

Processus de contact sur graphe et méthodes variationnelles de simplification de la structure spatiale

Nathalie Peyrard, dept. MIA - INRA - Avignon/Toulouse

Alain Franc, dept EFPA - INRA - Bordeaux

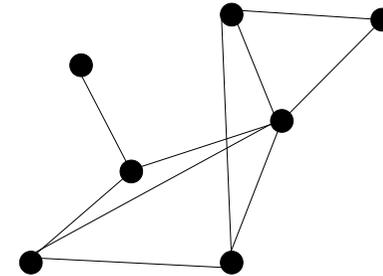
Ulf Dieckmann, projet ADN - IIASA - Vienne

Motivation

- Quelles sont les caractéristiques du réseau qui influencent l'évolution du processus? Comment?
- Rôle sur la transition de phase?
- Comment améliorer l'estimation de la valeur critique?

Structure des graphes

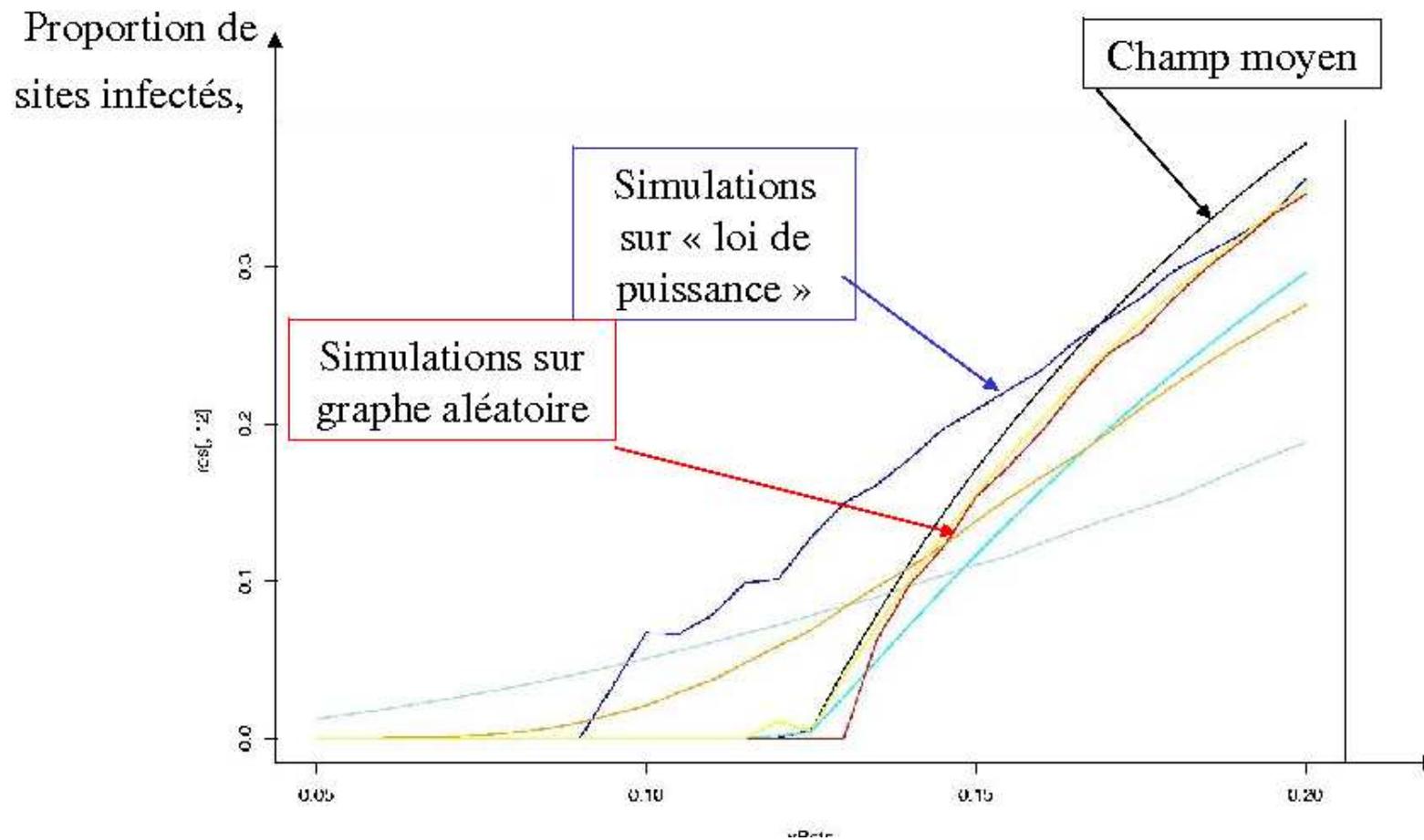
- Principales caractéristiques des graphes :
 - distribution des degrés, P
 - coefficient d'agrégation, C
 - diamètre, d
 - corrélation des degrés



