

Diffusion d'information et topologie des réseaux

Jean-Philippe Cointet, Camille Roth

*CREA — CNRS/Ecole Polytechnique, Paris, France
& Department of Sociology, University of Surrey, UK*



Algotel 29 mai - 1^{er} juin 2007
Ile d'Oléron

Modèles de diffusion d'informations

Point de vue théorique – mais aussi défi empirique

- Réseau social sous-jacent : cadre pour les interactions
- Règles d'interaction et de transmission d'informations

Question : de quelle manière les hypothèses plus ou moins stylisées des modèles traditionnels de topologie de réseaux modifient le phénomène de diffusion ?

1 Contexte

- La diffusion des connaissances
- Influence de la topologie

2 Influence de la topologie

- Modèle simulateur
- Les différentes topologies
- Résultats des simulations

3 Bilan

- Interprétation
- Résumé et perspectives

Vers une prise en compte de la structure

Des approches qualitatives de la diffusion

Coleman et al 1957 (adoption de pratiques médicales), Robertson, 1967 (innovation), Granovetter 1970 (threshold model), Rogers 1976 (adoption de nouveaux produits), Burt 1987 (contagion sociale)

Introduction progressive des effets de structures

Erdős-Rényi 1959

Impulsion formalisatrice

Lloyd & May 2001, Kempe et al. 2003

Vers une prise en compte de la structure

Des approches qualitatives de la diffusion

Coleman et al 1957 (adoption de pratiques médicales), Robertson, 1967 (innovation), Granovetter 1970 (threshold model), Rogers 1976 (adoption de nouveaux produits), Burt 1987 (contagion sociale)

Introduction progressive des effets de structures

Erdős-Rényi 1959

Impulsion formalisatrice

Lloyd & May 2001, Kempe et al. 2003

Vers une prise en compte de la structure

Des approches qualitatives de la diffusion

Coleman et al 1957 (adoption de pratiques médicales), Robertson, 1967 (innovation), Granovetter 1970 (threshold model), Rogers 1976 (adoption de nouveaux produits), Burt 1987 (contagion sociale)

Introduction progressive des effets de structures

Erdős-Rényi 1959

Impulsion formalisatrice

Lloyd & May 2001, Kempe et al. 2003

Résultats empiriques récents

En parallèle, mises à jour de propriétés topologiques non triviales

Watts & Strogatz 1998, Barabasi & Albert 1999

Le réseau “scale-free” nouvelle archétype

à la suite de Pastor-Satorras Vespignani 2001, par ex. Ganesh 2005

Résultats empiriques récents

En parallèle, mises à jour de propriétés topologiques non triviales

Watts & Strogatz 1998, Barabasi & Albert 1999

Le réseau “scale-free” nouvelle archétype

à la suite de Pastor-Satorras Vespignani 2001, par ex. Ganesh 2005

Validation empirique des modèles de diffusion

Mais... quid de la confrontation de ces formalismes avec la réalité ?

un voeu pieu... ? Rogers 1976 :

"For network analysis to fulfill its potential, however, I feel we must improve the methods of data gathering and measurement (...). Longitudinal panel designs for networks analysis of diffusion process are also needed ; along with field experiments, they help secure the necessary data to illuminate the over-time process of diffusion."

Validation empirique des modèles de diffusion

Mais... quid de la confrontation de ces formalismes avec la réalité ?

un voeu pieu... ? Rogers 1976 :

“For network analysis to fulfill its potential, however, I feel we must improve the methods of data gathering and measurement (...). Longitudinal panel designs for networks analysis of diffusion process are also needed ; along with field experiments, they help secure the necessary data to illuminate the over-time process of diffusion.”

Le processus de transmission

- N agents, une seule information.
 $c(t) \in \{0, 1\}^N$, c est croissante.

Modèle SI (Hethcote 2000))

- Proportion initiale λ d'agents "informés", choisis au hasard.
- *Interaction* : un agent cible i est choisi, et choisit un de ses voisins j .
- *Transmission* : si le voisin est informé, la cible adopte l'information. ($c_j(t) = 1 \Rightarrow c_i(t) = 1$)

Le processus de transmission

- N agents, une seule information.
 $c(t) \in \{0, 1\}^N$, c est croissante.

