

Projet L3

Modélisation de l'interaction au sein d'un troupeau d'ovis aries à l'aide des réseaux complexes

Barzyk Alexandre et Mistretta Nicolas
encadrant : M. Peruani

24 mai 2013

Table des matières

1	Introduction	2
2	Notre modèle	2
3	Travail Préalable	3
3.1	Erdos Reyni	3
3.2	Random Recursive Tree	5
4	L'expérience	5
4.1	Vue générale	5
4.2	Relation avec la surface initiale du troupeau	6
4.3	Relation avec la position initiale par rapport à l'initiateur	7
4.4	Mise en relation du nombre de départ et du temps de latence, l'effet "avalanche"	7
4.5	Simulation sous forme de réseau métrique et topologique	8
5	Analogie avec la théorie	10
5.1	Analogie avec ER	10
5.2	Analogie avec RRT	10
6	Conclusion	10

1 Introduction

L' *Ovis aries* est une espèce à l'instinct grégaire. A l'aide de la théorie des réseaux complexes, le but est de chercher les facteurs qui déterminent cet instinct et de confronter nos résultats à une des études déjà menés sur ces mammifères. Nous jouerons avec un troupeau de moutons cantonné dans un enclos. Nous connaissons le nombre de moutons et leurs positions initiales. L'expérience est de placer à un endroit de la nourriture et de stimuler un de nos moutons pour que celui-ci repère la nourriture. Une fois ceci fait les autres moutons auront deux choix, suivre notre champion ou rester sur place. Dans la quasi-totalité des cas nos moutons vont le suivre (d'après [1]) et on mesura le temps que chacun des moutons a mis avant de partir. A l'aide de ces quelques données on cherchera à établir un réseau pour le comparer à des réseaux typiques et en tirer des conclusions. Au préalable afin d'avoir les connaissances nécessaires pour les interpréter sous forme de réseaux, nous avons étudié les plus basiques afin d'en tirer des éléments qui les caractérisent. Pour cela nous avons construit numériquement ces réseaux puis comparé nos calculs théoriques aux résultats de ces simulations.

2 Notre modèle

On a décidé de modéliser les interactions entre ces moutons par un réseau complexe ("complex network") non dirigé et non pondéré. C'est à dire que chacun des moutons est lié à un ou plusieurs autre(s) mouton(s) de la même manière suivant une certaine topologie. Ici nous avons choisis trois modèles de réseaux différents : métrique, topologique et "fully connected".

- Métrique : les moutons interagissent avec tous les autres moutons situés à une distance "d".
- Topologique : ils interagissent seulement avec les "n" plus proches moutons (dans notre fig.1 "n"=2).
- Fully connected : tous les moutons sont connectés, le réseau est obligatoirement percolé(c'est à dire qu'on peut atteindre n'importe quel noeud à partir de n'importe quel autre. Exemple : sur la fig.1 les 2 derniers réseaux sont percolés).

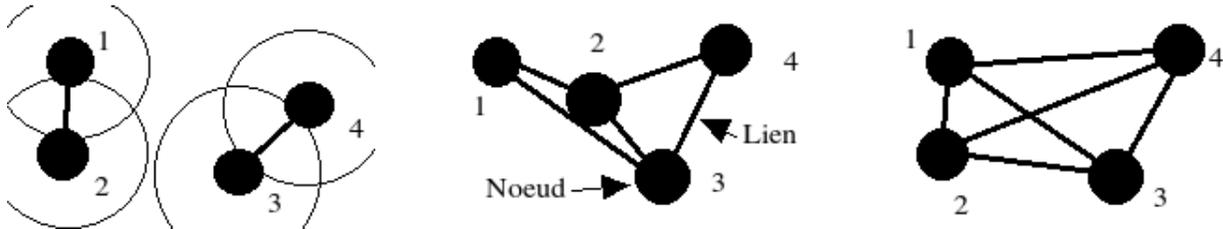


FIG. 1 – exemple de 3 réseaux dans l'ordre suivant : métrique, topologique (ici n=2) et "fully connected"

graphique	nombre de clusters	degrès des noeuds	profondeur à partir du noeud 1
métrique	2	4 de degrés 1	2
topologique	1	2 de degrés 3(num 2 et 3) et 2 de degrés 2	1 pour chaque cluster
fully	1	4 de degrés 3	1

La modélisation de la transmission de l'information se fait par le billet de deux états possibles affectés aux noeuds. Ce qu'on appellera par la suite par S ("susceptible") qui représente un mouton qui ne suit pas l'initiateur et D ("Departure") un mouton qui suit l'initiateur.

3 Travail Préalable

Pour un cours très bien détaillé voir le livre : [3] , nous nous en sommes servis comme support pour pouvoir étudier tout ce qui suit.

