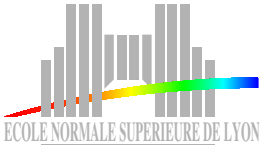
ÉCOLE NORMALE SUPÉRIEURE DE LYON
Département de Mathématiques et Informatique



RAPPORT DE STAGE

de 2^e année du Master de Recherche en Informatique Fondamentale
soutenu par

Sylvain Sené

le 30 juin 2005

Modèle de diffusion de la confiance pour les réseaux ad hoc

Directeur de stage : Michel Morvan

Année universitaire 2004-2005

Remerciements

Il est indéniable que l'intégralité de ce stage n'aurait pu se dérouler dans de si bonnes conditions sans l'aide et le soutien d'un certain nombre de personnes travaillant toutes au LIP. En effet, de nombreux chercheurs ont su être présents pour que le projet puisse avancer dans les meilleures conditions possibles.

Tout d'abord, je tiens particulièrement à remercier Michel Morvan, mon directeur de stage, qui a su me mettre à l'aise fort rapidement. Sa connaissance des systèmes complexes et les explications qu'il m'en a faites m'ont permis de rapidement m'accoutumer aux problèmes posés par le projet. Par ailleurs, il a su m'apprendre les méthodologies de recherche à adopter et a su être à mon écoute aux moments où j'en avais besoin afin de m'aider à prendre les meilleures décisions.

Je souhaite aussi remercier Emmanuelle Lebhar, Nazim Fatès et Nicolas Schabanel qui ont toujours su répondre aimablement à chacune de mes questions, notamment lorsque M. Morvan était absent. Par ailleurs, leur relecture de ce rapport se sont avérées fort appréciables.

Enfin, je tiens à ne pas oublier mes camarades stagiaires ainsi que toutes les personnes avec qui j'ai partagé des repas, des cafés et des discussions. Ces petits moments agréables et enrichissants sont pour beaucoup dans le plaisir que j'ai pris à travailler.

Table des matières

Introduction

"Espérance ferme que l'on place en quelqu'un, en quelque chose, certitude de la loyauté d'autrui" et *"conviction que les autres peuvent avoir de votre sincérité, de votre dévouement, de votre honnêteté."*
Dictionnaire de l'Académie française, 2005

Voici comment l'Académie française définit le terme de confiance. Il s'agit d'une notion assez naturelle que chaque être acquiert et utilise systématiquement pour décider si un échange avec un autre individu, quelle que soit sa nature, est envisageable (ou pas). Il faut remarquer que, de façon générale, la confiance que nous plaçons en une personne ne dépend pas uniquement de notre propre connaissance de celle-ci mais aussi de celles que possèdent les personnes qui nous entourent à son sujet. Ceci illustre l'existence de la diffusion de confiance dans la vie courante. Il semble donc raisonnable que les réseaux informatiques procèdent de manière identique.

Définition 0.1 [8] *Un* **réseau ad hoc** *(aussi appelé réseau ambiant ou spontané) est une collection de nœuds fortement mobiles, distribués et indépendants, capables de s'organiser et de constituer un réseau sans s'appuyer sur une infrastructure fixe.*

Les réseaux ad hoc ont commencé à être étudiés dans les années 1980 pour résoudre les problèmes inhérents aux communications sur les champs de bataille [16]. L'objectif était alors de développer un réseau radio entre des "mobiles" (i.e. des soldats, des véhicules...) dispersés sur un terrain ne possédant aucune infra-structure de communication. Un tel contexte interdit l'usage des réseaux cellulaires classiques dans lesquels chaque mobile est relié aux autres par l'intermédiaire d'une station centrale. L'approche indispensable pour mettre en place des réseaux ad hoc est donc totalement décentralisée, ce qui signifie que lorsqu'un individu A veut communiquer avec un autre individu B et que la portée radio de A n'atteint pas B, l'utilisation d'individus relais est indispensable. Par ailleurs, la mobilité du réseau implique une topologie changeante et une connectivité non garantie. Ces deux principales contraintes imposent la conservation de toute information au niveau local. En effet, chaque composante du réseau doit maintenir ses propres connaissances sur les autres.

Par ailleurs, lorsqu'on étudie les réseaux ad hoc, il s'avère intéressant d'étudier différents modèles de graphes capables de les représenter, notamment les **graphes complets** et les **graphes à distribution de degrés fixée suivant une loi de puissance**. Ces derniers sont un moyen de caractériser les réseaux d'interaction sociale dont les réseaux ad hoc sont généralement très proches.

Le stage que j'ai effectué s'inscrit dans le cadre de l'étude des systèmes complexes. Un **système complexe** est un système dans lequel des objets interagissent

les uns avec les autres via des règles d'interaction locales et qui produit des résultats visibles au niveau global qui ne peuvent être directement déduits de la simple juxtaposition des interactions locales.

Ce travail consiste en l'étude et la validation par simulations d'un modèle de diffusion de la confiance pour les réseaux ad hoc, de manière à évaluer les avantages et les inconvénients du principe de diffusion. Pour ce faire, notre objectif a été de mettre en place un protocole de confiance simple et dynamique. Dans ce protocole, la notion de confiance est à la fois celle en les personnes elles-mêmes mais aussi celle en les connaissances que ces dernières possèdent au sujet des autres. À titre d'exemple, nous pouvons noter qu'au cours d'une commission scientifique, certains chercheurs peuvent avoir une totale confiance en les qualités scientifiques de certains spécialistes mais ne pas avoir forcément confiance en leur avis sur les autres.

Le protocole étudié doit résister à de nouvelles attaques inhérentes aux réseaux ad hoc. Les simulations sont par conséquent d'une grande importance pour la validation de ce nouveau protocole.

Ce rapport présente les recherches et les résultats de quatre mois et demi de travail passés au sein de l'équipe Modèle de calcul et complexité (MC2) du Laboratoire de l'informatique du parallélisme (LIP) de Lyon. À la suite d'une brève présentation du projet dans lequel le protocole étudié sera intégré, nous évoquerons les concepts de confiance existant dans les réseaux ad hoc puis nous exposerons l'état de l'art concernant les réseaux sociaux. Nous continuerons par l'explication détaillée du modèle et du protocole. Enfin, nous terminerons en décrivant les simulations effectuées et en analysant leurs résultats.

1 Préliminaires et état de l'art

1.1 Le projet *Knowledge Authentication Ambient*

Le projet KAA [2, 6] de l'ACI "Sécurité et Informatique" [1], dans lequel ce stage se déroule, est un projet inter-disciplinaire regroupant des équipes de recherche en informatique, mathématiques et sciences sociales. Il vise à mettre en œuvre un système ambiant de sécurité basé sur les connaissances pour la gestion d'échanges entre des appareils communiquants autonomes mis en réseau de manière spontanée. Les communications entre systèmes autonomes, à savoir les objets de la vie quoti-dienne dotés de capacités de communication hétérogènes (i.e. les téléphones mobiles, les PDA, les ordinateurs portables...), ainsi que les échanges de données qu'ils peuvent effectuer constituent un enjeu technique mais également un véritable enjeu social.

L'idée majeure du projet KAA est de reprendre l'étude de l'existant en partant des notions fondamentales des mécanismes de gestion de la confiance issues des systèmes sociaux et d'en déduire un modèle technologique. Ceci passe par son étude mathématique grâce aux outils des systèmes dynamiques tels que les graphes d'interaction et son implantation dans une plate-forme expérimentale.

1.2 La confiance dans les réseaux ad hoc

La notion de confiance dans les réseaux ad hoc [8] s'est fondée sur celle des réseaux informatiques traditionnels, à savoir qu'elle s'est jusqu'à présent uniquement basée sur des aspects de sécurité (technologiques) et non sur des aspects so-ciologiques (humains). En effet, dans le monde des télécommunications, la confiance a été gérée par des modèles reposant sur la préalable connaissance des identités. Ainsi, la confiance entre deux parties n'est rendue possible que par la transmission antérieure d'informations. Cette condition amène résolument à des modèles binaires et contraignants imposant trois étapes : l'identification, l'authentification et le maintien de la confiance.

Ces trois clés de voûte de la notion actuelle de confiance dans les réseaux "classiques" sont un frein à l'utilisation de tels modèles pour les réseaux ambiants (i.e. ad hoc) dont les caractéristiques sont très différentes : des topologies fortement dynamiques, un passage à l'échelle incontrôlé et l'anonymat de la population. La question qui se pose alors est la suivante : comment rendre possible la validation de l'identité d'un appareil puis la reconnaissance de cette identité tout au long d'un échange dans les réseaux ad hoc ? De nombreux travaux précédents se sont attachés à l'étude de ce problème et ont proposé des solutions que nous pouvons regrouper en deux catégories : la première préconisant l'émulation d'un superviseur [7] (i.e. plusieurs nœuds du réseau jouent le rôle d'un unique superviseur) et la deuxième fondée sur l'absence de cette tierce partie [4].

Ces deux approches donnent une idée restreinte de l'ensemble des solutions proposées. Néanmoins, la notion de confiance dans les réseaux ad hoc n'est aujourd'hui

perçue que comme relative aux identités des objets mobiles et, à aucun moment, la recherche informatique ne s'est intéressée à son aspect sociologique. En effet, les hommes, en dehors de l'identité de la personne avec laquelle ils interagissent, qui est un facteur essentiel, donnent un poids très important à la qualité des échanges réalisés au cours du temps pour établir leur degré de confiance.

Ainsi, il semble que l'usage de la cryptographie ne soit pas un moyen unique et suffisant pour parvenir à établir des modèles de confiance dans les réseaux ambiants. Combiner les concepts de sécurité informatique à des modèles développés autour des concepts de communautés, toujours plus proches des modèles sociaux tels que la réputation, le crédit ou encore la recommandation, semble permettre de contourner les problèmes inhérents aux réseaux spontanés. Une telle solution ne serait d'ailleurs pas seulement une avancée dans ce type de réseaux mais pourrait également avoir un grand intérêt dans les réseaux "classiques" comme l'internet. Nous allons maintenant introduire la notion de réseau social et présenter des modèles permettant de les étudier car nous simulerons le protocole proposé sur de tels réseaux.

1.3 Les réseaux sociaux et les graphes aléatoires

1.3.1 Les réseaux sociaux

Définition 1.1 [12] *Un* **réseau social** *est un ensemble de personnes (ou de groupes de personnes) interagissant les unes avec les autres selon un certain modèle.*

Cette définition montre que les réseaux sociaux englobent de nombreux traits de la vie en société tels que l'amitié entre les individus, les relations d'affaires entre les sociétés ou encore les collaborations entre scientifiques et le *World Wide Web* (WWW)... De nombreuses expériences ont été réalisées sur ce type de réseaux. À titre d'exemple, la plus célèbre est certainement celle de Milgram sur les graphes petit-monde [9]. Ce dernier a observé des trajets de lettres entre des paires d'individus qui ne se connaissaient pas. Les individus ne devaient utiliser que leurs relations personnelles pour transmettre la lettre à son destinataire, sur lequel ils avaient des données partielles (profession, ville...). Les résultats ont montré qu'un quart des lettres sont arrivées mais que celles qui sont parvenues à destination ont utilisé un très petit nombre de relais, à savoir six en moyenne.

L'arrivée de l'internet chez les particuliers et son impact grandissant a renforcé l'idée que l'étude des réseaux sociaux est essentielle à la compréhension de nombreux phénomènes de société mais aussi au développement de réseaux de types nouveaux tels que les réseaux ad hoc. À la suite d'une présentation de l'évolution des recherches relatives à ce type de réseaux, nous introduirons brièvement les graphes à distribution de degrés fixée suivant une loi de puissance (cf. définition 1.4) qui sont un premier pas dans la direction d'une bonne représentation des réseaux sociaux. Par ailleurs, le fait que deux nœuds ont plus de chance d'être connectés s'ils ont un voisin en commun a été largement observé dans de tels réseaux.

1.3.2 Le modèle d'Erdös et Rényi

On représente de façon naturelle un réseau par un graphe $G = (S, A)$ où S est l'ensemble des sommets (les nœuds du réseau) du graphe et A l'ensemble de ses arêtes qui relient deux sommets entre eux. La théorie des graphes a été introduite au $XVIII^e$ siècle par L. Euler avec son célèbre problème des ponts de Königsberg. Au XX^e siècle, les problèmes liés à cette théorie ont plus été étudiés dans le domaine des statistiques et de l'algorithmique. Ainsi est apparue la notion de graphe aléatoire avec P. Erdös et A. Rényi.

À la fin des années 1950 et au cours des années 1960, Erdös et Rényi ont étudié l'un des premiers modèles théoriques des réseaux, les graphes aléatoires, dont ils ont donné plusieurs versions. La plus étudiée est notée $G_{n,p}$.