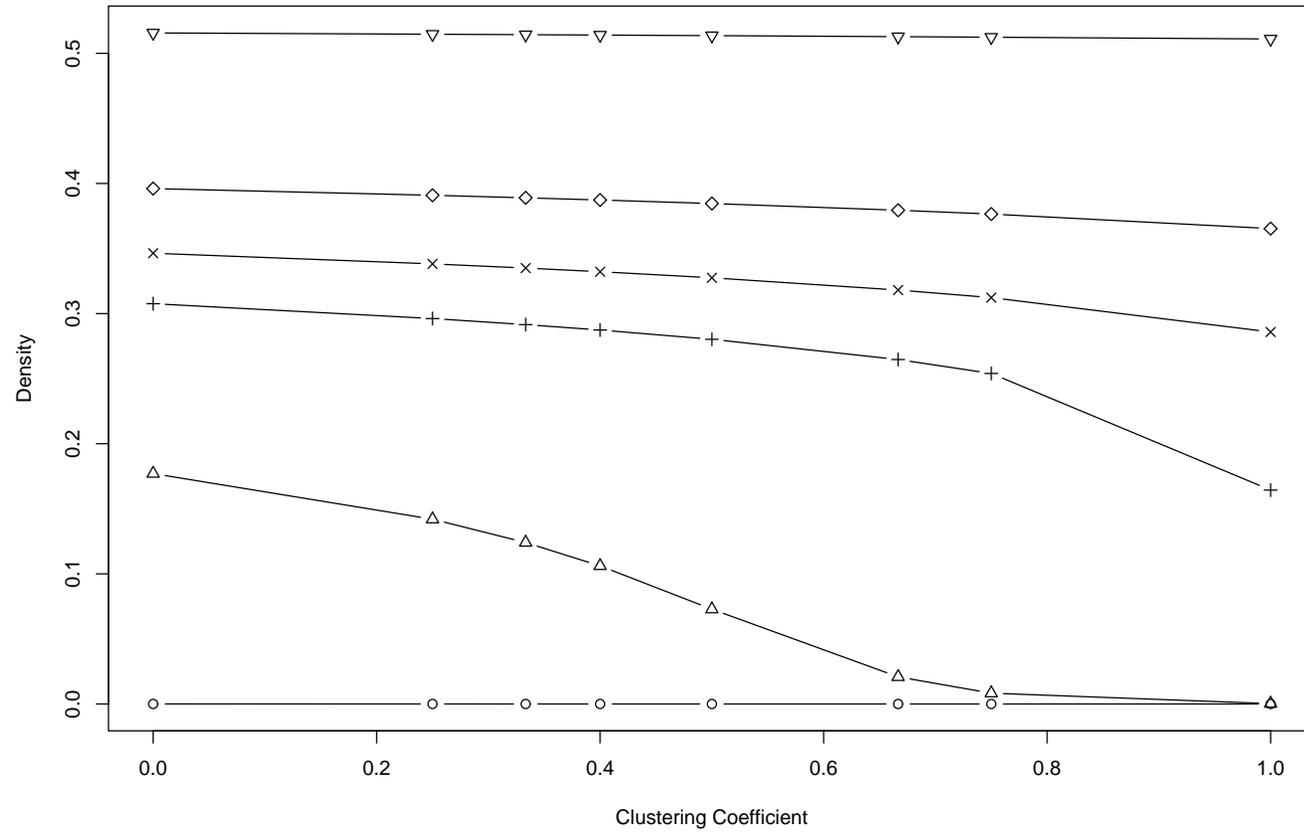
	Random Graph (Poisson)	Scale-free (Barabasi)	Graphes Réels
Petit d	×	×	×
Grand C	NON	NON	×
P , loi puissance	NON	×	×

Newman, M. E. J., 2003. *The structure and function of complex networks*. SIAM Review 45, 167-256

Effet de la distribution des degrés



Effet du coefficient d'agrégation



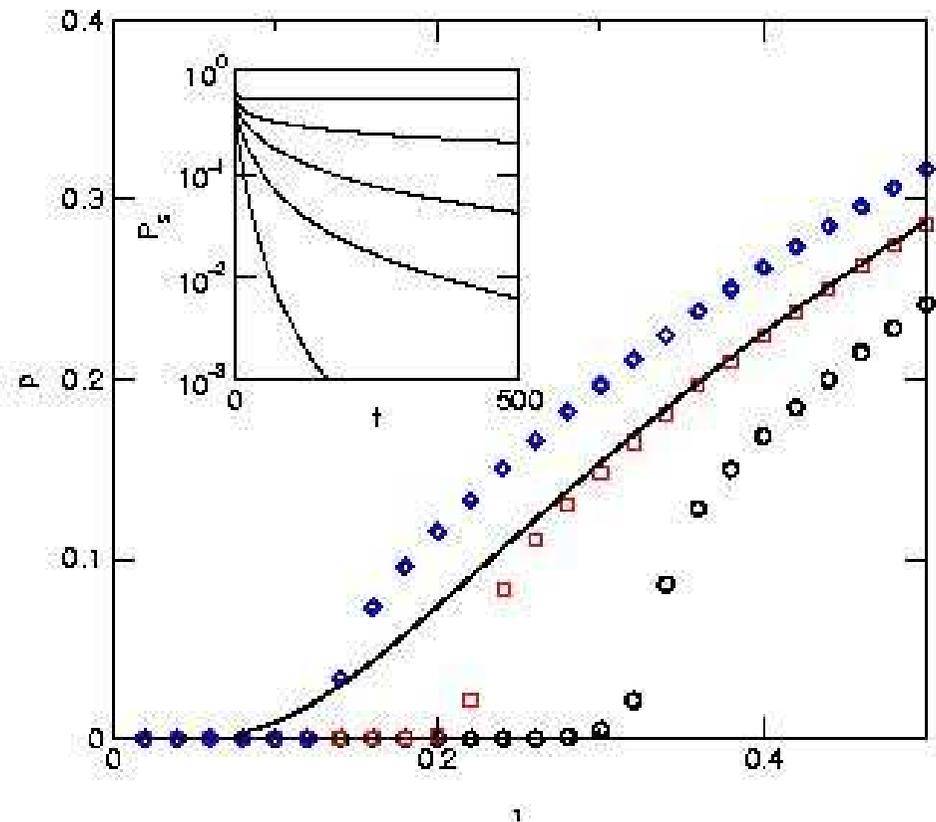
Effet de la corrélation des degrés

Deux graphes en loi de puissance

- un avec degrés non corrélés

- un avec hubs connectés préférentiellement aux sommets de faible degré

→ le second est un frein à la propagation



Eguiluz V. and Klemm K., 2002. Epidemic threshold in structured scaled free networks, *Phys. Rev. Let.* (89)

