choisir les éléments initialement propagateurs en connaissant q.

Pour $\beta = \frac{1}{4}$,

- Si $q \geq \frac{1}{2} \rightarrow \approx q$ des éléments de plus haut degré ;
- Si $q \leq \frac{1}{4} \rightarrow \approx 5\%$ des éléments de plus haut degré ;
- Une étude plus quantitative serait nécessaire pour $\frac{1}{4} < q < \frac{1}{2}$.

Plan

Conclusion

- On a un premier critère de choix des éléments initiaux

Conclusion

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

Conclusion

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

Conclusion

- 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)

Conclusion

- 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)

Conclusion

- 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 ?

Plan

Démonstration : 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.

Démonstration : 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 $(1-p) > q \leq 0$. Il existe donc des voisins de i vérifiant cette propriété, on a un z-cluster avec $z > q$.

Plan

Algorithme de Watts-Strogatz

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