Définition 1.2 [5] *Étant donnés $p, n \geq 0$, un* **graphe aléatoire** *$G_{n,p}$ est un graphe construit à partir de n sommets, en reliant chaque paire de sommets indépendamment avec probabilité p.*

Le graphe ainsi créé représente un réseau dans lequel toutes les $\frac{n(n-1)}{2}$ arêtes sont créées indépendamment avec une certaine probabilité p. L'une des propriétés intéressantes de ce type de graphe concerne le degré moyen des sommets, que nous noterons δ. Le nombre moyen d'arêtes étant égal à $\frac{n(n-1)p}{2}$, on peut déduire que le degré moyen est

$$\delta = \frac{n(n-1)p}{n} = (n-1)p \simeq np$$

où l'approximation finale peut être admise lorsque n est suffisamment grand. La distribution de degrés dans un tel graphe suit une loi binomiale de paramètres $n-1$ et p. Donc la probabilité que le degré δ_i d'un nœud i soit k est donnée par :

$$p_k = p(\delta_i = k) = C_{n-1}^k p^k (1-p)^{n-1-k}$$

ce qui peut s'approximer par une distribution de degrés suivant une loi de Poisson lorsque n tend vers l'infini (cf. [3]). D'où :

$$p_k \sim \frac{\delta^k e^{-\delta}}{k!}$$

Définition 1.3 *Une* **composante connexe** *est un sous-ensemble maximal de sommets d'un graphe dans lequel chacun des sommets est accessible à partir des autres par un chemin à travers le réseau. La* **composante géante** *d'un graphe est la composante connexe du graphe qui est composée du plus grand nombre de sommets.*

Une autre caractéristique remarquable est l'existence d'une transition de phase [1] avec l'augmentation de δ qui amène à la présence d'une composante géante. La formation d'une telle composante a été remarquée dans de nombreux réseaux réels.

[1] Une **transition de phase** représente une discontinuité dans la progression de l'état d'un système dynamique, qui perd alors un ensemble de propriétés pour en satisfaire de nouvelles.

Définition 1.4 *On dit que la distribution de degrés d'un graphe suit une* **loi de puissance** *lorsque la probabilité qu'un nœud quelconque ait un degré égal à k décroît polynomialement en fonction de k.*

Malgré la propriété commune de l'existence d'une composante géante des graphes aléatoires d'Erdös et Rényi et des réseaux réels, les recherches récentes ont montré que le modèle d'Erdös et Rényi est loin d'être une bonne représentation des réseaux réels. Il s'avère en effet que la caractéristique principale des réseaux sociaux est qu'ils s'illustrent par une distribution de degrés suivant une loi de puissance et non par une distribution de degrés suivant une loi de Poisson. À ce titre, de nombreuses expériences (cf. [12]) illustrant la distribution de degrés des réseaux réels communément étudiés (WWW, collaborations entre scientifiques...) ont obtenu une distribution de degrés qui s'approche effectivement d'une loi de puissance.

Définition 1.5 *Un réseau présente du* ***clustering*** *(ou transitivité) si la probabilité que deux nœuds soient connectés l'un à l'autre est plus grande lorsqu'ils possèdent un voisin commun. Le* **coefficient de *clustering*** *(ou interconnectivité locale), noté $\mathcal{C}$, se définit comme la probabilité moyenne que deux nœuds voisins d'un troisième soient également connectés l'un à l'autre.*

D'après [15], les réseaux sociaux présentent aussi un fort *clustering*. En effet, si l'on considère une étude réalisée par Pastor-Satorras et ses collaborateurs en 2001 sur les sites WWW reliés par des "hyper-liens" (i.e. 153127 sites ayant un degré moyen de 35.2), le coefficient de *clustering* obtenu est de 0.11 alors qu'il est de 0.00023 avec le modèle d'Erdös et Rényi ; celui observé en étudiant les collaborations entre directeurs de sociétés est de 0.59 dans la réalité et de 0.0019 avec le modèle $G_{n,p}$.

1.3.3 Les graphes à distribution de degrés fixée suivant une loi de puissance

Les graphes à distribution de degrés fixée suivant une loi de puissance, contrairement au modèle de graphes d'Erdös et Rényi, illustrent effectivement la propriété des réseaux sociaux, c'est-à-dire qu'une faible mais non négligeable proportion de nœuds possède un degré important.

Définition 1.6 *Le* **polylogarithme** *$Li_n(x)$, aussi connu sous le nom de fonction de Jonquière, est la fonction*

$$Li_n(x) = \sum_{k=1}^{\infty} \frac{x^k}{k^n}$$

Remarque 1.7 *Les caractéristiques de la fonction de Jonquière intéressantes pour le modèle sont notamment qu'elle s'annule pour $x = 0$ et que sa croissance est finie et monotone dans l'intervalle $[0; 1[$ pour toute valeur de n.*

La distribution de degrés suivant une loi de puissance utilisée dans [13, 14] est donnée par :

$$p_k = \begin{cases} 0 & \text{if } k = 0 \\ Ck^{-\tau}e^{-k/\kappa} & \text{if } k \geq 1 \end{cases}$$

où C, τ et κ sont des constantes. Les constantes τ et κ servent à la mise en place du seuil de coupure (*cutoff*) pour la transition de phase. La constante C est fixée à $[Li_\tau(e^{-1/\kappa})]^{-1}$.

En se basant sur une "fonction génératrice" $G_0(x)$ définie comme

$$G_0(x) = \sum_{k=0}^{\infty} p_k x^k$$

on obtient par substitution la nouvelle fonction génératrice suivante :

$$G_0(x) = \frac{Li_\tau(xe^{-1/\kappa})}{Li_\tau(e^{-1/\kappa})}$$

Le paramétrage de cette fonction grâce aux constantes τ et κ permet d'engendrer des graphes possédant ou non une composante géante. La figure 1 présente les seuils de coupure de ces constantes à partir desquels une composante géante se forme.

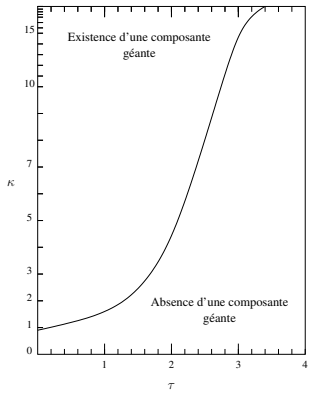


FIG. 1 – Transition de phase en fonction de τ et κ.

La procédure pour créer de tels graphes, fondée sur le couplage de demi-arêtes, est donnée dans l'annexe A. Une modification est aussi proposée pour augmenter le coefficient de *clustering*.

2 Le protocole de confiance

Le principe général du protocole de confiance vise à faire évoluer au cours du temps une relation de confiance entre les nœuds d'un réseau basée à la fois sur leurs propres connaissances et sur les connaissances des autres. Il repose en conséquence

sur trois points fondamentaux : la décentralisation des informations, leur diffusion et la maintenance d'historiques des interactions. La décentralisation est une conséquence directe de la nature des réseaux sur lesquels le modèle doit s'appliquer. En revanche, la diffusion comme la maintenance d'historiques proviennent directement de la nature même du modèle que nous voulons mettre en place. En effet, l'évolution des relations de confiance utilisant les connaissances de chacun, les nœuds doivent divulguer les renseignements qu'ils ont en leur possession. De plus, en partant du fait que la notion de confiance que nous voulons mettre en place est intimement liée à la qualité des interactions exécutées, il est indispensable que les interactions antérieures servent, nous verrons plus après dans quelle mesure, dans l'élaboration d'une nouvelle mesure de confiance.

Un protocole de confiance est dit **efficace** (nous parlerons d'**efficacité**) s'il vise prioritairement à éviter la réalisation d'interactions néfastes et **efficient** (nous parlerons d'**efficience**) s'il vise principalement à favoriser la réalisation d'interactions bénéfiques. L'objectif que nous nous sommes fixé est de créer un protocole de confiance basé avant tout sur l'efficacité et pas nécessairement sur l'efficience.

Par ailleurs, on peut évaluer les qualités et défauts du protocole selon plusieurs critères. Tout d'abord, nous pouvons vérifier le nombre de nœuds honnêtes réussissant à percevoir la malhonnêteté présente dans le système. Ensuite, nous pouvons nous intéresser à la moyenne du nombre d'interactions (bonnes et mauvaises) que les nœuds honnêtes reçoivent de tous les nœuds malhonnêtes. Enfin, nous pouvons aussi vérifier le nombre total d'interactions néfastes que les nœuds malhonnêtes ont réussi à fournir aux nœuds honnêtes. Nous observerons ces paramètres sur les graphiques résultant des simulations (cf. annexes).

Enfin, pour vérifier la pertinence du protocole, nous avons utilisé une approche par simulations. En effet, pour de simples raisons de moyens, nous n'avons pas pu mettre en réseau des dizaines, voire des centaines d'objets communiquants. Par conséquent, tout en respectant les règles liées au type de réseaux traités, nous avons créé un modèle de simulation abstrayant l'aspect distribué. L'objectif de ce modèle est, pour un temps, de ne pas se préoccuper des problèmes liés au réseau et, ainsi, de pouvoir se concentrer sur ceux relatifs au protocole de confiance lui-même. **Avec cette approche, nous voulons simuler des interactions entre des entités de manière qu'à la fin, l'ensemble de ces interactions ressemble à celles qui auraient eu lieu entre les éléments d'un réseau ad hoc**. Pour ce faire, nous fixons le réseau à l'avance (graphes complets, graphes à distribution de degrés fixée suivant une loi de puissance, graphes à distribution de degrés fixée suivant une loi de puissance ayant subi une augmentation du coefficient de *clustering*) et nous simulons des interactions ayant lieu uniquement entre des sommets adjacents. Nous faisons donc l'hypothèse que le graphe induit par les interactions dans un réseau ad hoc réel est de l'une de ces formes.

Cette partie vise à présenter dans un premier temps l'architecture du modèle de diffusion de confiance que nous allons étudier avant de nous intéresser au cœur

même du modèle, qui concerne la gestion de sa dynamique.

2.1 L'ossature du modèle

2.1.1 Une totale décentralisation

Une des nécessités du modèle est que, pour un nœud quelconque, chaque interaction le mettant en jeu soit traitée localement par celui-ci. Ainsi, toutes les informations utiles à l'évolution du système sont transmises par le réseau et sont ensuite récupérées puis utilisées par les usagers.

Ainsi, les informations qu'un nœud i conserve peuvent être rassemblées sous la forme d'un vecteur $\mathcal{V}_i$, que nous appellerons vecteur de confiance. Chaque nœud du système possède par conséquent son propre vecteur de confiance.

Définition 2.1 *Le j^e élément $\mathcal{V}_i[j]$ du* **vecteur de confiance** *$\mathcal{V}_i$, avec $i \neq j$, contient les informations que le nœud i possède sur le nœud j.*

Remarque 2.2 *Notons que l'élément $\mathcal{V}_i[i]$ ne contient pas d'information car un nœud ne conserve pas de notion de confiance sur lui-même ; en effet, un nœud n'interagit pas avec lui-même.*

Avec cette approche, les nœuds doivent conserver de façon permanente les informations qu'ils possèdent. Lorsqu'un utilisateur se déconnecte du réseau, il maintient alors encore tous les renseignements qu'il a pu obtenir jusqu'à ce moment. Ainsi, s'il décide de se reconnecter, alors, si la topologie du réseau n'a pas changé (i.e. la reconnexion l'amène à être entouré des mêmes voisins), il possède déjà des informations à leur sujet et peut choisir, si les relations de confiance entretenues auparavant l'y encouragent, de procéder à de nouveaux échanges. En revanche, si la topologie du réseau a changé (i.e. les voisins qu'il avait alors se sont déplacés depuis ou lui-même ne s'est pas reconnecté au même endroit que précédemment), il se retrouve avec de nouveaux voisins qu'il ne connait peut-être pas encore. Dans ce cas, il est possible que les données qu'il a conservées sur son ancien voisinage lui permettent de se créer une opinion sur ses nouveaux voisins car ces derniers ont potentiellement déjà pu interagir les uns avec les autres.

Les nœuds du réseau ne possédant pas de mémoire partagée, le réseau est le seul moyen permettant de partager des données. Ainsi, l'envoi régulier de messages est un point essentiel pour faire évoluer le système. À terme, le protocole de confiance que nous allons présenter devra limiter au maximum la taille des messages transmis sans que ces diminutions n'affectent les qualités du modèle. Néanmoins, bien qu'il s'agisse d'un point important, nous avons tout d'abord voulu nous assurer de la pertinence du protocole sans cette contrainte (i.e. les nœuds diffusent l'ensemble des informations qu'ils possèdent) car l'objectif de ce travail a été de vérifier la validité du protocole pour savoir si les recherches ont des chances d'aboutir à un protocole de confiance utilisable en pratique. Nous proposerons dans la conclusion quelques axes qui permettraient de modifier le protocole pour prendre en compte la limitation de la taille des échanges.

En prenant en compte les contraintes liées à cet aspect décentralisé, le modèle de simulation a été représenté par une matrice dans laquelle chaque ligne correspond au vecteur de confiance d'un nœud du système. Cette représentation matricielle permet donc de stocker l'ensemble des informations relatives au réseau d'entités étudié et, par conséquent, n'est autre qu'une représentation alternative du graphe du réseau.

2.1.2 Les critères de confiance

Lorsqu'on souhaite créer un protocole de confiance, l'une des premières questions à laquelle il est utile de trouver une réponse est : sur quoi voulons-nous qu'il se base ? Plus clairement, sur quoi se base un nœud du réseau pour décider d'effectuer ou non un échange avec un autre ? Quelles données conserve-t-il dans son vecteur de confiance ? Dans la suite de cette partie, nous allons nous intéresser à la cellule $V_i[j]$ du vecteur de confiance du nœud i. Elle conserve toutes les informations que le nœud i possède et utilise au sujet du nœud j.