Le modèle de processus de contact

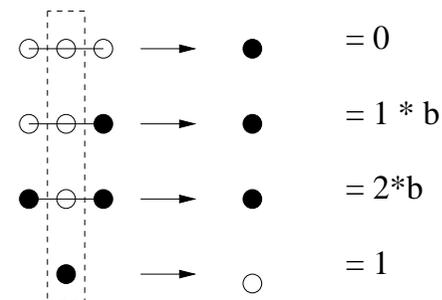
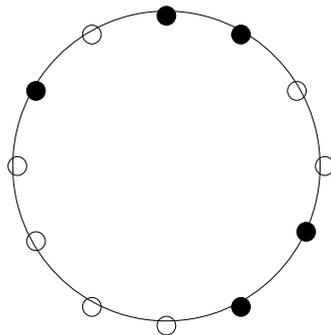
- stochastique,
- spatio-temporel
- 2 états (sain/malade),
- b : taux de contamination par site,
- transition de phase

$a_i = \text{nb de voisins de } i \text{ infectés}$

$$P(z_i^{t+dt} = 0 | z_i^t = 1) = dt$$

$$P(z_i^{t+dt} = 1 | z_i^t = 0, a_i = a) = a b dt$$

$$\rho(t) = P(z_i^t = 1)$$



Harris T.E., 1974. Contact interactions on a lattice, *Annals of Probability*, 2, 969-988

Equations des dynamiques d'ordre 1 et 2

- Pour un graphe à degré constant d :

$$\frac{d\rho}{dt} = (1 - \rho)t_{0 \rightarrow 1} - \rho$$

$$\frac{dP^{(1)}(01)}{dt} = (\rho - P^{(1)}(01)) + (1 - \rho - P^{(1)}(01))t_{00 \rightarrow 01}^{(1)} - P^{(1)}(01)(1 + t_{01 \rightarrow 11}^{(1)})$$

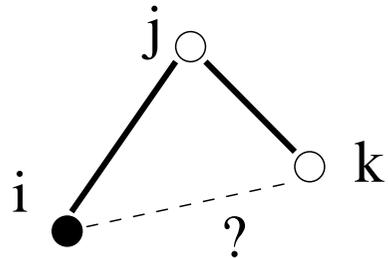
$$\frac{d\rho}{dt} = dbP^{(1)}(01) - \rho$$

$$\frac{dP^{(1)}(01)}{dt} = (\rho - P^{(1)}(01)) + b(d - 1)P^{(1,1)}(100)$$

$$-P^{(1)}(01)\left[1 + b + b(d - 1)\frac{P^{(1,1)}(100)}{1 - \rho - P^{(1)}(01)}\right]$$

Simplification de la complexité spatiale

- Principe : méthode d'approximation par **troncature de la hiérarchie** à un niveau choisi
- Fermeture à l'ordre 2, compromis entre précision et complexité
⇒ on approche les probabilités sur les triplets par une fonction des probas sur les singletons et sur les paires
- Plusieurs fermetures possibles dans la littérature



$$P^{(1,1)}(z_i, z_j, z_k) \approx C\tilde{P}^\Delta(z_i, z_j, z_k) + (1 - C)\tilde{P}^\wedge(z_i, z_j, z_k),$$

C : coefficient d'agrégation du réseau d'interaction

Dieckmann U., Law R. and Metz J.A. editors, 2000. The geometry of ecological interactions - Simplifying spatial complexity. Cambridge Studies in Adaptive Dynamics. Cambridge University Press, UK.

Objectif : Prise en compte des corrélations à longue distance dans la fermeture

- les solutions existantes de fermeture des moments ne prennent en compte que les corrélations à distance 1
- mauvaise approximation près du point critique (décroissance des corrélations en loi puissance au lieu d'exponentielle)

⇒ introduction des corrélations à plus grande distance

Méthodes variationnelles

- Origine = mécanique statistique, étude des systèmes de particules en interaction, minimum d'entropie
- Espace discret (grille, graphe)
- Utilisé en analyse d'image (estimation), écologie (recherche d'équilibre), modèles graphiques (inférence)

Méthodes variationnelles à l'ordre 1 et 2

- Méthode la plus naïve, à l'ordre 1 : approximation **champ moyen**
 - on approche l'influence des voisins par leur influence moyenne → indépendance des sites
- Approximation **par paire** (Sato, Iwasa, 2000)
 - même approximation pour un triangle fermé et ouvert