Pour bien comprendre la dynamique d'un système typique S(susceptible)- I(infecté) nous nous sommes familiarisés avec deux types de réseaux complexes qui peuvent servir de modèle pour nos expériences, Erdos Reyni et Random recursive tree.

NB : quand nous parlerons de ces deux systèmes on parlera de système S-I, car à la base nous étudions la propagation d'une "infection" dans une population. Quand on parlera de notre système sur les moutons on parlera plutôt de système S-D.

Pour commencer on utilise l'interaction suivante pour représenter la transmission :



\tag{2}

Par normalisation on aura dans le réseau $P_S + P_I = 1$, où P_S le pourcentage de noeuds susceptible d'être infecté, P_I le pourcentage de noeuds infecté et α la probabilité d'être infecté. Dans un tel système si celui-ci est percolé on est sûr que la totalité des noeuds sera dans l'état I au bout d'un certains temps. Sinon seul les "clusters" (un ensemble de noeuds percolé qui est isolé des autres noeuds du système voir tableau de la fig. 1) où l'information est émise seront dans un état I. Nous avons étudié deux systèmes standards pour bien comprendre l'évolution de la transmission. On déclare informatiquement AdjLink, la matrice principale qui détermine les connexion entre les noeuds. Cette matrice prendra plusieurs définition tout au long de notre projet suivant le sujet. Dans le cadre du travail préalable elle sera symétrique. Les indices des éléments de la matrice correspondent au numéro des noeuds et la valeur des éléments est 0(non connecté) ou 1(connecté). Les indices des éléments de la matrice correspondent au numéro des noeuds et la valeur des éléments est 0(non connecté) ou 1(connecté).

Exemple voici dans le même ordre les matrices AdjLink associées à nos 3 réseaux de la fig.1.

$$\text{AdjLink} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \text{AdjLink} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \text{AdjLink} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \tag{3}$$

3.1 Erdos Reyni

Le premier réseau est le système ER (Erdos Reyni) : On fixe un nombre N de noeuds puis on ajoute un nombre L de liens entre deux noeuds (plusieurs liens entre deux noeuds sont possibles). Ce système peut modéliser notre troupeau à condition que L soit grand pour qu'il y ait percolation car l'intégralité de ses individus est censé réagir. Comme le montre les simulations effectuées nous sommes dans un tel système qu'à partir d'une certaine valeur de L. Pour 32 individus 194 liens en moyenne. 2 d'après la figure 2.

Pour relever l'importance d'un noeud quelconque dans un réseau nous avons décidé de chercher

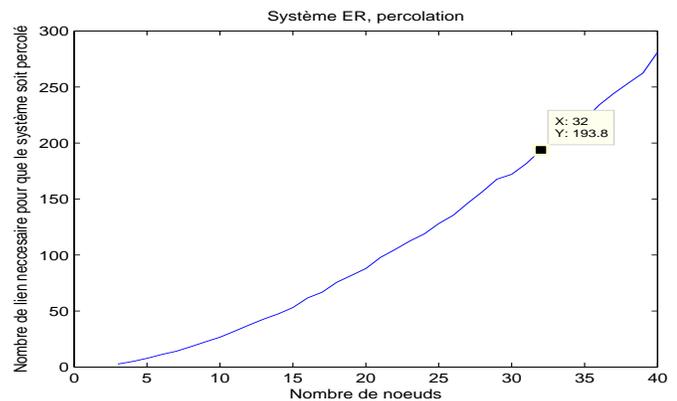


FIG. 2 – Valeur critique du nombre de liens pour un système ER

la probabilité d'être connecté à un nombre k de noeuds, appelée la distribution de degrés. On dira d'un noeud qu'il est de degrés k si il est lié à k noeuds. *En effet dans un réseau le degré d'un noeud est très important. Prenons l'exemple de l'épidémie de la grippe A(H1N1). Si nous modélisons sous forme d'un réseau complexe une ville ou un pays, tous les lieux hautements fréquentés (comme les aéroports) représenteraient des noeuds de degrés k très élevés devant les autres. C'est pourquoi on les ferme pour fragiliser au maximum le réseau et ralentir la progression de l'infection.*