Avant qu'un nœud décide d'interagir avec un autre, il va vérifier si la mesure de confiance qu'il a en celui-ci est favorable ou non. Cette mesure de confiance, que nous appellons dans le protocole la **note de confiance**, est donc la clé de voûte du système par laquelle l'évolution du système va être régie.

Comme nous l'avons évoqué au début du rapport, le modèle de confiance que nous voulons mettre en œuvre est fondé sur deux critères principaux. Les nœuds doivent non seulement utiliser leur propre expérience mais encore se servir de l'ensemble des connaissances des autres pour maintenir leurs informations. Nous allons à présent détailler ces deux critères.

• La liste des notes de confiance antérieures

Lorsque le nœud i souhaite interagir à l'instant t avec le nœud j, il possède un certain nombre d'informations personnelles à son sujet qu'il a lui-même mis à jour durant l'intervalle de temps $[0, t-1]$. Ces informations personnelles ont été acquises à chaque fois qu'il a interagi avec j auparavant. Par conséquent, il est légitime de penser que l'expérience personnelle sur laquelle se base le nœud i pour décider ou non d'interagir avec le nœud j correspond aux résultats des interactions antérieures qu'il a effectuées avec j. Ainsi, s'il s'aperçoit que j lui a envoyé plus de mauvaises que de bonnes informations dans le passé, son expérience personnelle signalera qu'une nouvelle interaction avec ce nœud a de fortes chances de ne pas s'avérer bénéfique pour lui et que j n'est donc pas digne de confiance.

À présent, la question qui se pose est de savoir combien d'interactions antérieures doivent être prises en compte pour que l'expérience personnelle soit un critère fiable. Tout d'abord, on peut imaginer utiliser l'ensemble de toutes les interactions déjà réalisées avec j. Cette idée ne semble pas convaincante dans le sens où j pourrait choisir de rester "honnête" pendant un très grand nombre d'interactions t, afin que la confiance que lui porte i augmente jusqu'à atteindre un niveau maximum, puis de lui envoyer un virus à l'interaction $t+1$. A contrario, on pourrait choisir de n'utiliser que la dernière interaction. Dans ce cas, si j est honnête et qu'il commet

une seule erreur, i le considèrera comme malhonnête. Par conséquent, il semble que la meilleure solution soit de dire que i doit se baser sur un nombre d'interactions antérieures égal à θ, avec $\theta > 1$ et θ borné par un seuil θ_{max}, pour éviter au maximum les effets de bord. Dans le modèle actuel, nous avons paramétré θ_{max} à 10. Cette expérience personnelle est conservée par chacun dans ce que nous appelons leur liste de notes de confiance antérieures.

- **La liste de confiance**

L'autre critère indispensable correspond aux "expériences extérieures". Il s'agit des connaissances des autres nœuds avec lesquels i a déjà interagi dans le passé. Pour mettre en place l'utilisation de ces expériences extérieures, les nœuds diffusent leurs connaissances en transférant à chaque interaction les notes de confiance qu'ils possèdent des nœuds qu'ils connaissent au nœud avec lequel ils communiquent et en récupérant, de la même manière, de celui-ci sa liste des notes de confiance. Une fois récupérée, la liste des notes de confiance est appelée la liste de confiance.

- **Les critères adjacents**

À ces trois critères "prioritaires" s'ajoutent trois autres attributs qui permet-tent d'affiner les relations de confiance entre les nœuds. Tout d'abord, à chaque note de confiance correspond un indice de confiance. Il sert principalement à permettre la distinction entre deux notes de confiance identiques. Un nœud A pourra donc distinguer la qualité d'un nœud B de celle d'un nœud C même s'il a une note de confiance identique pour ces derniers. Ceci est dû au fait de l'utilisation de l'historique des interactions précédentes. En effet, si l'on fixe l'hypothèse que l'historique ne conserve que les cinq dernières interactions et que les interactions sont notées sur une échelle de 0 à 1 (0 pour les mauvais échanges, 1 pour les bons, $0,5$ le seuil à partir duquel on dit qu'un nœud fait confiance à un autre), on peut estimer qu'un nœud ayant toujours reçu la même note $0,7$ et étant donc évalué à $0,7$ est plus propice à l'exécution d'un nouvel échange qu'un nœud évalué lui-aussi à $0,7$ dont l'historique, par exemple $\{1; 0,2; 0,8; 1; 0,5\}$, montre qu'il a déjà été "malhonnête".

De même, il existe un indice de confiance pour la liste de confiance. Cet indice permet à un nœud de déterminer si les listes de confiance qu'il reçoit sont crédibles ou non.

Enfin, à chaque diffusion de sa liste de notes de confiance, chaque nœud envoie aussi sa liste d'indices de confiance en ces notes de confiance.

Notation 2.3 *Le nœud i possède par conséquent six attributs relatifs au nœud j.*

la note de confiance	$V_i[j].NC \in [0,1]$
l'indice de confiance en cette note de confiance	$V_i[j].NCNC \in [0,1]$
la liste des notes de confiance antérieures	$V_i[j].LNCA$ *avec* $V_i[j].LNCA[k] \in [0,1]$ *et* $k \in [0, \theta_{max}]$
la liste des notes de confiance de j	$V_i[j].LC$ *avec* $V_i[j].LC[k] \in [0,1]$ *et* $k \in [0,n]$
la liste des indices de confiance que j possède en ses notes de confiance	$V_i[j].ILC$ *avec* $V_i[j].ILC[k] \in [0,1]$ *et* $k \in [0,n]$
l'indice de confiance en la liste de notes de confiance de j	$V_i[j].NCLC \in [0 \dots 1]$

2.2 La gestion de sa dynamique

2.2.1 Les premiers pas

Avant de présenter la méthode d'initialisation du modèle, il est important de noter que les attributs de confiance donnés ont tous des valeurs réelles comprises entre 0 et 1. Seules les interactions sont évaluées de manière binaire. Elles prennent la valeur 0 si l'échange n'est pas profitable et 1 s'il s'avère bénéfique pour le demandeur. Les notes données aux interactions sont donc différenciées des mesures de confiance. Nous verrons plus tard qu'en réalité, le calcul des notes de confiance est, en partie, réalisé en fonction des notes données aux interactions. Nous verrons à ce propos comment nous parvenons à obtenir des mesures de confiance non binaires.

Notation 2.4 *Nous appelerons* **note négative** *un attribut de confiance dont la valeur est incluse dans* $[0; 0,5[$ *et* **note positive** *un attribut de confiance dont la valeur est incluse dans* $[0,5; 1]$.

Une fois le modèle de simulation du réseau créé, il est initialisé en fonction des connexions entre les nœuds. Ainsi, lorsque deux nœuds i et j sont voisins de premier degré, les deux cellules $\mathcal{M}_t(i,j)$ et $\mathcal{M}_t(j,i)$ (avec $t = 1$) du modèle de simulation sont initialisées. Dans ce cas, les notes de confiance, les indices de confiance en les notes de confiance et en les listes de confiance prennent la première valeur possible d'une note positive, à savoir $0,5$. Les autres attributs ne sont initialisés que lors de la première interaction de i et j, c'est-à-dire lorsqu'ils diffuseront les listes des notes de confiance et des indices de confiance en ces notes.

2.2.2 Le maintien des données et leur évolution

En dehors des listes de notes de confiance et d'indices en ces notes qui sont diffusées et récupérées à chaque interaction par les nœuds par simple copie à l'emplacement correspondant dans leur vecteur de confiance, les autres attributs sont mis à jour au fil du temps par des fonctions que nous allons expliciter dans cette partie. Les sous-parties qui suivent vont donc illustrer comment faire évoluer les attributs du système à la suite d'une interaction, c'est-à-dire comment transformer

l'état du système pour le faire passer de son état courant, à l'instant t, à un nouvel état valide à l'instant $t+1$. Notons que ces fonctions sont en constant affinage au vu des simulations réalisées dont nous parlerons dans la partie 3.

2.2.2.1 Les notes de confiance

Définition 2.5 *À l'instant t, un nœud i décide d'interagir avec un nœud j si et seulement s'il possède une note de confiance positive au sujet de j. Plus formellement, nous notons :*

$$Inter_t(i,j) = vrai \iff \mathcal{M}_t(i,j).NC \geq 0,5$$

Remarque 2.6 *Deux nœuds voisins i et j ont l'un à propos de l'autre des informations totalement indépendantes dans le sens où ce n'est pas parce que i décide d'interagir avec j que j décidera forcément d'interagir avec i.*

Les notes de confiance représentent le point critique du modèle puisque c'est uniquement à partir de celles-ci que les nœuds vont décider d'effectuer ou non un échange avec leurs voisins. En outre, ce sont ces notes qui forment la composante principale de la dynamique du modèle.

Nous avons vu, au début de la partie 2, que la relation de confiance du modèle est fondée sur les connaissances de chaque acteur du système. Plus clairement, lorsqu'un nœud A souhaite interagir avec un autre B, il va se fier non seulement à ce qu'il connait personnellement du nœud B en question mais aussi à ce que les autres nœuds avec lesquels il a déjà interagi pensent du nœud B. Il va donc évaluer ses propres connaissances en se basant sur la note donnée à l'echange venant d'avoir lieu, que nous noterons α, mais aussi sur les notes des θ dernières interactions, avec $\theta \leq \theta_{max}$, contenues dans sa liste de notes de confiance antérieures. La moyenne, notée PK pour *personal knowledges*, de ces $\theta+1$ notes est alors calculée :

$$PK_{t+1} = \frac{\alpha + \sum_{k=1}^{\theta} \mathcal{M}_t(i,j).LNCA[k]}{\theta+1}$$

Cette évaluation personnelle du nœud B n'est pas suffisante pour donner un sens au concept de diffusion de la confiance. C'est pourquoi nous ajoutons ici les connaissances des autres en qui A a notamment une forte confiance (80%) en les listes de confiance récupérées, qui ont une note négative au sujet de B et qui ont suffisament confiance en cette note négative (60%). C'est le procédé que nous avons choisi pour obtenir un protocole de confiance basé sur l'efficacité et non sur l'efficience. Là encore, une moyenne, notée EK pour *external knowledges*, est calculée :

$$EK_{t+1} = \frac{1}{\rho} \sum_{\substack{k=1 \\ S*}}^{n} \mathcal{M}_t(i,k).LC[j]$$

avec S^* l'ensemble des conditions paramétrables que $\mathcal{M}_t(i,k)$ doit respecter, fixé à :

$$S^* = \left\{ \begin{array}{c} (k \neq i,j) \\ (\mathcal{M}_t(i,k).NCLC > 0{,}8) \\ (\mathcal{M}_t(i,k).LC[j] < 0{,}5) \\ (\mathcal{M}_t(i,k).ILC[j] \geq 0{,}6) \end{array} \right\}$$

dans le protocole actuel et ρ le nombre de sommets respectant ces conditions. Le fait de ne se servir des connaissances des autres que lorsque ceux-ci ont un point de vue négatif du nœud avec qui l'interaction doit avoir lieu est une méthode utile pour se rendre compte plus rapidement de la malhonnêteté des nœuds que si on utilisait à la fois les avis négatifs et positifs. Par conséquent cette méthode permet de diffuser plus rapidement la perception de nœuds ne servant pas l'intérêt de la "communauté". Le protocole mis en place est donc effectivement un protocole efficace et non efficient.

La fonction permettant de faire évoluer les notes de confiance, c'est-à-dire de calculer la nouvelle valeur (au temps $t+1$) de la note de confiance qu'un nœud i a en un nœud j à partir de l'ancienne (au temps t), après l'exécution d'une nouvelle interaction est donc la suivante :

$$\mathcal{M}_{t+1}(i,j) = \frac{\varepsilon.[\theta \times IK_{t+1}] + \overbrace{\zeta.[\theta_{max} \times EK_{t+1}]}^{(E)}}{\varepsilon.\theta + \underbrace{\zeta.\theta_{max}}_{ssi(E)\neq 0}}$$

avec ε et ζ deux facteurs permettant de paramétrer le poids que l'on veut donner à la connaissance des autres par rapport à celle que possèdent les nœuds eux-mêmes. Ainsi, plus le facteur ζ sera grand par rapport au facteur ε et plus la perception de malhonnêteté sera rapide. Il convient donc de trouver le bon équilibre entre ces deux facteurs pour ne négliger aucun type de connaissance par rapport à l'autre. Dans le protocole actuel, ces paramètres ε et ζ sont respectivement fixés à 1 et 2.

Remarque 2.7 *Il est utile de constater que, en cas d'égalité de ces deux paramètres de pondération, les connaissances externes ont toujours un poids supérieur ou égal aux connaissances personnelles. Ceci se vérifie par le fait que θ est majoré par θ_{max} et s'illustre lors des θ_{max} premières interactions, c'est-à-dire lorsque la liste des notes des interactions précédentes n'est pas encore remplie. Il est raisonnable de penser que lorsque des nœuds ont peu d'informations au sujet d'un autre, ils s'en remettent plus facilement à l'avis des autres.*

Ce mode d'évolution des notes de confiance montre que le modèle de diffusion de confiance développé ne peut être moins performant (par rapport aux critères d'évaluation évoqués au début de la partie 2) qu'un modèle semblable n'implantant pas le principe de diffusion. En effet, dans le cas où une interaction entre un nœud X et un nœud Y est possible, lorsque X met à jour sa note de confiance en Y, il ne

peut qu'obtenir une note inférieure ou égale à celle issue de la seule prise en compte des interactions précédentes qu'il a effectuées avec Y. Ceci est la conséquence du fait qu'il ajoute à cette première note uniquement les notes des nœuds qui ont une opinion négative à propos de Y et en qui il a une forte confiance en la liste de confiance.