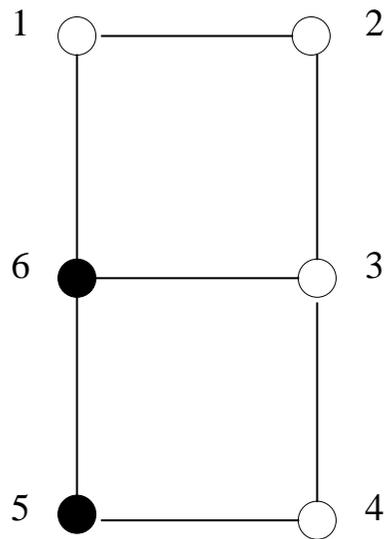
$$\tilde{P}^{\Delta}(z_i, z_j, z_k) = \frac{P^{(1)}(z_i, z_j)P^{(1)}(z_j, z_k)}{P(z_j)}$$

- Approximation de **Bethe** (Morita, 1994)
 - égal à l'approx. par paire lorsque le triangle est ouvert

$$\tilde{P}^{\Delta}(z_i, z_j, z_k) = \frac{P^{(1)}(z_i, z_j)P^{(1)}(z_j, z_k)P^{(1)}(z_i, z_k)}{P(z_i)P(z_j)P(z_k)}$$

Généralisation à un ordre quelconque

- Approximation de **Kikuchi** :



Exemple à l'ordre 4:

$$P(z_1, \dots, z_6) = \frac{P(z_1, z_2, z_3, z_6)P(z_6, z_3, z_4, z_5)}{P(z_6, z_3)}$$

- Application : modèles graphiques

Yedidia J., Freeman, W. and Weiss Y., 2001, Bethe free energy, Kikuchi approximations and belief propagation algorithms, MERL technical report.

Méthodes de fermeture des moments

- Origine : physique, transitions de phase
- Espace continu
- **Power 1** (Bolker and Pacala, 1997, Dieckman and Law, 2000)

$$\begin{aligned}\tilde{P}^{\Delta}(z_i, z_j, z_k) &= P(z_i)P^{(1)}(z_j, z_k) + P(z_j)P^{(1)}(z_i, z_k) \\ &\quad + P(z_k)P^{(1)}(z_i, z_j) - 2P(z_i)P(z_j)P(z_k)\end{aligned}$$

- **Power 2** (Dieckmann and Law, 2000)

$$\begin{aligned}\tilde{P}^{\Delta}(z_i, z_j, z_k) &= \frac{1}{2} \left(\frac{P^{(1)}(z_i, z_j)P^{(1)}(z_j, z_k)}{P(z_j)} + \frac{P^{(1)}(z_i, z_j)P^{(1)}(z_i, z_k)}{P(z_i)} \right. \\ &\quad \left. + \frac{P^{(1)}(z_i, z_k)P^{(1)}(z_j, z_k)}{P(z_k)} - P(z_i)P(z_j)P(z_k) \right)\end{aligned}$$

Extension aux corrélations à distance 2 : exemple sur Bethe

- Bethe “classique”, l'équation exacte non fermée

$$\begin{aligned}\frac{d\rho}{dt} &= dbP^{(1)}(01) - \rho \\ \frac{dP^{(1)}(01)}{dt} &= (\rho - P^{(1)}(01)) + b(d-1)P^{(1,1)}(100) \\ &\quad - P^{(1)}(01)\left[1 + b + b(d-1)\frac{P^{(1,1)}(100)}{1 - \rho - P^{(1)}(01)}\right]\end{aligned}$$

est fermée en approchant $P^{(1,1)}(100)$ par

$$P^{(1,1)}(100) \approx C \frac{P^{(1)}(01)^2 P^{(1)}(00)}{\rho(1-\rho)^2} + (1-C) \frac{P^{(1)}(01)P^{(1)}(00)}{1-\rho}$$

Extension aux corrélations à distance 2 : exemple sur Bethe

- Approximation Bethe quelles que soient les distances (géodésiques) entre les sommets :

$$\tilde{P}^{(d_{ij}, d_{jk}, d_{ik})}(z_i, z_j, z_k) = \frac{P^{(d_{ij})}(z_i, z_j)P^{(d_{jk})}(z_j, z_k)P^{(d_{ik})}(z_i, z_k)}{P(z_i)P(z_j)P(z_k)}$$

- si d_{ik} est > 2 , on remplace $P^{(d_{ik})}(z_i, z_k)$ par $P(z_i)P(z_k)$

Extension aux corrélations à distance 2 : exemple sur Bethe

- Rappel, on cherche une approximation de la forme

$$\tilde{P}^{(1,1)}(z_i, z_j, z_k) = C\tilde{P}^\Delta(z_i, z_j, z_k) + (1 - C)\tilde{P}^\wedge(z_i, z_j, z_k)$$

⇒ L'approximation $\tilde{P}^\Delta(z_i, z_j, z_k)$ ne change pas

⇒ l'approximation $\tilde{P}^\wedge(z_i, z_j, z_k)$ devient

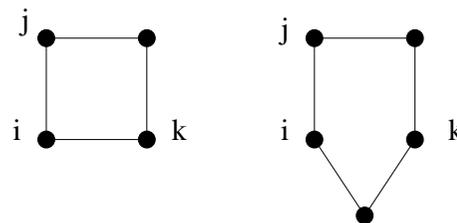
$$\tilde{P}^{(1,1,2)}(z_i, z_j, z_k) = \frac{P^{(1)}(z_i, z_j)P^{(1)}(z_j, z_k)P^{(2)}(z_i, z_k)}{P(z_j)P(z_i)P(z_k)}$$