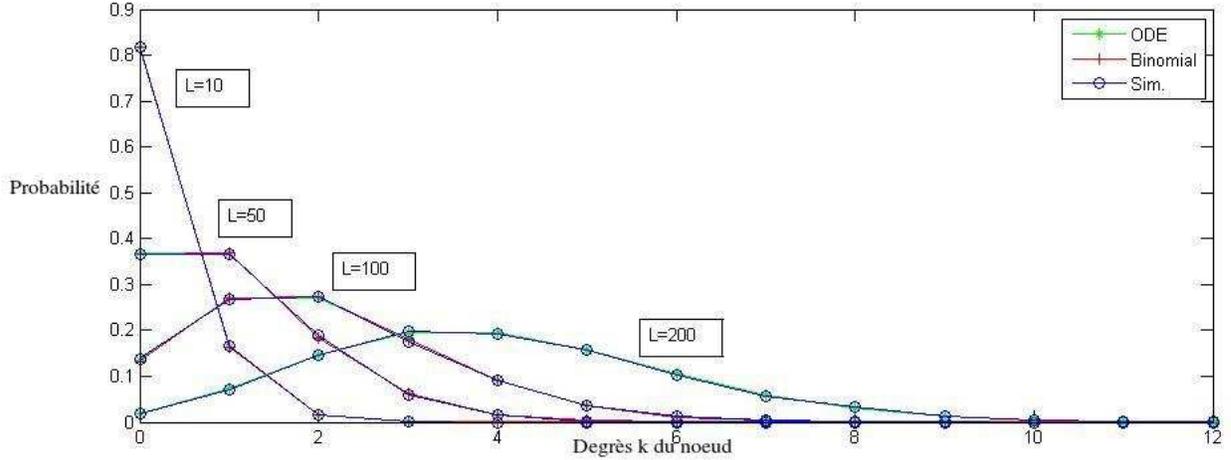


FIG. 3 – La distribution de degrés pour un système ER, on fait varier L le nombre de liens dans un système à 100 noeuds

La figure 3 correspond à la fonction de distribution de degrés résolu de 3 manières différentes. Ces courbes ont été tracées à l'aide des formules suivantes (*on ne développera pas les calculs, ils le seront dans notre site*) :

On pose : N le nombre de noeuds, L le nombre de liens, q la probabilité de connexion entre 2 noeuds, P_k la probabilité d'avoir un noeud de degrés k et n_k le nombre de noeud de degrés k

On a par définition : $\sum P_k = 1$, $q = \frac{2L}{N(N-1)}$ et $n_k = NP_k$

– La courbe ODE ("ordinary differential equation") : on a en premier lieu résolu l'équation différentielle suivante qui découle de notre système :

$$\frac{dn_k}{dt} = \alpha_{k-1}n_{k-1}(t) - \alpha_k n_k(t) \quad (4)$$

$$(5)$$

On aura α indépendant du degré et du temps donc $\alpha_k = \alpha_{k-1} = 2/N$ [9]. Le résultat donne la formule de Poisson :

$$\frac{n_k(t)}{N} = \frac{\left(\frac{2t}{N}\right)^k e^{-\frac{2t}{N}}}{k!} \quad (6)$$

– La courbe "binomiale" correspond à la résolution numérique de la loi binomiale basique.

$$P_k = \binom{N-1}{k} q^k (1-q)^{N-k-1} \quad (7)$$

Qui à pour moyenne de degrés [9] $\langle k \rangle = \frac{2L}{N}$

- La courbe "sim" correspond à la distribution de degrés observé sur nos simulations de réseaux. On voit que nos 3 courbes se superposent donc nos résultats sont cohérents.

Une autre valeur très importante après le degré du noeud est la profondeur d'un réseau. C'est à dire le nombre de liens nécessaire à un noeud pour arriver au noeud le plus éloigné de lui voir 1. Celle-ci influence énormément la vitesse de propagation d'une infection. Dans le cas des réseaux S-I nous prendrons toujours la profondeur à partir du premier noeud où l'infection se déclare.

Comme on le voit sur la figure 4 si l'on considère un système percolé de 32 noeuds notre profondeur est comprise entre 1 et 3. L'information va donc circuler très vite dans un tel système puisqu'il y a peu d'intermédiaires.

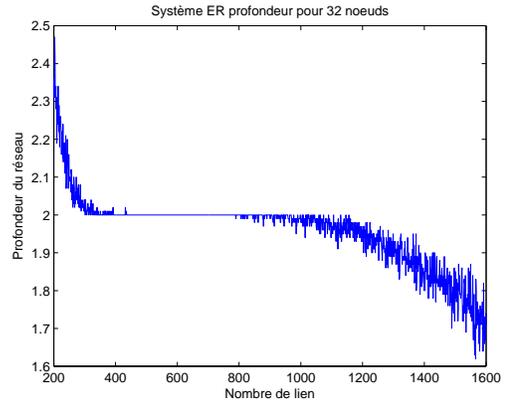
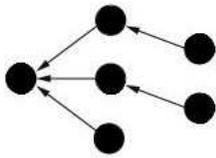


FIG. 4 – La profondeur

3.2 Random Recursive Tree



Le deuxième est un système RRT (Random Recursive Tree). L'algorithme de croissance du RRT consiste à ajouter progressivement des noeuds munis d'un ou plusieurs liens. Les paramètres du RRT sont le nombre de noeuds N et le nombre de liens associés à chaque noeud qui est ajouté au système.

Cependant nous n'avons préalablement étudié que dans le cas où l'on pose un noeud avec seulement un lien (les flèches montrent à qui le noeud s'est relié). de ce fait les résultats ne sont pas exploitables pour notre troupeau et nous développerons les résultats sur notre site (les calculs sur ce type de réseau seront aussi présentés).