2.2.2.2 Les indices de confiance en les notes de confiance

L'évolution des indices de confiance en les notes de confiance est déterminée par deux attributs, la nouvelle note α donnée à l'interaction venant d'avoir lieu et la liste des notes des interactions antérieures. Ces indices de confiance sont donc des avis personnels sur la qualité de la note de confiance.

Notation 2.8 *E et σ désigneront l'espérance et l'écart-type de façon usuelle.*

Ainsi, après avoir calculé $E(\mathcal{M}_t(i,j).LNCA)$ et $\sigma(\mathcal{M}_t(i,j).LNCA)$, nous procédons en nous basant sur un seuil ω_{NCNC} et en vérifiant l'appartenance ou non de la nouvelle note α à l'intervalle de valeurs

$$]E(\mathcal{M}_t(i,j).LNCA) - \omega_{NCNC} \dots E(\mathcal{M}_t(i,j).LNCA) + \omega_{NCNC}[$$

En cas d'appartenance à cet intervalle, nous ferons croître l'indice, dans le cas contraire, l'indice de confiance sera dévalué ; dans les deux cas, ces mises à jour seront proportionnelles à $\sigma(\mathcal{M}_t(i,j).LNCA)$ comme le montre l'algorithme 2 de l'annexe B. Dans le modèle, le seuil ω_{NCNC} est fixé à $0,15$ pour que l'indice soit augmenté uniquement lorsque la note donnée à la dernière interaction est proche de la moyenne des notes de interactions précédentes et le paramètre η est fixé à $\frac{1}{5}$ pour dévaluer au maximum de $\frac{1}{5}$ cet indice de confiance.

2.2.2.3 Les indices de confiance en les listes de confiance

Les indices de confiance en les listes de confiance sont quant à eux mis à jour de manière différente. En effet, contrairement à ceux portant sur les notes de confiance, lorsqu'un nœud i vient d'interagir avec un nœud j, la mise à jour effectuée ne porte pas sur l'attribut $\mathcal{M}_{t+1}(i,j).NCLC$ mais sur tous les $\mathcal{M}_{t+1}(i,k).NCLC$, avec $k \in [1 \dots n]$ et $k \neq i,j$. En dehors de ce point, la méthode utilisée est proche de celle discutée dans la partie 2.2.2.2 en raison de l'utilisation d'un seuil d'évaluation, que nous noterons ω_{NCLC}, fixé à $0,2$ dans le modèle. Ainsi, nous vérifions l'appartenance ou non de chaque $\mathcal{M}_{t+1}(i,k).LC[j]$ à l'intervalle de valeurs

$$]\alpha - \omega_{NCLC} \dots \alpha + \omega_{NCLC}[$$

Si $\mathcal{M}_{t+1}(i,k).LC[j]$ appartient à cet intervalle, on augmente l'indice de confiance que porte i en la liste de confiance de k ; dans le cas contraire, on le réduit. Notons que le modèle permet actuellement à ce niveau de réduire plus rapidement l'indice que ce qu'il permet de l'accroître, et ceci toujours par souci de meilleure perception de la malhonnêteté et d'efficacité du protocole. En effet, dans le modèle actuel, une augmentation ne peut dépasser $0,1$ alors qu'une baisse est forcément comprise dans $[0,1;0,2]$ grâce au paramètre ξ comme l'illustre l'algorithme 3 de l'annexe B.

3 Les simulations et l'analyse de leurs résultats

Afin de valider ou d'invalider le modèle présenté dans la partie 2, une phase de simulations s'avère indispensable. L'objectif principal de cette étape est de montrer les avantages du modèle avec diffusion des connaissances que nous avons développé par rapport à un autre n'utilisant pas ce concept de diffusion. Plus précisément, notre but est de réussir à évaluer les capacités d'un tel modèle et de les comparer avec celles observées dans un modèle sans diffusion de confiance. Pour ce faire, un simulateur a été implanté.

Cette partie vise à énoncer brièvement les possibilités offertes par ce simulateur puis à présenter en détails les diverses simulations réalisées et les résultats obtenus.

3.1 Le simulateur

Créer un logiciel totalement paramétrable conforme aux besoins définis dans le cahier des charges a été un point clé de cette phase de simulations. En effet, ce simulateur a permis de faire évoluer et d'affiner le modèle au fur et à mesure de ses exécutions.

Tout d'abord, notons qu'à chaque lancement, le logiciel exécute deux simulations similaires en tout point sur deux modèles différents, l'une uniquement basée sur les connaissances personnelles des nœuds et l'autre implantant le concept de diffusion. Ceci permet d'effectuer une comparaison précise des deux modèles sur chaque type de tests. Les résultats fournis sont représentés sous la forme de graphiques de manière à faciliter leur lecture. Le procédé de lecture de ces graphiques est présenté en annexes à la page *iv*.

Il permet aussi d'effectuer des tests sur trois différents types de graphes qui sont les principaux à nous intéresser : les graphes complets (réseaux en clique), les graphes à distribution de degrés fixée suivant une loi de puissance et les graphes à distribution de degrés fixée suivant une loi de puissance ayant subi une augmentation du coefficient de *clustering* (les plus proches des réseaux sociaux).

Par ailleurs, il permet d'intégrer dans le réseau des nœuds honnêtes et malhonnêtes mais aussi des nœuds ayant des opinions totalement erronées sur les autres (i.e. une opinion générée aléatoirement sans tenir compte de ses propres connaissances ni de celles des autres). De plus, il offre la possibilité de créer des groupes en raison de la forte présence de communautés dans les réseaux sociaux. Lorsque des nœuds d'un même groupe interagissent ensemble, ils se transmettent des données correctes. Lorsqu'ils interagissent avec des nœuds extérieurs, ils transmettent des données en fonction de leur nature première (honnête ou malhonnête). En revanche, lorsqu'ils diffusent leurs connaissances aux nœuds extérieurs, leur objectif est de leur faire croire en l'honnêteté des membres du groupe auquel ils appartiennent.

3.2 Les simulations et l'analyse de leurs résultats

Cette partie présente les premiers résultats que nous avons pu observer. Nous verrons en quoi ces simulations corroborent ou pas les attentes que nous avions en commençant ce stage. Pour ce faire, nous distinguerons les premières simulations réalisées, qui ont eu pour objectif de valider la cohérence du modèle, des autres dans lesquelles nous avons testé la résistance du modèle à des attaques particulières.

Remarque 3.1 *Notons que le modèle de simulation ne permet pas d'envisager des exécutions mettant en œuvre des milliers ou des dizaines de milliers de nœuds en raison de sa complexité mémoire. En conséquence, les simulations présentées dans la suite de ce rapport simulent toutes un million d'interactions ayant lieu sur un réseau de cent nœuds.*

Remarque 3.2 *Nous avons pour le moment décidé que les probabilités que les nœuds effectuent de mauvais échanges étaient paramétrables. En revanche, actuellement, nous posons la contrainte que tous les nœuds de même nature ont une probabilité égale de fournir de mauvaises informations.*

Le premier résultat que nous avons obtenu est que, si les interactions réalisées forment à la fin un graphe à distribution de degrés fixée suivant une loi de puissance sans augmentation du coefficient de *clustering*, le nombre de liens entre les nœuds est trop faible pour que la diffusion soit performante. Par conséquent, le protocole de confiance n'est pas utile. Dans la suite de cette partie, nous ne présenterons donc pas de simulations dans ce contexte.

3.2.1 La cohérence générale du protocole

Notation 3.3 *Nous dirons que le modèle est* **cohérent** *lorsque, en l'absence de nœud malhonnête, les nœuds se rendent compte de cette absence et ont confiance les uns en les autres et lorsque, lors de la présence d'un unique nœud malhonnête, celui-ci est repéré par l'ensemble des autres acteurs du système.*

La cohérence du protocole est le premier point qu'il était indispensable de tester avant de s'intéresser à des caractéristiques plus particulières émanant de la diffusion de la confiance. En effet, il était avant tout essentiel de valider le comportement général du protocole, c'est-à-dire la manière dont il évolue lors de la présence d'une très faible proportion de nœuds malhonnêtes.

Nous nous sommes ensuite attachés à montrer que le comportement du modèle de diffusion de confiance élaboré au cours de ce stage était valide pour les réseaux en clique lors de la présence d'un unique nœud malhonnête grâce aux simulations A (cf. figure 2) et B (cf. annexe C), dans lesquels les probabilités que le nœud malhonnête livre de mauvaises informations sont respectivement 80% et 20%. Les résultats illustrent de nettes différences entre les deux modèles, notamment lorsque le nœud malhonnête possède une faible probabilité de mal interagir avec les autres acteurs

du système. Tout d'abord, le modèle diffusant la confiance permet à tous les nœuds honnêtes de repérer les malhonnêtes ayant un faible taux d'erreur, ce que ne permet pas le modèle n'implantant pas la diffusion. De plus, la vitesse de perception de malhonnêteté est elle aussi plus rapide. En effet, par le calcul, nous remarquons que le nombre total de mauvaises interactions nécessaires que le nœud malhonnête (ayant 20% de chance de mal interagir) transmet pour qu'il soit considéré comme tel est de 45 pour le modèle avec diffusion et de 1343 pour le modèle sans diffusion (cf. simulation B). Ce nombre de 45 interactions est très intéressant puisqu'il permet de voir que, comme il existe 99 nœuds honnêtes, certains n'ont pas besoin de recevoir une mauvaise information du malhonnête pour se faire une opinion de sa nature. Le comportement du modèle avec diffusion est donc validé pour les réseaux en clique et s'avère plus performant que celui du modèle sans diffusion.

Les simulations C et D de l'annexe C montrent quant à elles les résultats obtenus sur des graphes à distribution de degrés fixée suivant une loi de puissance ayant subi une augmentation du coefficient de *clustering* de 33% [2]. La première place un nœud malhonnête sur le nœud de plus haut degré et la seconde sur le nœud de plus bas degré. Dans ces deux simulations, afin d'obtenir des différences les plus significatives possibles entre les deux modèles, les nœuds malhonnêtes ont chacun une probabilité de 20% de ne pas satisfaire les nœuds honnêtes [3]. Les résultats obtenus correspondent aux attentes. D'une part, lorsque le nœud malhonnête est placé sur le nœud de plus haut degré, le fait qu'il soit connecté à plusieurs autres nœuds permet une bonne diffusion des mesures de confiance et on note ainsi une grande performance du protocole de diffusion par rapport à l'autre. Le nœud malhonnête est repéré au terme de 14 mauvaises interactions au lieu de 472 dans le modèle sans diffusion. D'autre part, si le nœud de plus bas degré est malhonnête, alors, comme nous nous en doutions, l'intérêt du modèle de diffusion de la confiance ne s'illustre pas en raison de la connexion du nœud malhonnête avec un unique nœud honnête. La diffusion de confiance est par conséquent dans ce cas précis inutile. Le comportement du protocole de diffusion de confiance étudié est donc validé pour le modèle de réseaux sociaux utilisé même si sa performance est dépendante de la topologie.

Nous avons donc pu voir que la perception de malhonnêteté est plus ou moins rapide en fonction de la probabilité qu'ont les nœuds malhonnêtes de fournir de mauvaises informations. Par ailleurs, les comportements du modèle ne sont pas équivalents selon que le réseau est un graphe complet ou un graphe non complet. La particularité des graphes non complets est que, selon la place des nœuds, les informations qu'ils diffusent vont être plus ou moins vite perçues par les autres. En effet, si un nœud malhonnête est entouré de nœuds de haut degré, la diffusion des opinions de ses voisins sera plus importante que s'il est entouré de nœuds de bas degré. Par ailleurs, lorsqu'un nœud malhonnête est de degré maximum, il doit

[2] Dans la suite du rapport, le facteur d'augmentation du coefficient de *clustering* a été toujours été fixé à 33%

[3] Dans la suite de ce rapport, les résultats évoqués correspondent d'ailleurs tous à des simulations dans lesquelles les nœuds malhonnêtes ont une faible probabilité de fournir de mauvaises informations (20%).