Modèle SI (Hethcote 2000))

- Proportion initiale λ d'agents "informés", choisis au hasard.
- *Interaction* : un agent cible i est choisi, et choisit un de ses voisins j .
- *Transmission* : si le voisin est informé, la cible adopte l'information. ($c_j(t) = 1 \Rightarrow c_i(t) = 1$)

Observables

Proportion d'agents informés

$$\rho(t) = \frac{1}{N} \sum_{i=1}^N c_i(t)$$

(donc $\rho(0) = \lambda$)

Autres observables ?

- *Motifs* : délai d'information en fonction de la distance, densité des informés...
- *Conditions initiales* : choix particulier des informés initiaux (e.g. "super-spreaders")

Observables

Proportion d'agents informés

$$\rho(t) = \frac{1}{N} \sum_{i=1}^N c_i(t)$$

(donc $\rho(0) = \lambda$)

Autres observables ?

- *Motifs* : délai d'information en fonction de la distance, densité des informés...
- *Conditions initiales* : choix particulier des informés initiaux (e.g. "super-spreaders")

Les réseaux réels et leurs versions appauvries

“Réseau réel” (RN)

- 1 Réseau de collaboration scientifique (Medline, zebrafish, 2000–2004)
- 2 Réseau d'interlock dans les CA (Theyrule, 2005)

Les modèles

- “Scale-Free” (SF) : conservation du degré de chaque noeud
- “Erdos-Renyi” (ER) : même nombre de liens et d'agents
- “Complete network” (CN) : même nombre d'agents
- “Event-Based” (EB) : graphe à structure événementielle sous-jacente

Les réseaux réels et leurs versions appauvries

“Réseau réel” (RN)

- 1 Réseau de collaboration scientifique (Medline, zebrafish, 2000–2004)
- 2 Réseau d'interlock dans les CA (Theyrule, 2005)

Les modèles

- “Scale-Free” (SF) : conservation du degré de chaque noeud
- “Erdos-Renyi” (ER) : même nombre de liens et d'agents
- “Complete network” (CN) : même nombre d'agents
- “Event-Based” (EB) : graphe à structure événementielle sous-jacente

Les réseaux réels et leurs versions appauvries

“Réseau réel” (RN)

- 1 Réseau de collaboration scientifique (Medline, zebrafish, 2000–2004)
- 2 Réseau d'interlock dans les CA (Theyrule, 2005)

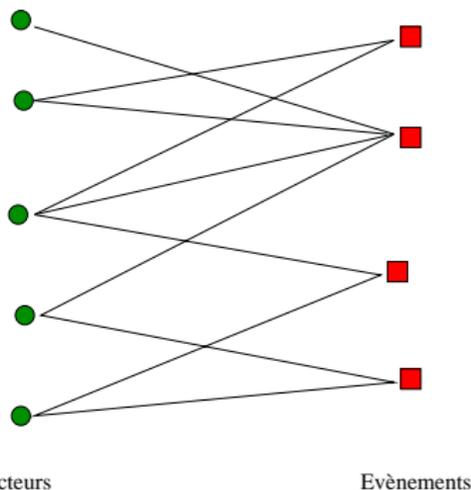
Les modèles

- “Scale-Free” (SF) : conservation du degré de chaque noeud
- “Erdos-Renyi” (ER) : même nombre de liens et d'agents
- “Complete network” (CN) : même nombre d'agents
- “Event-Based” (EB) : graphe à structure événementielle sous-jacente

Reconstruction bipartite

Processus

- réseau bipartite original
- dont on extrait les distributions de degrés respectives des acteurs et des événements
- afin de simuler des graphes bi-partites avec ces mêmes distributions



Reconstruction bipartite

Processus

- réseau bipartite original
- dont on extrait les distributions de degrés respectives des acteurs et des événements
- afin de simuler des graphes bi-partites avec ces mêmes distributions