Nous allons maintenant comparer les résultats issus des simulations informatiques de ces deux modèles avec l'expérience.

4 L'expérience

Résumé de l'expérience :

- Objectif : observer la transmission de l'information dans le troupeau et vérifier si une étude menée [1] sur un groupe de petite taille est utilisable pour décrire un groupe de taille plus importante.
- Sujets : 32 moutons dont un initiateur dans un enclos. Vérifier et comparer si les hypothèses d'interactions métrique et topologique sont valables au sein du troupeau.
- Données : les positions initiales de chacun des moutons et le temps de latence.
- Description : une cible attire notre initiateur puis progressivement les moutons se mettent à suivre l'initiateur. On étudie le temps de latence de chacun des moutons.



4.1 Vue générale

En premier lieu nous avons tracé pour avoir une vue générale les 6 graphes suivants pour chacune des 22 manipulations à l'aide de nos données (Fig. 5).

NB : on utilisera souvent le terme de latence à la place de temps de départ.

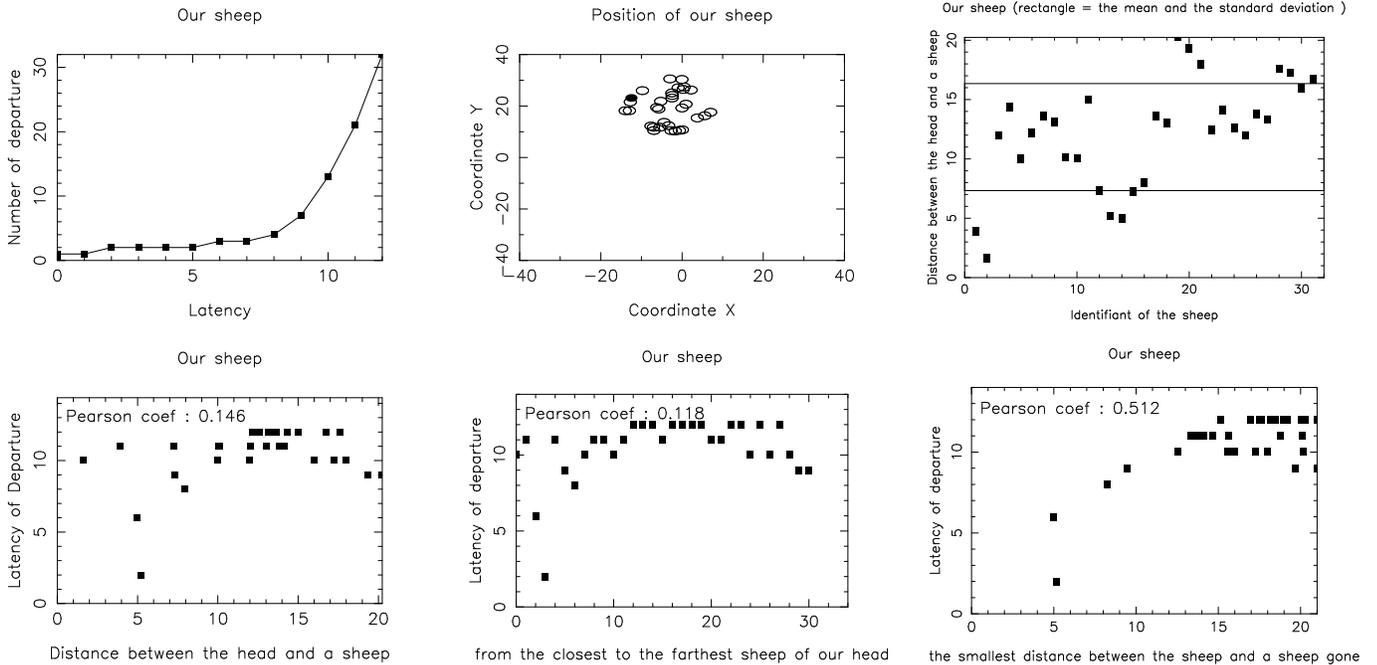


FIG. 5 – manipulations numéro 5 de

- Le premier graphe de la Fig. 5 représente D, le nombre de moutons qui suivent, en fonction du temps.
- Le deuxième de la Fig. 5 la position spatiale des moutons dans notre enclos.
- Le troisième de la Fig. 5 la distance entre l'initiateur et un mouton en fonction de l'identifiant de celui-ci dans notre fichier texte (de 2 à 32). Les deux droites correspondent à la moyenne \pm l'écart type.
(Pour les 3 derniers graphes de la Fig. 5 chaque points représente un mouton)
- Le quatrième de la Fig. 5 : en X la distance entre un mouton et l'initiateur et en Y la latence.
- Le cinquième de la Fig. 5 : en X le numéro des moutons indexés du moins éloigné de notre initiateur au plus éloigné et la latence en Y .
- Le sixième de la Fig. 5 : en X la distance entre un mouton et le mouton le plus proche parti avant lui et en Y la latence.