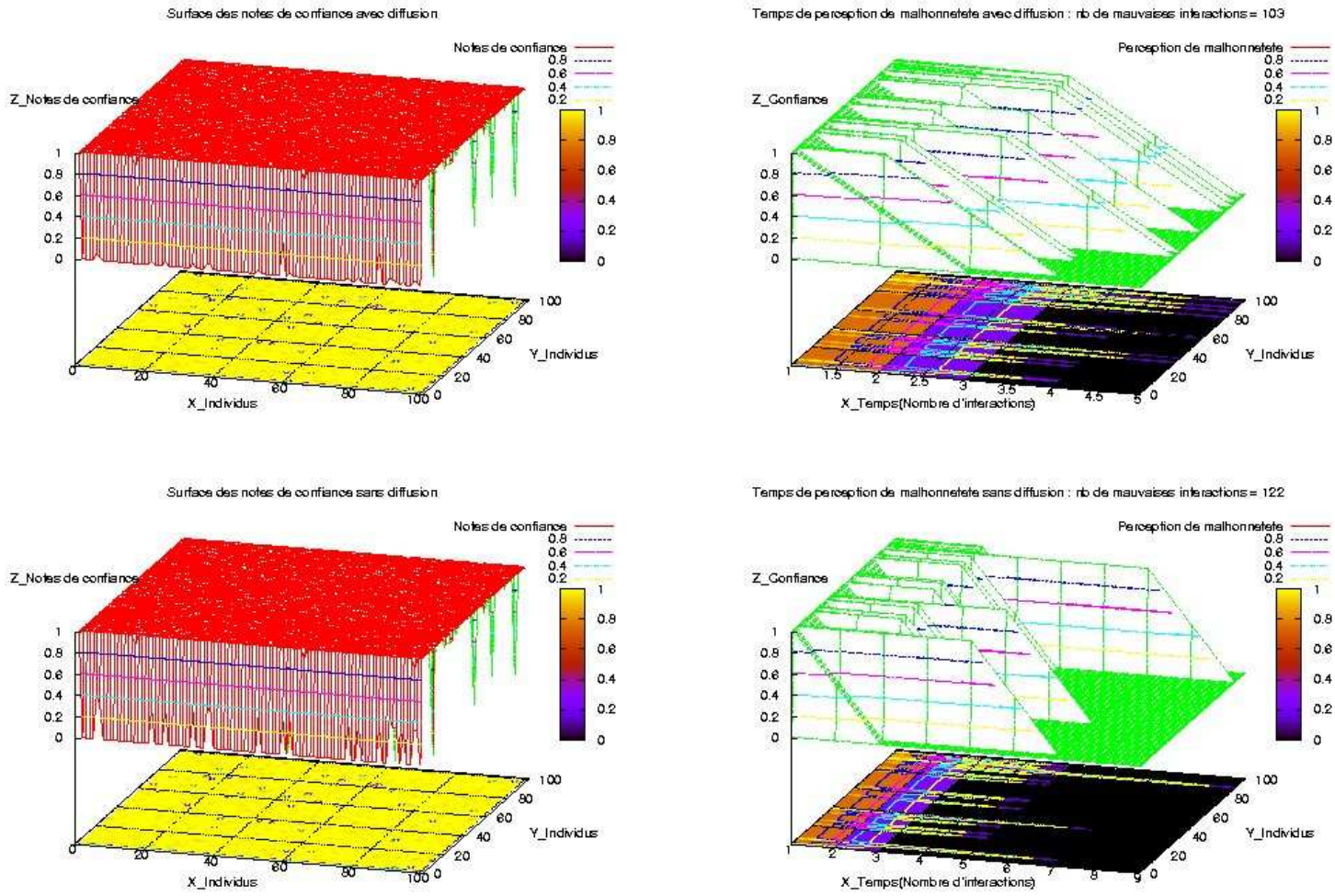


FIG. 2 – Simulation A : simulations réalisées sur un graphe complet de 100 nœuds. Seul l'individu 1 est malhonnête. Il possède 80% de chance de faire une mauvaise interaction.

être perçu comme tel par un grand nombre de nœuds. Il est donc logique que l'on puisse observer des différences significatives dans les résultats obtenus avec le protocole de diffusion et celui sans diffusion. En effet, avec diffusion des connaissances, les expériences extérieures sur lesquelles se base un nœud pour affecter une mesure de confiance ont rapidement du poids car les connaissances des autres sont reçues de manière régulière et fréquente. Ainsi, si le nœud malhonnête possède une faible probabilité d'effectuer de mauvaises interactions, il y a de fortes chances pour que les expériences extérieures signalent plus rapidement aux nœuds honnêtes la malhonnêteté du nœud en question que les expériences personnelles. En revanche, si le nœud

malhonnête est un nœud de bas degré, la rapidité de perception de malhonnêteté inhérente aux expériences personnelles est du même ordre que celle provenant des expériences extérieures, ce qui empêche de constater des différences notoires entre les deux modèles. Ceci s'explique par la faiblesse de la diffusion des connaissances induite par le faible degré du nœud malhonnête. Nous pouvons donc affirmer que l'intérêt de la diffusion des connaissances permettant de distinguer les nœuds malhonnêtes des nœuds honnêtes est fonction du degré des nœuds malhonnêtes. Par conséquent, la topologie du réseau, et particulièrement la place des nœuds malhonnêtes, sont deux critères fondamentaux qui amènent à des variations conséquentes des résultats. Nous parlerons plus après de **topologie-dépendance**.

3.2.2 La résistance aux attaques

Une fois la cohérence du modèle vérifiée, nous nous sommes intéressés à sa résistance aux attaques concertées. Dans un premier temps, cette partie présente le comportement du modèle de diffusion lors d'attaques de groupe de nœuds exclusivement malhonnêtes puis illustre ses réactions face à des attaques plus "intelligentes".

3.2.2.1 Les coalitions "simples"

L'objectif de cette partie est double : montrer la performance du protocole de diffusion face à une attaque d'un groupe de nœuds malhonnêtes de taille réaliste (on estime que, de manière générale, un réseau réel ne comporte pas plus de 10% de nœuds malhonnêtes) et montrer à partir de quelle taille un groupe malhonnête réussit à destabiliser le système. .

Les deux simulations E (cf. figure 3) et F (cf. annexe D), l'une sur un graphe complet et l'autre sur un graphe à distribution de degrés fixée avec augmentation du coefficient de *clustering*, réalisées ici pour comparer le comportement du modèle de diffusion de confiance à celui du modèle sans diffusion, ont mis en jeu un groupe [4] de dix individus malhonnêtes placés en fonction des degrés des nœuds de manière à occuper aussi bien des nœuds de haut degré que des nœuds de bas degré dans le cas des graphes non-complets.

Au vu des résultats de la simulation E, le modèle de diffusion de confiance apporte une reconnaissance parfaite des nœuds malhonnêtes d'un même groupe pour un réseau en clique, contrairement au modèle n'implantant pas le concept de diffusion. Tous les nœuds honnêtes du système se rendent effectivement compte de la coalition car ils possèdent tous une note négative à propos des nœuds malhonnêtes. On explique la reconnaissance imparfaite de la malhonnêteté quand la diffusion n'est pas utilisée car, comme les nœuds malhonnêtes ont seulement 20% de chances d'effectuer des mauvaises interactions, certaines listes des notes de confiance antérieures sur lesquelles se basent les expériences personnelles des nœuds ne contiennent à aucun moment suffisamment de notes négatives pour que les nœuds honnêtes puissent conclure en la malhonnêteté des nœuds formant la coalition.

[4] Les caractéristiques des groupes sont exposées dans le dernier paragraphe de la partie 3.1.

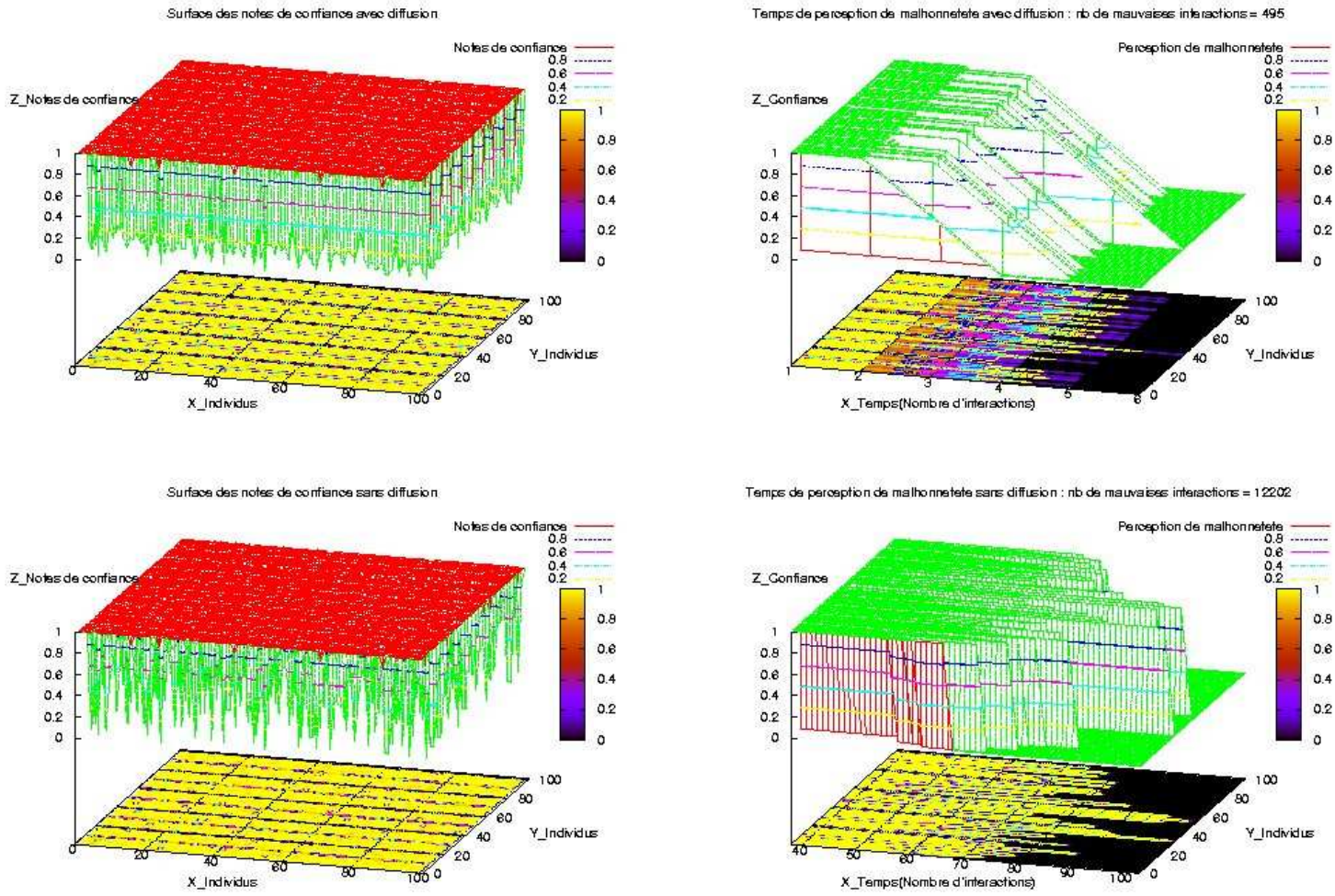


FIG. 3 – Simulation E : simulations réalisées sur un graphe complet de 100 nœuds mettant en œuvre un groupe de 10 individus malhonnêtes possédant 20% de chance de faire une mauvaise interaction. Ce groupe est composé des individus 10, 20, 30, 40, 50, 60, 70, 80, 90 et 100.

Par ailleurs, une fois encore, la vitesse de perception de malhonnêteté est plus rapide car les dix nœuds malhonnêtes n'effectuent en tout que 495 mauvais échanges plutôt que 12202 dans le cas où la diffusion des connaissances n'est pas activée. De plus, la simulation F a permis quant à elle de valider le protocole étudié sur les graphes à distribution de degrés fixée suivant une loi de puissance et ayant subi une augmentation du coefficient de *clustering* car le modèle de diffusion permet en l'occurrence un repérage de la malhonêteté en trois fois moins de mauvaises interactions. Ces différences de vitesses de perception de malhonnêteté des deux modèles

sont donc moins importantes mais sont la conséquence du positionnement des malhonnêtes sur des nœuds de degrés différents. La diffusion s'avère par conséquent moins rapide et les nœuds honnêtes doivent donc plus se baser, et ce pendant plus longtemps, sur leurs propres avis avant que les expériences extérieures leur permettent de se rendre compte de la malhonnêteté.

Ensuite, en nous intéressant au nombre de malhonnêtes devant se regrouper pour détourner le protocole de confiance de son comportement normal et prendre ainsi le contrôle du système, nous avons à nouveau constaté des divergences manifestes et avons ainsi pu vérifier l'intérêt du modèle. Tout d'abord, grâce à l'ensemble des simulations évoquées précédemment, il faut noter que, dans le modèle de diffusion, lorsqu'un seul nœud malhonnête est présent et essaie de perturber le système, il y parvient à partir du moment où il fait en sorte que son comportement se rapproche de celui d'un nœud honnête. En diminuant le nombre des mauvaises interactions qu'il réalise, il réussit à ne pas être reconnu par tous les nœuds du système. À ce propos, on a pu voir que ce n'était pas le cas en utilisant le protocole diffusant les mesures de confiance. Par conséquent, il devient intéressant d'évaluer quel pourcentage de la population les nœuds malhonnêtes regroupés en coalition doivent-ils occuper pour réussir à dérégler le comportement du modèle de diffusion. Nous avons pu remarquer que ce pourcentage se situe entre 60% (cf. simulation G) et 70% (cf. simulation H) pour les graphes complets. Pour les graphes à distribution de degrés fixée suivant une loi de puissance et ayant subi une augmentation du coefficient de *clustering* étudiés pour lesquels le protocole est topologie-dépendant, nous avons obtenus des résultats similaires. Toutefois, lorsque le système est composé de 90% de nœuds malhonnêtes appartenant à un même groupe, si ces derniers sont les nœuds de plus bas degrés, le système n'est pas perturbé et les nœuds honnêtes réussissent à repérer les malhonnêtes (cf. simulation I).