Extension aux corrélations à distance 2 : exemple sur Bethe

$$\begin{aligned}\frac{d\rho}{dt} &= dbP^{(1)}(01) - \rho \\ \frac{dP^{(1)}(01)}{dt} &= (\rho - P^{(1)}(01)) + b(d-1)\tilde{P}^{(1,1)}(100) \\ &\quad - P^{(1)}(01)\left[1 + b + b(d-1)\frac{\tilde{P}^{(1,1)}(100)}{1 - \rho - P^{(1)}(01)}\right] \\ \frac{dP^{(2)}(01)}{dt} &= -P^{(2)}(01)(t_{01\rightarrow 11}^{(2)} + t_{01\rightarrow 00}^{(2)}) \\ &\quad + P^{(2)}(00)t_{00\rightarrow 01}^{(2)} + P^{(2)}(11)t_{11\rightarrow 01}^{(2)}\end{aligned}$$

Extension aux corrélations à distance 2 : exemple sur Bethe

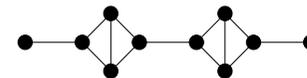
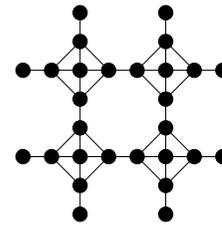
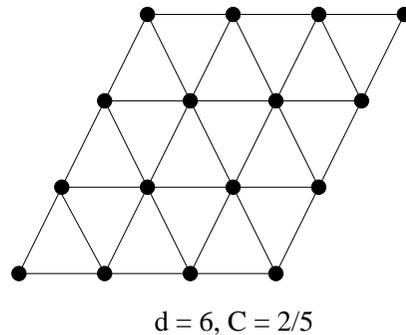
$$\tilde{t}_{01 \rightarrow 11}^{(2)} = b(d-1) \frac{\tilde{P}^{(1,2)}(101)}{P^{(2)}(01)} + b \frac{\tilde{P}^{\wedge}(011)}{P^{(2)}(01)}$$



$$\begin{aligned} \tilde{P}^{(1,2)}(z_i, z_j, z_k) = & \alpha_1 \frac{P^{(1)}(z_i, z_j) P^{(2)}(z_j, z_k) P^{(1)}(z_i, z_k)}{P(z_i) P(z_j) P(z_k)} \\ & + \alpha_2 \frac{P^{(1)}(z_i, z_j) P^{(2)}(z_j, z_k) P^{(2)}(z_i, z_k)}{P(z_i) P(z_j) P(z_k)} + \alpha_3 \frac{P^{(1)}(z_i, z_j) P^{(2)}(z_j, z_k)}{P(z_j)} \end{aligned}$$