Chacun de nos troupeaux occupe un espace initial différent. Nous avons donc essayé de mettre en relation cette surface occupée avec la transmission de l'information.

On définit par $Pearson = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$ avec X et Y deux variables aléatoires.

4.2 Relation avec la surface initiale du troupeau

On voit sur (Fig. 6) bien que la latence moyenne, qu'elle soit prise de notre initiateur ou du premier qui le suit, n'a pas de lien direct avec la surface occupée initialement par notre troupeau (coefficient de corrélation très petit devant 1 dans les deux cas). Ce qui mène bien à la même

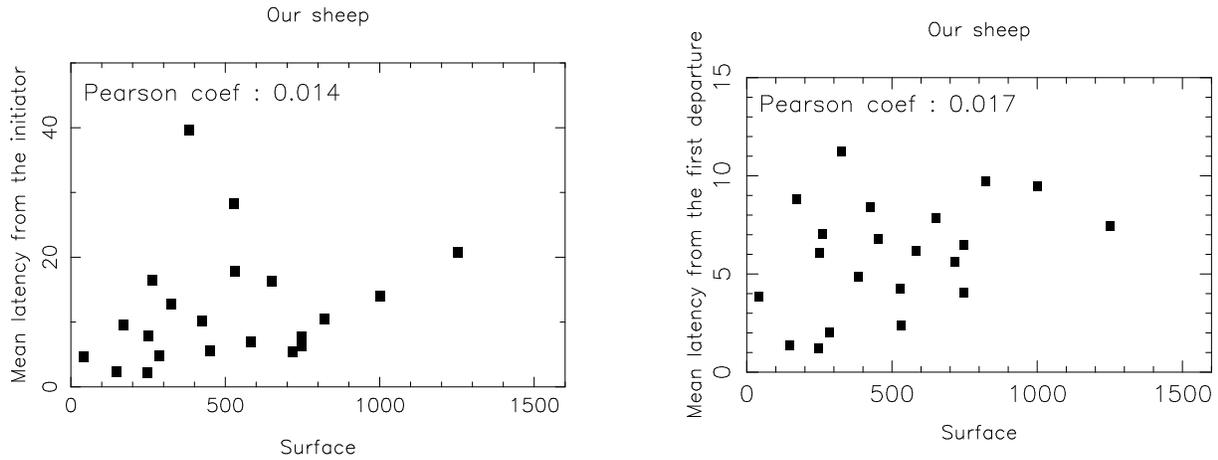


FIG. 6 – A gauche on moyenne la latence à droite on moyenne la latence en à partir du premier départ

conclusion que l' étude prise en référence [1]. Cependant on ne peut vérifier l'hypothèse qu'un groupe avec un plus grand nombre de sujets aurait un temps de latence moyen plus grand.

4.3 Relation avec la position initiale par rapport à l'initiateur

En observant les graphes 4 et 5 (Fig. 5) de toutes nos manipulations (coefficient de Pearson encore très bas), on a pu en conclure que la position initiale de l'initiateur par rapport aux autres moutons, que ce soit métrique ou topologique, n'influence pas notre temps de propagation de l'information comme on peut le voir sur ces deux graphes (Fig. 7) :

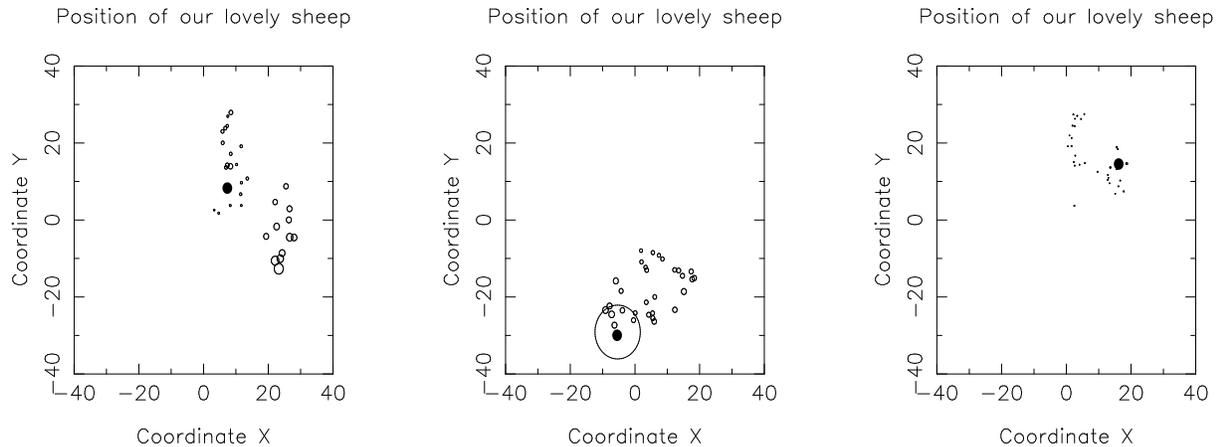


FIG. 7 – Les positions des moutons des expériences 18 19 et 20 sont représentées avec des cercles inversement proportionnelles à leurs temps de départ. L'initiateur est le cercle plein