Cette perturbation du système sur une clique est due au fait que les nœuds malhonnêtes monopolisent les échanges et empêchent ainsi les nœuds honnêtes de suffisamment communiquer les uns avec les autres pour que la diffusion soit satisfaisante. En effet, ces derniers ne se transmettent que très peu leurs connaissances et ont par conséquent du mal à repérer rapidement la malhonnêteté des autres. Dans un réseau en clique de 100 nœuds, à chaque interaction, un arc du graphe a une chance sur 9900 d'être choisi (on différencie une demande d'interaction de i à j et de j à i). Lorsque trente nœuds seulement sont honnêtes, la probabilité qu'une interaction soit effectuée entre deux nœuds honnêtes est de $\frac{870}{9900} = 0.0878$. Lorsque quarante nœuds sont honnêtes, la probabilité qu'une interaction ait lieu entre deux nœuds honnêtes est multipliée par deux ($\frac{1560}{9900} = 0.1575$). Donc, dans le deuxième cas, la diffusion continue à avoir du poids car elle reste dans le même ordre de grandeur que la probabilité qu'ont les nœuds malhonnêtes à fournir de mauvaises informations mais, dans le premier cas, la diffusion devient négligeable.

Les coalitions simples sont donc plus faciles à réaliser quand le protocole de confiance n'utilise pas le principe de diffusion des connaissances. Elles ne représentent

en revanche pas de réel danger avec le protocole de diffusion de confiance tant que la proportion de nœuds malhonnêtes n'est pas fortement majoritaire. Cependant, on peut estimer que la durée de vie d'un réseau où plus de 60% des acteurs sont malhonnêtes est très faible et que son existence dans la réalité est très peu probable.

3.2.2.2 Des attaques plus précises

- **La perception d'un "cheval de Troie"**

Définition 3.4 *Un cheval de Troie est un programme informatique qui contient des fonctions cachées et qui s'exécute en toile de fond dès son arrivée sur un ordinateur à l'insu de son utilisateur.*

Une fois les résultats relatifs aux attaques simples analysés, nous avons cherché à valider le protocole sur des attaques plus complexes tout à fait envisageables dans les réseaux ad hoc comme l'existence d'un "cheval de Troie" par analogie aux virus informatiques. Nous appelons ici cheval de Troie un nœud réellement malhonnête protégé par d'autres nœuds du même groupe se faisant passer pour honnêtes auprès des nœuds réellement honnêtes. Le but des protecteurs du cheval de Troie est de toujours donner aux nœuds vraiment honnêtes des informations positives à propos du cheval de Troie. Les simulations réalisées ici se fondent sur l'hypothèse d'une population réaliste, à savoir que la quantité de nœuds malhonnêtes (ou protégeant un nœud malhonnête) ne dépasse pas 10% de l'ensemble des acteurs du système.

Dans un premier temps, les simulations ont porté sur des réseaux en clique. Nous avons ainsi pu remarquer que, malgré les fausses informations diffusées par les nœuds protégeant le cheval de Troie, le modèle étudié parvient à percevoir ce dernier comme un nœud malhonnête plus rapidement que le modèle sans diffusion, notamment lorsque les nœuds malhonnêtes ont un taux d'erreur inférieur à 60% comme le montre la simulation J de la figure 4. Ainsi, le fait que les protecteurs du cheval de Troie diffusent à son sujet de mauvaises informations aux nœuds honnêtes ne permet pas de les destabiliser car le modèle tel qu'il a été défini (i.e. efficace et non efficient) fait en sorte que ce type de mensonges ne soit jamais pris en compte.

De plus, en traitant la rapidité de perception comme le nombre total de mauvaises interactions effectuées par le cheval de Troie, le modèle de diffusion est environ deux fois plus rapide que le modèle sans diffusion de confiance lorsque le nœud malhonnête a une probabilité de 60% de ne pas satisfaire les autres acteurs du système et environ trente fois plus rapide lorsque cette probabilité est égale à 20% (cf. simulation K). Une telle probabilité d'effectuer de mauvaises interactions implique que le cheval de Troie n'est pas reconnu par la totalité des nœuds honnêtes du système lorsqu'aucune diffusion n'est réalisée, ce qui montre une nouvelle fois l'intérêt qu'ont les nœuds à diffuser leurs connaissances comme le préconise le modèle étudié.

En ce qui concerne les graphes à distribution de degrés fixée suivant une loi de puissance et ayant subi une augmentation du coefficient de *clustering*, les résultats correspondent aux attentes que nous en avions malgré la topologie-dépendance.

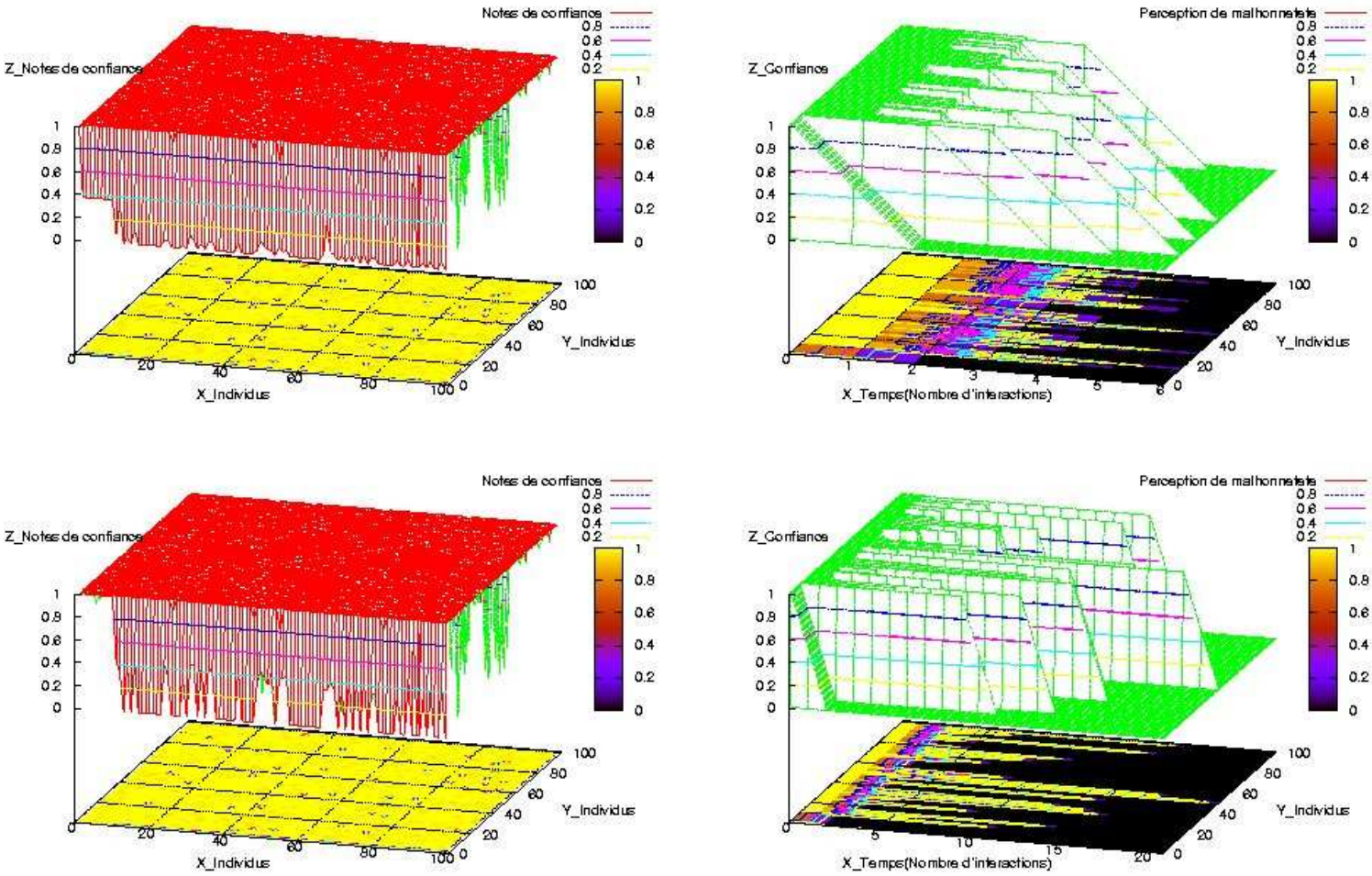


FIG. 4 – Simulation J : simulations réalisées sur un graphe complet de 100 nœuds regroupant un cheval de Troie et 9 individus le protégeant en diffusant aux individus honnêtes de mauvaises informations à son sujet. Le cheval de Troie est positionné sur l'individu 1, appartient au groupe composé des personnes $20, 30, 40 \ldots 100$ et possède 60% de chances de faire une mauvaise interaction.

Selon la place du cheval de Troie, mais aussi celle de ses "protecteurs", les résultats diffèrent. Cependant, tous permettent de valider la cohérence et la performance du modèle de diffusion de confiance car les résultats obtenus avec celui-ci sont nettement meilleurs que ceux découlant de l'utilisation du protocole sans diffusion. À titre d'exemple, la simulation L montre comment réagissent les nœuds honnêtes d'un réseau social face à un cheval de Troie placé sur le cinquième nœud de plus haut degré et protégé par neuf autres nœuds, à savoir $15, 25, 35 \ldots 95$. La simulation M, quant à elle, illustre la présence d'un cheval de Troie sur le nœud de plus haut degré.

Dans les deux cas, les simulations présentent des différences significatives entre les deux modèles. En effet, les perceptions du cheval de Troie sont au moins deux fois plus rapides lors de l'utilisation du modèle de diffusion de confiance. La simulation L montre en effet que, avec le modèle de diffusion de confiance, 269 mauvaises interactions sont effectuées par le nœud malhonnête avant que celui-ci soit repéré comme tel par la totalité des nœuds honnêtes alors que le modèle sans diffusion nécessite 635 mauvaises interactions. Ces différences ne sont pas aussi flagrantes que celles auxquelles on pouvait s'attendre. A contrario, lorsque le nœud réellement malhonnête est celui qui a le plus de voisins (cf. simulation M), on remarque des différences très importantes entre les deux modèles. Le modèle de diffusion de confiance permet une perception de malhonnêteté en seulement 19 mauvaises interactions alors que l'autre modèle en a besoin de 360, d'où une rapidité environ vingt fois supérieure du modèle de diffusion. En conséquence, le protocole étudié s'avère fiable et performant pour ce type d'attaque et la vitesse de perception s'avère du même ordre que celle observée lors de la présence d'un unique individu malhonnête dans un réseau social (cf. partie 3.2.1).

- **La réactivité face à une "bombe logique"**

Définition 3.5 *Une bombe logique est un programme informatique qui contient des fonctions cachées et qui se déclenche à un instant défini et s'exécute ensuite en toile de fond sur un ordinateur à l'insu de son utilisateur.*

Toujours par analogie aux virus informatiques, une autre attaque plus pernicieuse à envisager est la présence d'un nœud jouant le rôle d'une bombe logique. Contrairement à un cheval de Troie qui, par sa nature, est malhonnête de manière continue dès son arrivée sur le réseau, la difficulté émanant de la présence d'une bombe logique est que les autres acteurs du système vont avoir le temps de se faire une opinion sur celle-ci avant son déclenchement et vont donc la percevoir comme un nœud honnête. Notre objectif est donc de voir combien de mauvaises interactions effectuées par la bombe logique sont nécessaires après son déclenchement pour que les opinions positives que les nœuds du système possèdent à son sujet se transforment en opinions négatives. Nous appellerons ce nombre d'interactions le temps de réactivité.

Pour commencer, nous avons étudié le temps de réactivité du système face à une bombe logique sur un réseau en clique. Dans ces simulations, nous avons choisi de déclencher la bombe logique placée sur le nœud de plus haut degré et protégé par les nœuds $20, 30, 40 \ldots 100$ à la moitié de l'exécution, à savoir à la 500000^e interaction. À la vue des résultats, il semble que les modèles ne s'avèrent pas performants face à une telle attaque (cf. simulation N de la figure 5). En effet, 19 nœuds honnêtes parviennent à repérer la bombe avec le modèle de diffusion et seulement 12 y parviennent avec le modèle analogue n'offrant pas la diffusion.