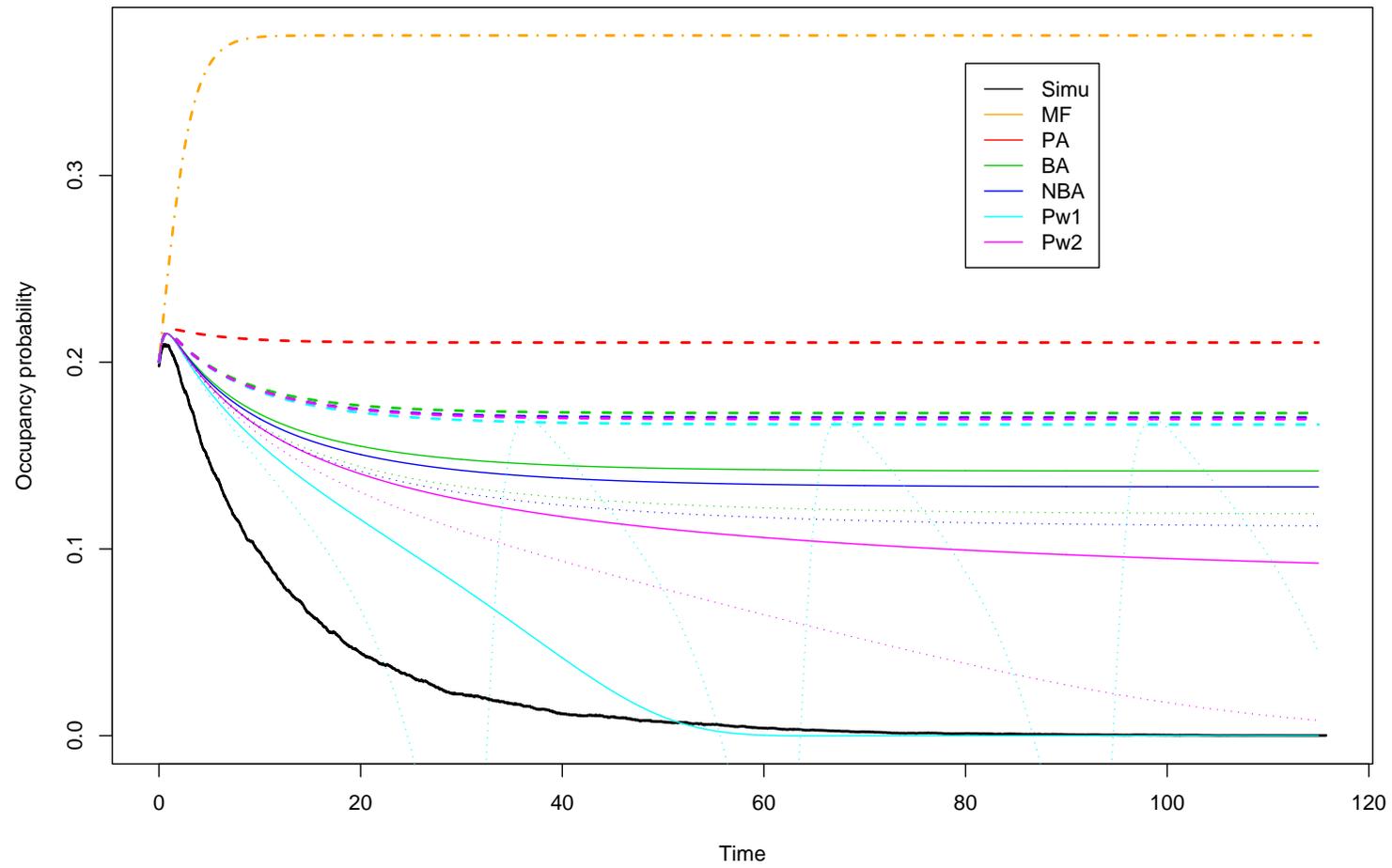
Résultats

- Comparaison de
 - simulations du processus de contact
 - approximation en champ moyen (indépendance spatiale)
 - Pair approximation (PA), Bethe (BA), Bethe normalisé (NBA), Power 1 (Pw1) et Power 2 (Pw2) avec corrélation à distance 1, 2 et 3
- Sur les graphes réguliers suivants



Près du point critique ...

$$d = 4$$



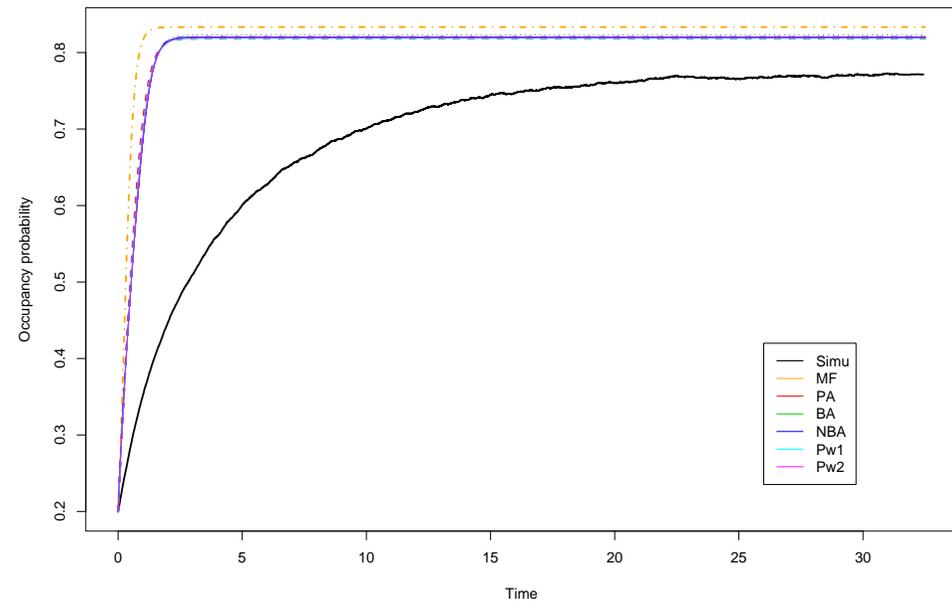
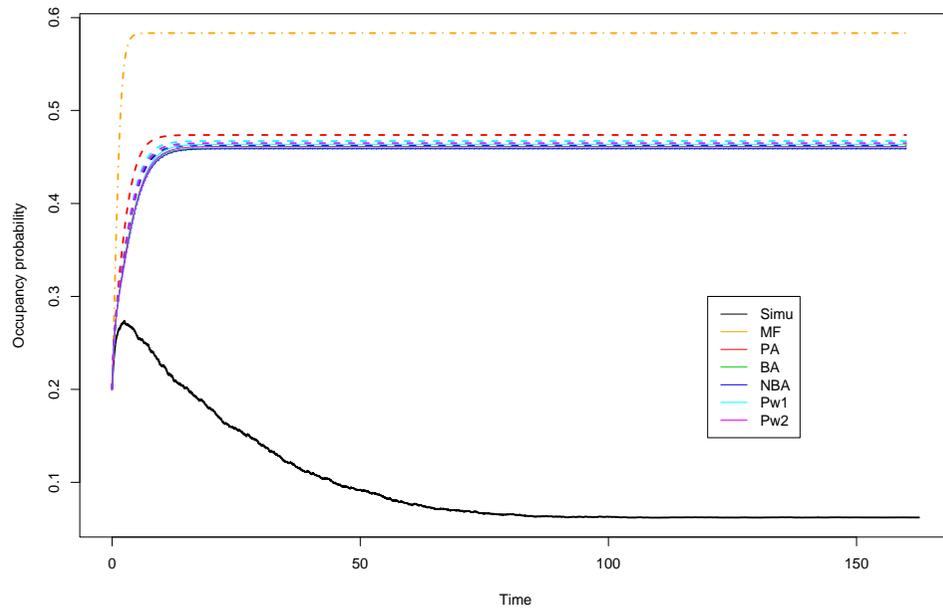
Près du point critique ...

- Conclusions

- Performances relatives communes aux trois graphes
- Toutes les approximations surestiment la proba d'être infecté
- Champ moyen : le plus mauvais ...comme attendu
- La prise en compte successive des corrélations à distance 1, puis 2, puis 3 améliore beaucoup l'approximation juste en dessous du point critique
- Pw1 ou Pw2 paraissent les approximations les plus performantes

... et en s'éloignant du point critique

$$d = 3$$



... et en s'éloignant du point critique

- Conclusions

- L'apport de la prise en compte des corrélations est de plus en plus faible
- Proche du point critique mais au dessus, gros écart entre simulations et approximations
- Cet écart tends vers 0 quand le paramètre de contamination augmente

Approximation par paire et de Bethe pour un graphe quelconque

- Qu'est-ce qui change?