Acteurs

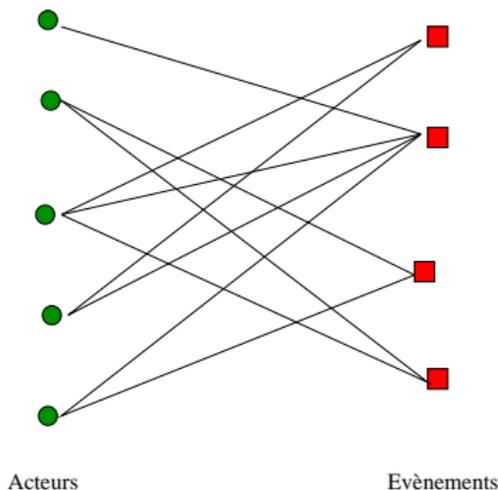


Evènements

Reconstruction bipartite

Processus

- réseau bipartite original
- dont on extrait les distributions de degrés respectives des acteurs et des événements
- afin de simuler des graphes bi-partites avec ces mêmes distributions



paramètres topologiques

pour chaque réseau, on évalue :

- la distribution de degrés,
- le coefficient de clustering (local) : c_3 ,
- le *coefficient de diamants* : “ la proportion moyenne de voisins communs dans le voisinage d'un noeud ” :

$$c_4(i) = \frac{\sum_{i_1=1}^{k_i} \sum_{i_2=i_1+1}^{k_i} \kappa_{i_1, i_2}}{\sum_{i_1=1}^{k_i} \sum_{i_2=i_1+1}^{k_i} [(k_{i_1} - \kappa_{i_1, i_2})(k_{i_2} - \kappa_{i_1, i_2}) + \kappa_{i_1, i_2}]}$$

où κ_{i_1, i_2} désigne le nombre de noeuds que les j_1^{eme} & j_2^{eme} voisins de i ont en commun.

paramètres topologiques

pour chaque réseau, on évalue :

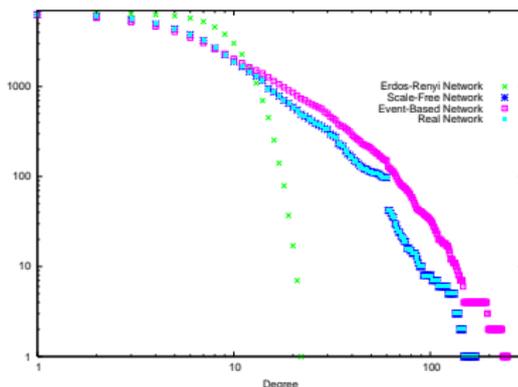
- la distribution de degrés,
- le coefficient de clustering (local) : c_3 ,
- le *coefficient de diamants* : “ la proportion moyenne de voisins communs dans le voisinage d'un noeud ” :

$$c_4(i) = \frac{\sum_{i_1=1}^{k_i} \sum_{i_2=i_1+1}^{k_i} \kappa_{i_1, i_2}}{\sum_{i_1=1}^{k_i} \sum_{i_2=i_1+1}^{k_i} [(k_{i_1} - \kappa_{i_1, i_2})(k_{i_2} - \kappa_{i_1, i_2}) + \kappa_{i_1, i_2}]}$$

où κ_{j_1, j_2} désigne le nombre de noeuds que les j_1^{eme} & j_2^{eme} voisins de i ont en commun.

Résumé des paramètres topologiques

	RN1	SF1	ER1	CN1	EB1
agents N	6453				
links M	$6.74 \cdot 10^4$			$2.08 \cdot 10^7$	$7.62 \cdot 10^4$
density d	.00162			1	.00183
degree dist.	power-law tail		Poisson	—	power-law tail
$\langle c_3 \rangle$.827	.00539	.00199	1	.753
$\langle c_4 \rangle$.284	.000444	.000261	1	.0443