Pourquoi les nœuds honnêtes ne parviennent pas tous à identifier la bombe logique ? Cela est dû au nombre trop petit d'interactions effectuées après le déclenchement de la bombe logique. Lorsqu'on déclenche le

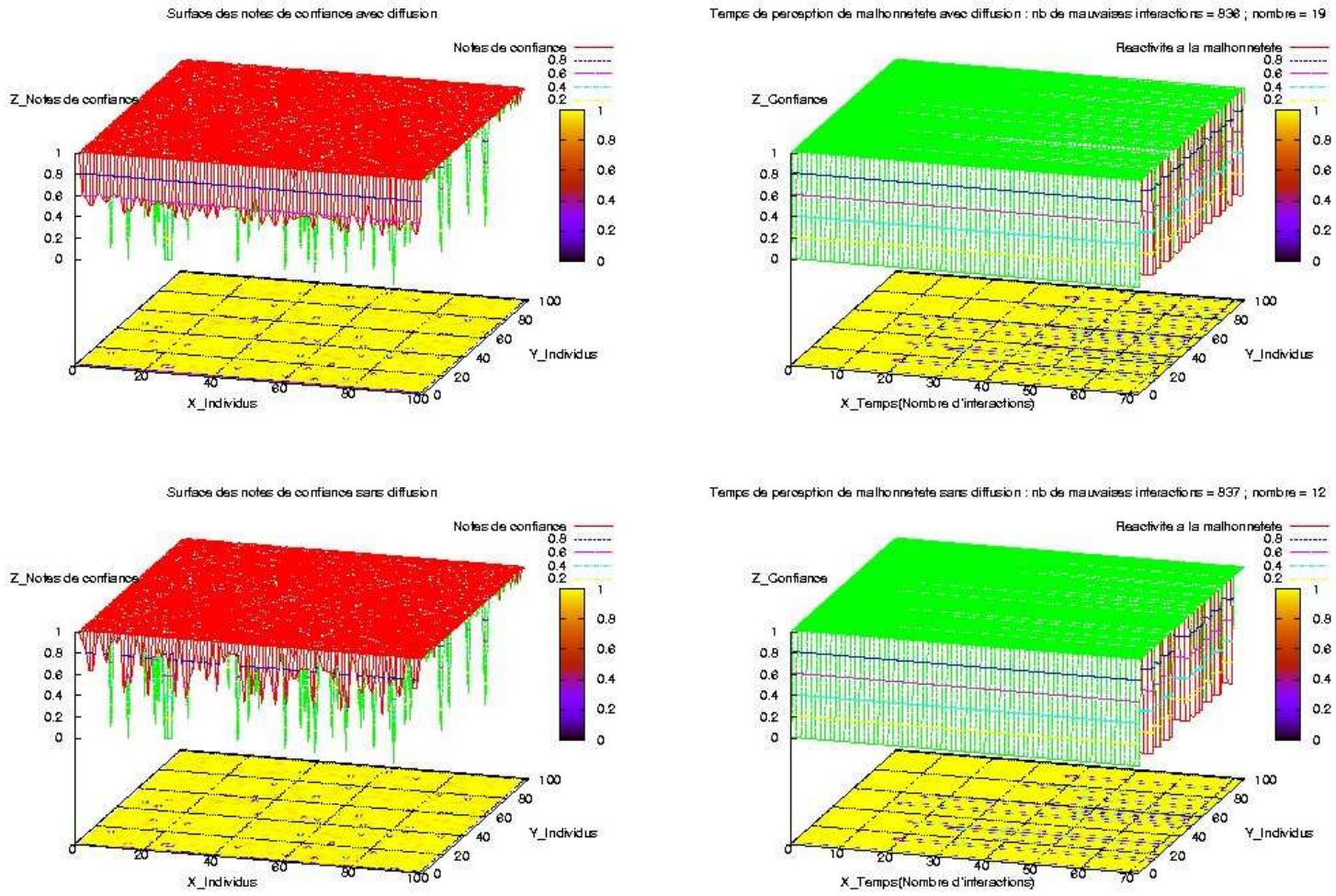


FIG. 5 – Simulation N : simulations réalisées sur un graphe complet de 100 nœuds regroupant une bombe logique et 9 individus la protégeant en diffusant aux individus honnêtes de mauvaises informations à son sujet. La bombe logique est positionnée sur l'individu 1 et appartient au groupe composé des personnes $20, 30, 40 \ldots 100$. Elle est déclenchée à la 500000^e interaction et possède 20% de chances de faire une mauvaise interaction.

comportement malhonnête de la bombe au bout de 500000 interactions, il ne reste plus que 500000 interactions avant la fin de la simulation. Or, les interactions sont choisies aléatoirement et il en existe $n(n-1)$ possibles, soit 9900 pour un système de cent nœuds, d'où par approximation environ 10000. Par conséquent, une interaction d'un nœud avec la bombe logique est potentiellement choisie cinquante fois, ce qui ne suffit pas pour reconnaitre la malhonnêteté lorsque la probabilité de la

bombe logique à fournir une mauvaise interaction est faible. Pour vérifier cette hypothèse, nous avons lancé une simulation en déclenchant la bombe logique au bout de 100000 interactions (cf. simulation O). Dans ce cas, le modèle sans diffusion permet à 27 nœuds d'identifier la bombe alors 75 nœuds du modèle avec diffusion la reconnaissent, ce qui valide l'hypothèse. Si l'on s'intéresse au nombre de mauvaises interactions réalisée par la bombe, on obtient des résultats intéressants. En effet, elle parvient à en donner 1051, soit en moyenne $11,68$ à chacun des individus. En tenant compte du fait que tous les nœuds honnêtes avaient, avant le déclenchement de la bombe, des avis positifs sur celle-ci et que la bombe a 20% de chances d'effectuer de mauvaises interactions, ces résultats sont satisfaisants. Le modèle est donc performant face à une attaque d'une bombe logique sur un réseau en clique.

Les résultats obtenus sur les graphes à distribution de degrés fixée suivant une loi de puissance et ayant subi une augmentation du coefficient de *clustering* (cf. simulation P) sont moins pertinents bien qu'ils permettent toutefois de valider le protocole. En effet, les résultats obtenus avec le protocole de diffusion sont avantageux par rapport à ceux observés avec l'autre modèle. Cependant, ils sont loin d'être performants. On remarque effectivement que, même si ses 15 voisins parviennet à la repérer, la bombe logique réussit à leur envoyer 419 fois des mauvaises données. Ses voisins reçoivent donc en moyenne $27,93$ mauvaises interactions, ce qui est environ trois fois plus que dans un réseau en clique. Ceci s'explique une fois encore par la topologie qui empêche les informations de se diffuser rapidement.

Conclusion et perspectives

Ces quatre mois et demi de stage passés au sein de l'équipe MC2 du LIP ont été une nouvelle occasion de travailler dans le domaine de la recherche informatique et notamment dans celui des systèmes complexes. Les recherches réalisées m'ont permis de cotoyer des chercheurs de divers horizons. En effet, j'ai pu assister aux groupes de travail organisés au Centre d'innovation en télécommunications et intégration de services (CITI) au cours desquels des sujets de tout ordre (informatique et mathématique, social, économique ou encore juridique) étaient discutés. Un projet comme KAA soulève de nombreuses questions dans ces différents domaines et les points de vue donnés par chacun de ses participants ne permettent pas d'y apporter des réponses triviales. Réussir à satisfaire tous les acteurs du projet n'est pas chose facile mais c'est pourtant l'un des objectifs de KAA et plus généralement de l'étude des systèmes complexes.

Le protocole de confiance que nous avons étudié consiste à bien intégrer le fait qu'une entité du système peut avoir confiance en une autre sans pour autant faire confiance à son avis qu'elle a sur les autres. En outre, il vise à réaliser des mesures de confiance selon deux critères. Le premier est la propre connaissance qu'un nœud X a en celui avec qui il souhaite interagir (Y) et le second est l'ensemble des connaissances qu'il a reçues des autres ayant déjà interagi avec Y et en qui il a confiance en l'avis. Cette double évaluation permet d'affiner les mesures de confiance et empêche par ailleurs à un nœud malhonnête de s'en prendre de manière répétée aux acteurs du système car ces derniers s'inviteront mutuellement à ne plus lui faire confiance en diffusant leurs connaissances. En ce sens, ce protocole de confiance est fondé sur l'efficacité plus que sur l'efficience.

Ce travail montre que l'utilisation d'une telle diffusion peut présenter un intérêt. D'une part, la cohérence des mesures de confiance est garantie par le modèle. Nous avons montré qu'un système dans lequel les individus malhonnêtes sont en nombre très limité reste viable pour ceux qui souhaitent satisfaire leur entourage et qui restent donc intègres au cours du temps. D'autre part, les simulations mettant en jeu une grande proportion de malhonnêteté ont illustré une grande robustesse de ce modèle de diffusion de confiance, ce qui n'est pas le cas d'un modèle semblable mais ne permettant pas la divulgation des connaissances par les acteurs du système ; ceci est mis en exergue par sa résistance à une population composée d'autant d'individus honnêtes que d'individus malhonnêtes, même lorsque ces derniers décident de limiter leur malhonnêteté à vingt pour cent de leurs interactions. Ensuite, en nous intéressant à des attaques plus réfléchies comme la présence d'un individu jouant le rôle d'un cheval de Troie protégé par une communauté, nous avons réussi à valoriser les qualités inhérentes à ce concept de diffusion. Ces trois types d'attaques s'avèrent par conséquent sans réel danger pour les réseaux en clique et les réseaux sociaux (modélisés ici par les graphes à distribution de degrés fixée suivant une loi de puissance ayant subi une augmentation du coefficient de *clustering*). Néanmoins, l'existence d'une bombe logique protégée par un groupe sur un réseau social

reste actuellement une attaque envisageable par un groupe de nœuds souhaitant perturber le système. Malgré la fiabilité du protocole étudié face à une telle attaque sur un réseau en clique, il sera nécessaire de chercher comment parvenir à rendre le modèle robuste à cette attaque sur un réseau social.

Les recherches effectuées ont donc donné des résultats positifs dans l'ensemble et incitent à penser que les recherches futures auront de grandes chances d'aboutir à un protocole de diffusion de confiance fiable, robuste et performant pour les réseaux ad hoc. Une perspective intéressante est qu'un tel protocole de confiance pourrait s'appliquer non seulement aux réseaux ad hoc pour lesquels il a été étudié mais aussi aux autres réseaux informatiques. En effet, on peut penser que ce protocole pourrait avoir un grand intérêt dans Internet ou dans les réseaux P2P. Il pourrait par exemple servir à mesurer la confiance que les utilisateurs placent dans les sites et pourrait ainsi diffuser la reconnaissance de sites "mal-intentionnés" mettant implicitement à disposition des virus à télécharger. Cependant, avant de parvenir à créer un tel protocole, et ce malgré les qualités déjà observées sur le modèle que nous avons intégralement développé, de nombreuses améliorations sont à envisager. Nous en citerons trois qui nous paraissent être les principales.

La première à laquelle nous souhaitons nous intéresser est la réduction de la taille des messages transmis. La **"légèreté"** du protocole est un point primordial pour éviter que le réseau soit surchargé. Ainsi, l'objectif que nous voulons atteindre est de trouver le parfait équilibre permettant de réduire au maximum le nombre d'informations transmises entre les nœuds du réseau tout en conservant une diffusion des connaissances suffisante pour que le protocole reste fiable et robuste. Plusieurs solutions sont à examiner. Tout d'abord, comme le protocole est basé sur l'efficacité, nous pourrons limiter les transferts aux seules informations indiquant la malhonnêteté des nœuds. Cela ne posera pas de problème car, en dehors des connaissances personnelles, seules les notes négatives des nœuds extérieurs sont prises en compte lors du calcul d'une nouvelle note de confiance. Mais peut-être est-il possible de réduire encore plus la diffusion ? On peut imaginer que les nœuds pourraient uniquement diffuser un certain pourcentage des informations négatives. Ce pourcentage diminuerait avec l'augmentation de la taille du système. On peut aussi imaginer que les nœuds diffuseraient uniquement les informations concernant leur plus proches voisins.

La deuxième consisterait à faire varier l'ensemble des paramètres du protocole pour voir s'il est possible, par ces simples transformations, d'obtenir un protocole efficient et non plus efficace. De même, il semblerait intéressant de rendre le protocole à la fois efficace et efficient. En effet, parvenir à créer un protocole de confiance permettant d'assurer l'efficacité et l'efficience serait un point positif car combiner ces deux caractéristiques pour n'en garder que leurs avantages respectifs pourrait amener à un protocole de confiance d'autant plus fiable et performant.

Enfin, un dernier point qu'il serait intéressant d'étudier est de faire en sorte que l'indice en les listes des notes de confiance ne soit plus un indice général (i.e.

il mesure actuellement la confiance qu'un nœud possède au sujet de la liste dans son intégralité) mais un indice plus spécifique à chacun des nœuds contenus dans cette liste. Ce changement pourrait peut-être rendre le protocole plus performant ou, au contraire, peut-être permettrait-il à de nouvelles attaques de destabiliser le système ?

Bibliographie

[1] Site officiel de l'ACI "sécurité et informatique". http ://acisi.loria.fr/.

[2] Site officiel du projet KAA. http ://citi.insa-lyon.fr/kaa.

[3] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1) :47–97, 2002.

[4] L. Buttyan, S. Capkun, and J. Hubaux. The quest for security in mobile ad hoc networks. In *Proceeding of the* ACM *Symposium on Mobile Ad Hoc Networking and Computing*, 2001.

[5] P. Erdös and A. Rényi. On random graphs. *Publicationes Mathematicae*, 6 :290–297, 1959.

[6] L. Guihéry, V. Legrand, J. Morêt-Bailly, M. Morvan, J.-P. Neuville, J. Pousin, A. Rabagny, S. Ubéda, and F. Valois. Scientific description of the KAA project. ACI "Sécurité et Informatique", 2004.