4.4 Mise en relation du nombre de départ et du temps de latence, l'effet "avalanche"

On constate qu'il existe une valeur seuil. Le troupeau entier se met à suivre seulement si au moins entre 3 et 4 moutons décident de partir dans un bref laps de temps. Apparemment les départs

isolés (c'est à dire un unique départ durant un intervalle de temps non bref) n'ont pas d'impact direct sur la fonction de probabilité. Par manque de manipulations, on ne peut pas vérifier si à partir d'un certain seuil de départs isolés l'effet avalanche est encore présent

De même, vu que l'on considère notre système comme statique on ne peut vérifier si il s'agit du nombre de mouton en mouvement qui affecte la fonction de probabilité.

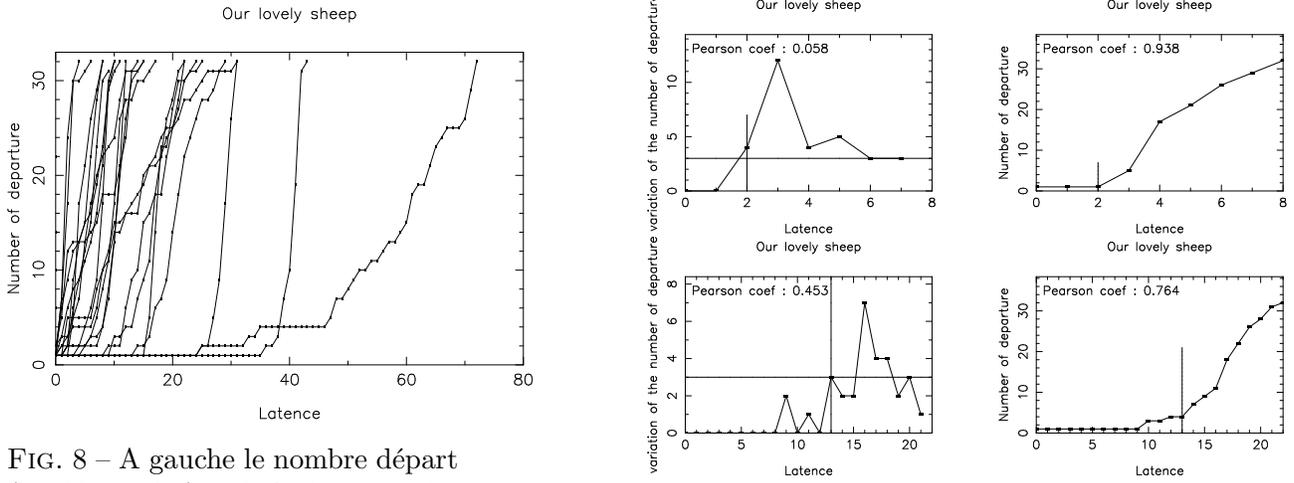


FIG. 8 – A gauche le nombre départ (les 22 manip.) et à droite sa variations (manip.6 et 21) en fonction du temps.

4.5 Simulation sous forme de réseau métrique et topologique

Notre simulation repose sur un principe très simple. Nous prenons les positions initiales de chacun de nos moutons puis nous en bâtissons un réseau.

- Réseau métrique : notre matrice $AdjLink$ décrite en éq. 3 se transforme en une matrice contenant les distances par chacun de nos moutons.
- Réseau topologique : ici $AdjLink$ se transforme en une matrice contenant l'ordre de nos moutons classé suivant leurs distances entre eux. exemple : la matrice $AdjLink$ associé au réseau "fully connected" de la fig. 1 :

$$AdjLink(Topologique) = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 1 & 0 & 2 & 3 \\ 3 & 2 & 0 & 1 \\ 3 & 2 & 1 & 0 \end{pmatrix} \quad AdjLink(Métrique) = \begin{pmatrix} 0.0 & 1.0 & 2.6 & 3.0 \\ 1.0 & 0.0 & 2.2 & 3.2 \\ 2.6 & 2.2 & 0.0 & 2.0 \\ 3.0 & 3.2 & 2.0 & 0.0 \end{pmatrix} \quad (8)$$

Comme on le constate dans le cas topologique la matrice n'est plus symétrique. En effet le fait d'être voisin est independant de la distance nous séparant.