- dépendance entre états, entre degrés et entre états et degrés

- quantités d'intéret : ρ_k et $P^{kl}(z_i, z_j) \equiv P(z_i, z_j | \text{deg}(i) = k, \text{deg}(j) = l), \forall k, j.$

- Multiplication de la taille du système.

- Equations exactes

$$\frac{d\rho_k}{dt} = (1 - \rho_k)t_{0 \rightarrow 1}^k - \rho_k$$

$$\frac{dP^{kl}(00)}{dt} = P^{kl}(01) + P^{kl}(10) - P^{kl}(00)(t_{00 \rightarrow 01}^{kl} + t_{00 \rightarrow 01}^{lk})$$

Approximation par paire

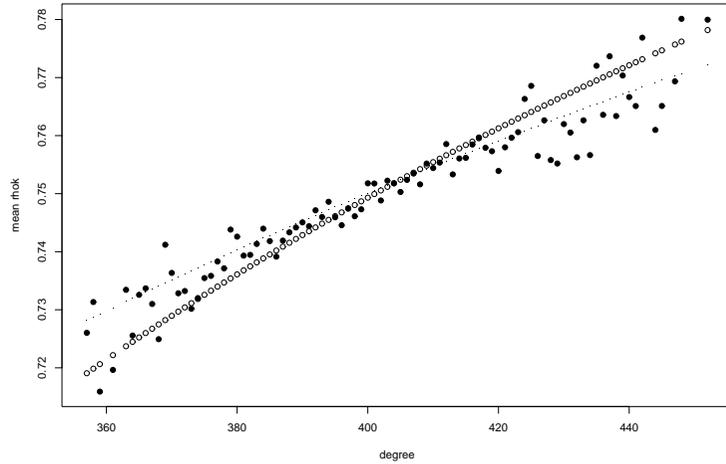
- Approximation des transitions

$$t_{0 \rightarrow 1}^k = b \frac{k P^{kl}(01)}{1 - \rho_k}$$
$$t_{00 \rightarrow 01}^{lk} \approx b \frac{(k-1) P^{kl}(01)}{1 - \rho_k}$$

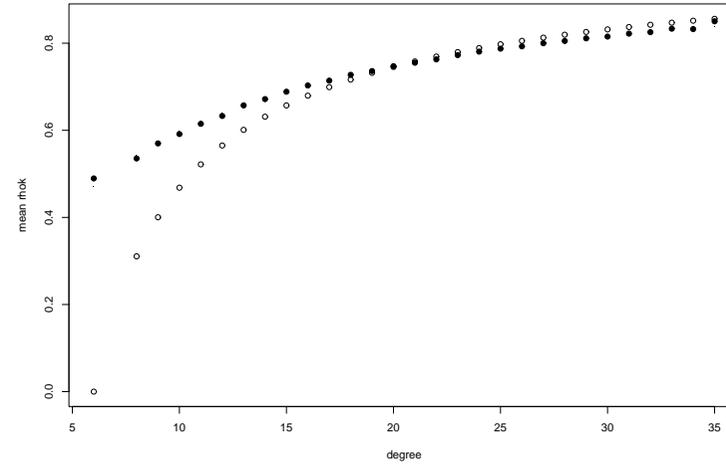
- Solution explicite à l'équilibre

$$\alpha_k = kb - \frac{k}{k-1}$$
$$\rho_k = 1 - \frac{1}{1 + \alpha_k} \forall k > 1$$