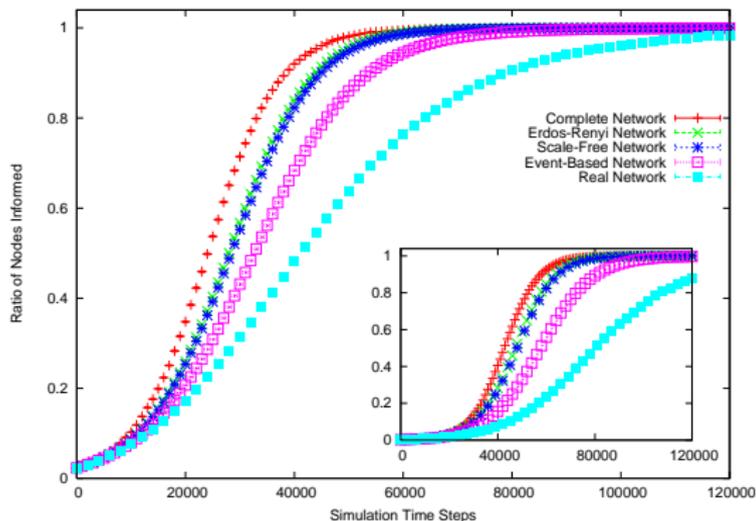


(valeurs
moyennées sur
1000 réseaux)

Resultats

Vitesse de convergence

- Plus les réseaux sont "réalistes" et plus lente est la convergence,
- ER et SF ont des dynamiques semblables,
- EB intermédiaire entre RN et ER/SF

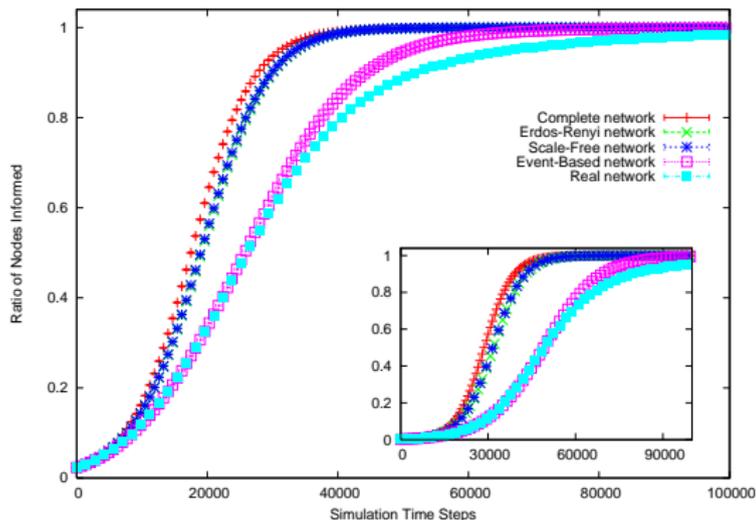


Medline

Resultats

Vitesse de convergence

- Plus les réseaux sont "réalistes" et plus lente est la convergence,
- ER et SF ont des dynamiques semblables,
- EB intermédiaire entre RN et ER/SF



They rule

Qualitativement...

Confirmation de résultats qualitatifs

La transitivité ralentit la propagation de l'information :

"if one tells a rumor to all his close friends, and they do likewise, many will hear the rumor a second and third time, since those linked by strong ties tend to share friends"

(Granovetter 1973)

i.e., les voisinages recouvrants ralentissent la diffusion

Résumé et quelques questions

topologie et diffusion

Rôle secondaire de la distribution de degrés, mais rôle prépondérant de la densité locale - structure de communautés.

modèles réalistes de morphogenèse

EB = meilleur modèle de reconstruction mais insuffisant pour *Medline...*

autres classes de réseaux

Réseaux dont la structure sous-jacente n'est pas bi-partite ?

Résumé et quelques questions

topologie et diffusion

Rôle secondaire de la distribution de degrés, mais rôle prépondérant de la densité locale - structure de communautés.

modèles réalistes de morphogenèse

EB = meilleur modèle de reconstruction mais insuffisant pour *Medline...*

autres classes de réseaux

Réseaux dont la structure sous-jacente n'est pas bi-partite ?

Résumé et quelques questions

topologie et diffusion

Rôle secondaire de la distribution de degrés, mais rôle prépondérant de la densité locale - structure de communautés.

modèles réalistes de morphogenèse

EB = meilleur modèle de reconstruction mais insuffisant pour *Medline...*

autres classes de réseaux

Réseaux dont la structure sous-jacente n'est pas bi-partite ?