Ensuite il nous a fallut définir la fonction de probabilité " μ " déterminant si le mouton suit ou pas. Notre but étant de vérifier si la fonction donnée par notre étude de référence[1] est applicable ici.

$$\mu(D, S) = \alpha \frac{D^\beta}{S^\gamma} \quad (9)$$

Avec $\alpha = 0.19$ $\beta = 1.16$ $\gamma = 0.6$ provient de la page 4 de [1]

Nous avons juste rajouté une fonction égale à 1 ou 0 qui dépend de la topologie de notre système ou de la distance entre nos moutons. On a donc pour notre simulation les deux fonctions de probabilités suivante :

$$\text{forme métrique} \quad \mu(D, S) = \begin{cases} \alpha \frac{D^\beta}{S^\gamma} & \text{si } d_i < d \\ 0 & \text{si } d_i \geq d \end{cases} \quad (10)$$

$$\text{forme topologique} \quad \mu(D, S) = \begin{cases} \alpha \frac{D^\beta}{S^\gamma} & \text{si } n_i < n \\ 0 & \text{si } n_i \geq n \end{cases} \quad (11)$$

Où d_i la distance entre 2 moutons, d le rayon d'interaction, n_i le numéro du mouton dans notre matrice $AdjLink$ (topologique) en éq. 8, n le nombre de connections qu'a un individu.

En d'autres termes : si on est assez proche alors on peut avoir la possibilité d'être infecté sinon non.

Enfin on trace les graphes suivants En pointillé notre expérience et en trait continu nos simulations.

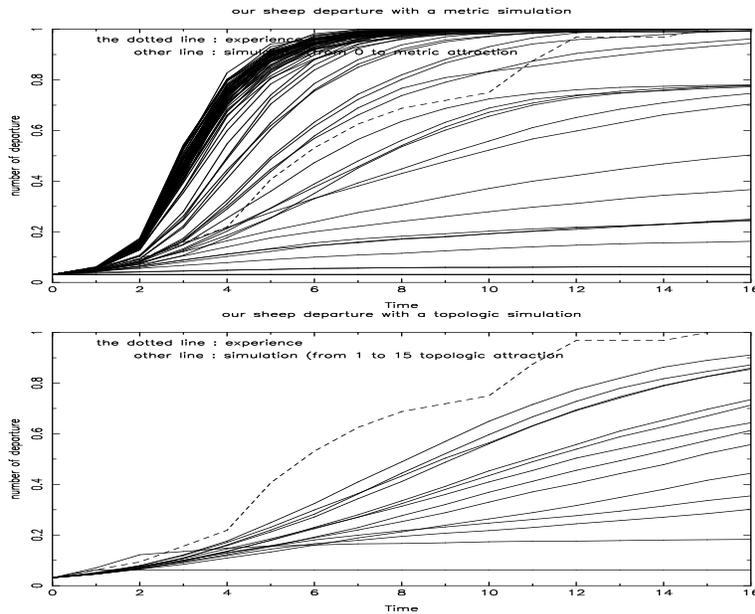


FIG. 9 – manipulations 10, interprétation métrique en haut puis topologique en bas droite (d : rayon d'interaction, n : nombre de connections par individu)

On fait varier d , ou n . Chacune des courbes correspond à une valeur différente de ce paramètre. On peut constater que nos simulations s'approchent de l'expérience réelle dans certaines manipulations (ici une manipulation affichée). Mais dans certains cas la simulation ne marche pas du tout. On peut en tirer que la position initiale de chacun des individus influe beaucoup sur le comportement du troupeau. Le modèle "fully connected" ici est moins représentatif. C'est mis en évidence sur le graphe à interaction métrique quand on a une valeur de d très grande. Donc ces résultats montrent qu'une représentation de la distribution spatiale de ces mammifères par un réseau complexe n'est pas évidente mais reste possible. Ces résultats ne remettent pas en question ceux de notre étude de référence [1], bien au contraire leur fonction de probabilité semble être un modèle acceptable si on rajoute un paramètre de portée à celui-ci.

On peut donc en conclure que

$$\mu = \mu(D, S, f) \quad (12)$$

$$\text{avec } f = f(\text{conditions initiales}, d, n) \quad (13)$$

On devrait même d'après nos premiers résultats sur l'effet "avalanche" supposer que

$$\mu = \mu(D, \frac{dD}{dt}, S, f) \quad (14)$$

5 Analogie avec la théorie

Cette étude nous montre des similitudes avec nos deux premiers réseaux et les liens entre chacun des moutons au sein du troupeau. Nous avons cherché à mettre en évidence que les connexions entre les moutons simuler avec une loi de probabilité à dépendance métrique était du type ER. Pour une interprétation métrique des interactions le modèle ER semble le mieux approprié et pour une interaction topologique RRT semble plus cohérent.