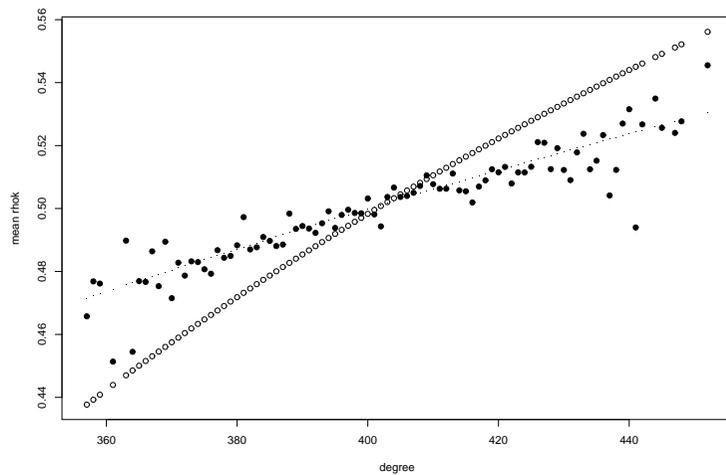
Résultats sur graphe aléatoire



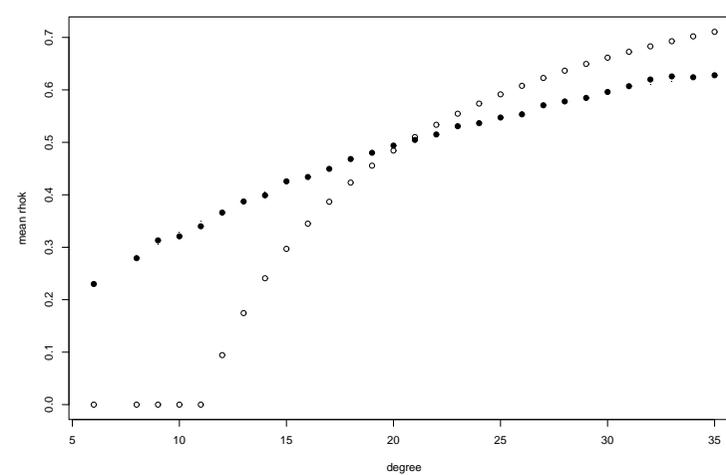
$$E[d] = 400, b.E[d] = 4$$



$$E[d] = 20, b.E[d] = 4$$

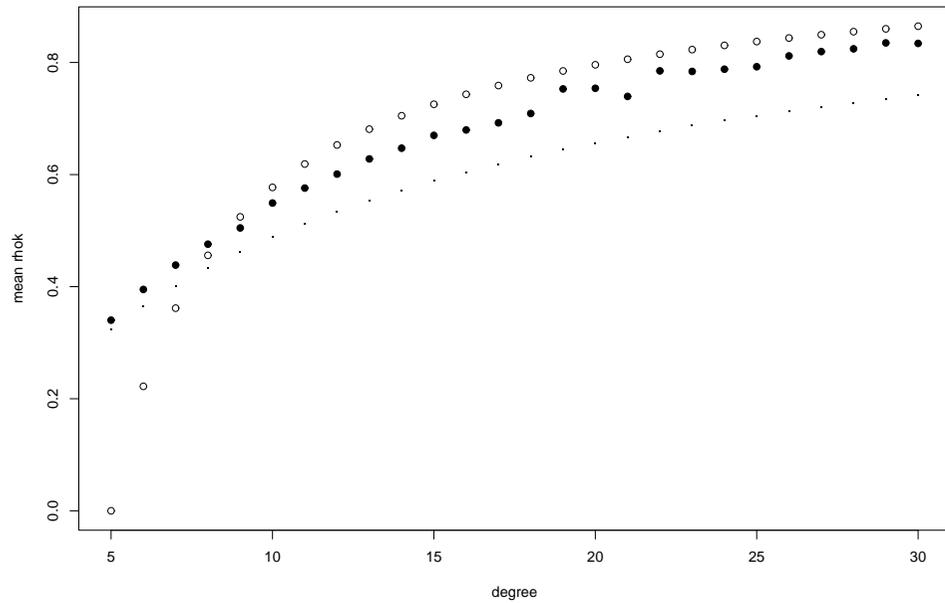


$$E[d] = 400, b.E[d] = 2$$

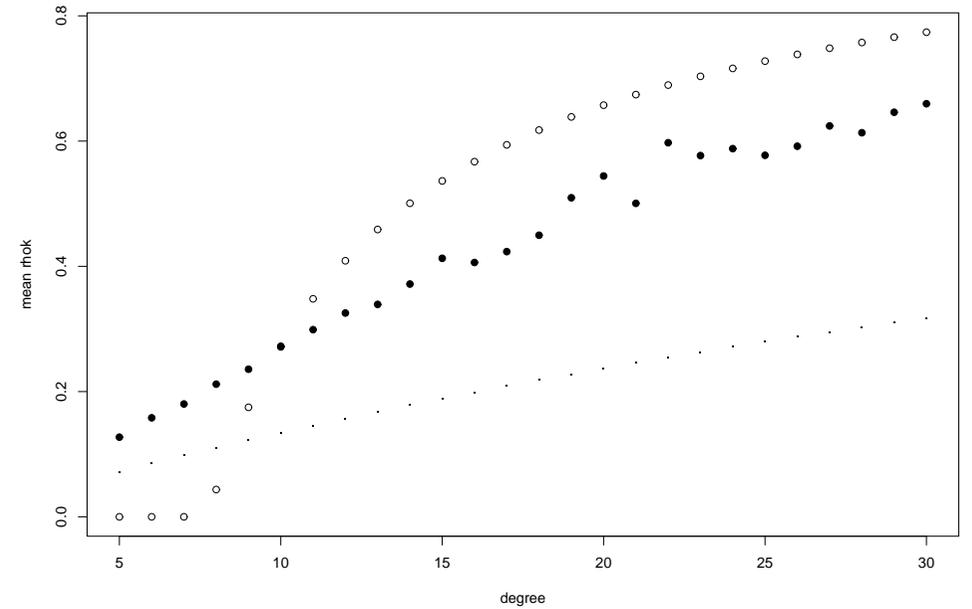


$$E[d] = 20, b.E[d] = 2$$

Résultats sur graphe en loi de puissance



$$E[d] = 9, b.E[d] = 2$$



$$E[d] = 9, b.E[d] = 1.2$$

Potentiel et limites de l'approximation par paire

- Réseaux proches de l'hypothèse d'un système "well-mixed": le champ moyen suffit
- Réseaux plus réalistes : l'approximation par paire améliore
- Mais pas suffisant pour une bonne approximation du seuil critique
- Comment prendre en compte d'autres caractéristiques du graphe : avec l'approximation de Bethe

$$t_{00 \rightarrow 01} \approx b \frac{d-1}{1-\rho} \left(C \frac{P(01)^2}{\rho(1-\rho)} + (1-C)P(01) \right)$$

Au delà ...

- Justification théorique de candidats pour fermeture des moments/méthodes variationnelles
- Extension au cas des graphes valués et à plus de deux états : plus réalistes (effet ville-village pour la coqueluche et la rougeole)