5.1 Analogie avec ER

Effectivement on a pu constater que si l'interaction est de type métrique entre nos moutons, on peut relier notre troupeau à un modèle ER. De plus avec les hypothèses de départ on sait que le système de notre troupeau est percolé, ainsi on a pu prouver que [9] :

$$\frac{\langle k^2 \rangle}{\langle k \rangle} \geq 2 \text{ où } \langle k \rangle \text{ est la moyenne de degrés des noeuds et } \langle k^2 \rangle \text{ la variance} \quad (15)$$

De plus la propagation de l'information ce faisant de manière très rapide on a un système dont la profondeur est petite et le nombre de liens les reliant est très grand devant leurs nombres. Pour un système de 32 noeuds le nombre de lien à l'entier 194 (voir figure 2)

Par manque de temps nous ne sommes pas allés plus loin mais à partir d'ici le but était de prouver **mathématiquement** que l'on pouvait utiliser le réseau théorique type ER pour décrire notre "réseau" de mouton. Le premier grand pas est de calculer la probabilité de connexion (c'est à dire la probabilité qu'il existe un lien entre 2 moutons) De ce fait on aurait pu remonter au nombre de liens unissant tous nos moutons, notre distribution de degrés k, de cette distribution on en aurait déduit le temps moyen de propagation théorique et la profondeur. Une fois ceci trouvé il aurait fallut vérifier si elles concordaient avec nos quelques affirmations de départ. Si c'est le cas alors notre réseau ER pourrait être un modèle approximatif de l'interaction entre nos moutons.

5.2 Analogie avec RRT

Pour RRT on aurait du approfondir l'étude théorique en changeant le processus de création du réseau. Au lieu de disposer un seul lien avec un noeud nous aurions disposé un noeud avec "i±p" lien(s) (avec i et p à faire varier pour créer des irrégularité dans notre réseaux). Nous pensons que ce modèle s'approche plus de la simulation topologique pour nos moutons. Nous ne l'avons pas non plus prouvé formellement mais l'instinct nous pousse à le supposer.

Ensuite une fois cela effectué on aurait déterminé les entiers i et p pour qu'ils concordent le mieux au expérience, de la on aurait cherché la probabilité de connexion. Puis après la distribution de degrés, la profondeur etc...(a peu près les mêmes caractéristiquent que ER)

6 Conclusion

Finalement d'après nos résultats l'instinct grégaire n'est pas facile à modéliser mathématiquement du moment où le nombre d'individu est assez grand. Ils nous aurait fallut plus de temps mais surtout plus de données sur la dynamique de nos expériences après l'instant initial. Nos conclusions sont que d'une part ,l'étude de référence[1] est cohérente pour un petit nombre de moutons. D'autre part un mixte entre l'interaction métrique et topologique à l'intérieur du troupeau, qui dépend des conditions initiales, semblerait mieux décrire la dynamique d'un troupeau de taille plus significative. Il faudrait par exemple travailler par exemples sur les conditions initiales ou faire une fonction de portée métrique dans une interaction topologique. Pour affiner la dynamique de l'information on pourrait travaillé avec un système S-D-S, dont nous avons préalablement étudié les caractéristiques et dans laquelle un mouton dans l'état D peut passer à l'état S. Ou encore rajouter un nombre seuil de départ pour déclencher "l'avalanche", c'est à dire la mise en mouvement en un temps bref de la quasi totalité du troupeau dans la direction de l'initiateur.

En remerciant Monsieur Eric Aristidi pour ses réponses très rapides aux mails et son aide pour les problèmes de codages C et surtout à notre responsable de projet monsieur Fernando Peruani qui a été très présent lors de l'ensemble du semestre.

Références

- [1] Pillot MH, Gautrais J, Arrufat P, Couzin I, Bon R, Deneubourg JL, "Scalable Rules for Coherent Group Motion in a Gregarious Vertebrate" PLoS ONE, e14487(2011).
- [2] M. E. J. Newman, I. Jensen, R. M. Ziff, "Percolation and epidemics in a two-dimensional small world", Phys. Rev. E, volume 65, 021904 (2002).
- [3] G. Caldarelli, A. Vespignani, "Large Scale Structure and Dynamics of Complex Networks : From Information Technology to Finance and Natural Science (Complex Systems and Interdisciplinary Science)".
- [4] F. Peruani and L. Tabourier, "Directedness of information flow in mobile phone communication networks", PLoS ONE ; e28860 (2011).
- [5] M. E. J. Newman, R. M. Ziff , "Convergence of threshold estimates for two-dimensional percolation", Phys. Rev. E 66, 016129 (2002).
- [6] A. Srivastava, B. Mitra, N. Ganguly, F.Peruani, "Correlations in complex networks under attack", Phys. Rev. E 86, 036106 (2012).
- [7] J. Park, M. E. J. Newman , "The statistical mechanics of networks" (2004).
- [8] O. Melchert, "Percolation thresholds on planar Euclidean relative-neighborhood graphs" Phys. Rev. E 87, 042106 (2013)
- [9] voir notre site web où tout les calculs seront détaillé.