[7] J. Kong, S. Lu, H. Luo, P. Zerfos, and L. Zhang. Providing robust and ubiquitous security support for mobile ad hoc networks. In *International Conference on Network Protocols*, pages 251–260, 2001.

[8] V. Legrand, F. Nait-Abdesselam, and S. Ubéda. Établissement de la confiance et réseaux ad hoc : un état de l'art. In *Sécurité et Architecture Réseaux*, 2003.

[9] S. Milgram. The small world problem. *Psychology Today*, 2 :60–67, 1967.

[10] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms*, 6 :161–179, 1995.

[11] M. E. J. Newman. Random graphs as models of networks. In S. Bornholdt and H. G. Schuster, editors, *Handbook of Graphs and Networks*, pages 35–68. Wiley-VCH, 2003.

[12] M. E. J. Newman. The structure and function of complex networks. *Reviews of Society for Industrial and Applied Mathematics*, 45(2) :167–256, 2003.

[13] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Physical Review E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, 64, 2001.

[14] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graph models of social networks. In *Proceedings of the National Academy of Sciences*, volume 99, 2002.

[15] S. H. Strogatz and D. J. Watts. Collective dynamics of 'small-world' networks. *Nature*, 393 :440–442, 1998.

[16] P. Tortelier. Les réseaux ad-hoc. France Télécom Recherche et Développement.

ANNEXES

A L'algorithme de création des graphes à distribution de degrés fixés suivant une loi de puissance

Dans [11], Newman présente l'algorithme de Molloy et Reed [10] pour engendrer de tels graphes. On commence par fixer une séquence de degrés. Plus formellement, à chaque sommet $i = 1 \dots n$, on affecte un degré spécifique δ_i en utilisant la distribution de degrés donnée dans la partie 1.3.3.

Une fois cette séquence calculée, la méthode à suivre est la suivante : tout d'abord, on donne à chaque sommet i un nombre δ_i de "talons" ou "demi-arêtes" (i.e. d'extrémités d'arêtes émergeant du sommet). Ensuite, on choisit ces talons uniformément et séquentiellement par paires sur des sommets différents. On relie ces paires pour obtenir des arêtes complètes. Cet algorithme est illustré dans la figure 6 et sera celui utilisé dans le modèle de diffusion de confiance étudié au cours de ce stage pour engendrer les graphes modélisant les réseaux sociaux. Comme nous allons le voir, cet algorithme est très intéressant car il ne favorise aucune solution par rapport à une autre.

Théorème A.1 *[10] Cet algorithme engendre* **uniformément** *des graphes à distribution de degrés fixée suivant une loi de puissance si le nombre total de demi-arêtes est pair.*

Ce théorème signifie que tous les graphes potentiellement engendrables à partir de la distribution de degrés fixée ont autant de chance les uns que les autres d'être engendrés par cet algorithme.

Remarque A.2 *En raison des $\delta_i!$ permutations possibles des talons émergeant du i^e sommet, il y a $\Pi_i \delta_i!$ différentes manières d'engendrer le graphe. Ce facteur est toutefois constant tant que la séquence de degrés est fixée.*

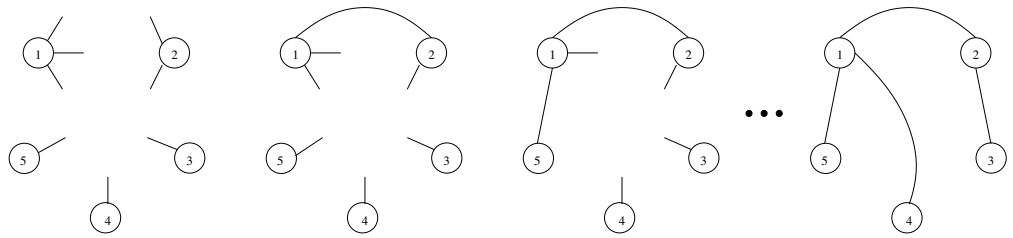


FIG. 6 – Exemple de génération d'un graphe aléatoire de cinq sommets suivant une loi de puissance avec la distribution de degrés $\{\delta_i\} = \{3, 2, 1, 1, 1\}$.

Remarque A.3 *Notons que, dans certains cas, les talons peuvent ne pas tous être reliés par paires de sommets différents et sont alors supprimés. On peut par exemple considérer le graphe de la figure 6 avec la distribution de degrés* $\{\delta_i\} = \{3, 2, 2, 1, 1\}$. *Par ailleurs, on empêche aussi la création d'arêtes multiples et de boucles reliant deux demi-arêtes d'un même nœud entre elles. Dans ces cas, on perd la propriété d'uniformité. Une méthode existe pour la conserver. Elle consiste à relancer tout l'algorithme autant de fois que nécessaire jusqu'à obtenir un graphe créé uniformément. Ce procédé étant très coûteux en temps, nous utilisons la méthode alternative visant à supprimer les arêtes restantes. Ceci n'est pas vraiment gênant car nous verrons que nous sommes dans tous les cas amenés à perdre l'uniformité pour se rapprocher au mieux des réseaux sociaux.*

Cet algorithme ne suffit pas pour obtenir un graphe s'approchant le plus possible des réseaux sociaux de la réalité. Il ne permet effectivement pas d'obtenir des graphes avec des coefficients de *clustering* tels qu'en présentent les réseaux sociaux.

Pour réussir à obtenir des graphes plus proches des réseaux sociaux, nous avons choisi la solution qui consiste à appliquer l'augmentation du coefficient de *clustering* sur le graphe à distribution de degrés fixée suivant une loi de puissance déjà créé en utilisant l'algorithme 1 qui vise à augmenter d'un facteur c le coefficient d'interconnectivité locale d'un graphe.

Algorithme 1 : Augmentation de facteur c du coefficient de clustering d'un graphe à distribution de degrés fixée suivant une loi de puissance $\mathcal{G}$.

```
début
    Soit 𝒢 un graphe;
    Soit c ∈ [0,1] le facteur d'augmentation du coefficient de clustering;
    pour chaque sommet s de 𝒢 faire
        pour chaque paire d'arête [sx],[sy] faire
            Soit tmp un nombre tiré aléatoirement entre 0 et 1;
            si c < tmp alors Créer l'arête [xy];
        fin
    fin
fin
```

Cette méthode ne conserve pas la distribution de degrés fixée. De plus, rien ne garantit que la distribution de degré finale reste uniforme parmi les graphes ayant une distribution de degrés fixée et un certain coefficient de *clustering*. Cependant, une méthode garantissant la conservation de la distribution de degré uniforme n'a pas encore été trouvée à ce jour. Il s'agit d'un problème ouvert relatif aux réseaux sociaux et le traiter n'était pas l'objet de ce stage. Toutefois, la solution proposée donne une méthode qui nous permet d'obtenir des graphes plus proches des réseaux sociaux tout en utilisant un facteur d'augmentation constant du coefficient de *clustering*.

B Les algorithmes de dynamique du modèle

Algorithme 2 : Dynamique des indices de confiance en les notes de confiance

début
  si $\alpha \in]E(\mathcal{M}_t(i,j).LNCA) - \omega_{NCNC} \dots E(\mathcal{M}_t(i,j).LNCA) + \omega_{NCNC}[$
  alors
    $X = (1 - \sigma(\mathcal{M}_t(i,j).LNCA)) \times (\omega_{NCNC} - |\alpha - E(\mathcal{M}_t(i,j).LNCA)|)$;
    $\mathcal{M}_{t+1}(i,j).NCNC = \mathcal{M}_t(i,j).NCNC + X$;
  sinon
    $X = (1 - \sigma(\mathcal{M}_t(i,j).LNCA)) \times \eta|\alpha - E(\mathcal{M}_t(i,j).LNCA)|$;
    $\mathcal{M}_{t+1}(i,j).NCNC = \mathcal{M}_t(i,j).NCNC - X$;
  fin
fin

Algorithme 3 : Dynamique des indices de confiance en les listes de confiance

début
  pour chaque k *tel que* $\mathcal{M}_{t+1}(i,k).LC[j]$ *existe* **faire**
    si $\mathcal{M}_{t+1}(i,k).LC[j] \in]\alpha - \omega_{NCLC} \dots \alpha + \omega_{NCLC}[$ **alors**
      $X = \omega_{NCLC} - (|\mathcal{M}_{t+1}(i,k).LC[j] - \alpha)|$;
      $\mathcal{M}_{t+1}(i,k).NCLC = \mathcal{M}_t(i,k).NCLC + X$;
    sinon
      si $|\mathcal{M}_{t+1}(i,k).LC[j] - \alpha)| == \omega_{NCLC}$ **alors**
        $\mathcal{M}_{t+1}(i,k).NCLC = \mathcal{M}_t(i,k).NCLC - 0,1$;
      sinon
        si $|\mathcal{M}_{t+1}(i,k).LC[j] - \alpha)| == 1$ **alors**
          $\mathcal{M}_{t+1}(i,k).NCLC = \mathcal{M}_t(i,k).NCLC - 0,2$;
        sinon
          $X = \xi|\mathcal{M}_{t+1}(i,k).LC[j] - \alpha - 0,1| - 0,1$;
          $\mathcal{M}_{t+1}(i,k).NCLC = \mathcal{M}_t(i,k).NCLC - X$;
        fin
      fin
    fin
  fin
fin

À la suite d'un tableau récapitulatif de toutes les simulations évoquées dans ce rapport (cf. page suivante), les annexes illustrent les résultats obtenus ayant permis de valider le modèle de diffusion de la confiance pour les réseaux ad hoc. Ces résultats proviennent directement du simulateur qui prend en charge leur gestion graphique par l'intermédiaire d'un appel au logiciel *Gnuplot* à la fin de la phase de calcul qui simule les interactions.

Ces simulations ont toutes été réalisées sur des réseaux (complets ou non) de 100 nœuds. De plus, elles ont toutes exécuté un million d'interactions.

Les résultats obtenus sont tous présentés de manière identique. Ainsi, en lisant ces graphiques, la partie du haut représente les résultats obtenus avec le modèle de diffusion de confiance et la partie du bas fournit ceux observés avec un modèle semblable à celui étudié mais n'implantant pas le concept de diffusion des connaissances. On retrouve deux types de représentations de surfaces : les surfaces des notes de confiance et les surfaces de temps de perception de malhonnêteté. Toutes ces représentations sont décomposées en deux zones distinctes pour faciliter leur lecture. La première zone illustre la surface qui correspond aux informations calculées par le simulateur ayant été données à *Gnuplot* pour en réaliser le tracé. Selon les simulations effectuées, les surfaces obtenues peuvent s'avérer peu lisibles. C'est pourquoi, dans une deuxième zone des représentations, à savoir sur le plan défini par ((0X)(0Y)), les courbes de niveau caractéristiques des surfaces ont été ajoutées.

Par ailleurs, dans le cas des graphes s'approchant des réseaux sociaux, les nœuds sont numérotés par degrés décroissants. Ainsi, le nœud de numéro 1 est celui de plus haut degré et celui de numéro 100 possède le plus petit nombre de voisins.

• Les surfaces de notes de confiance

Elles illustrent les notes de confiance que les nœuds situés sur l'axe (0X) possèdent au sujet des nœuds de l'axe (0Y). Les notes sont données par l'echelle de l'axe (0Z).

• Les surfaces de temps de perception de malhonnêteté

Dans le titre de ces représentations est donné le nombre total de mauvaises interactions données au nœuds honnêtes par les nœuds malhonnêtes. Pour les simulations d'attaque d'une bombe logique, un renseignement supplémentaire est donné, le nombre de nœuds honnêtes réussissant effectivement à repérer la bombe logique.

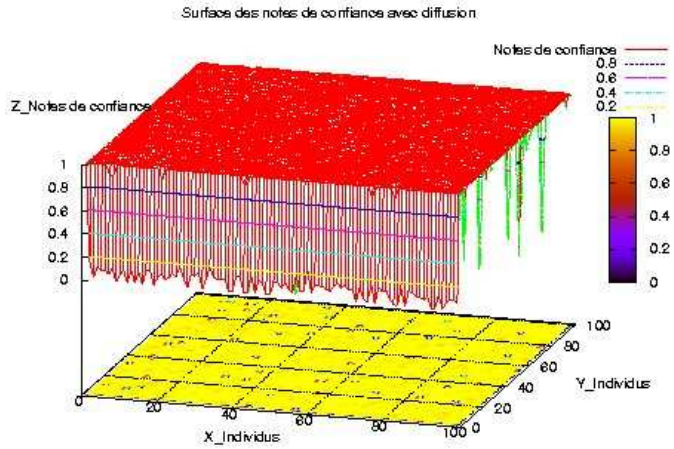
Les représentations graphiques indiquent le nombre d'interactions (bonnes ou mauvaises) moyen (cf. axe (0X)) que les nœuds honnêtes (axe (0Y)) effectuent avec les nœuds malhonnêtes avant de repérer leur malhonnêteté. Lorsqu'un nœud honnête a confiance, la valeur est de 1 sur l'axe (0Z) ; elle passe à 0 quand l'ensemble des nœuds malhonnêtes est perçu comme tel. Ainsi, si dix nœuds sont malhonnêtes et que la représentation montre que le nœud 50 les a repérés en 5 interactions, cela signifie qu'il a effectué en moyenne 5 interactions avec chacun des malhonnêtes avant de tous les considérer comme malhonnêtes.

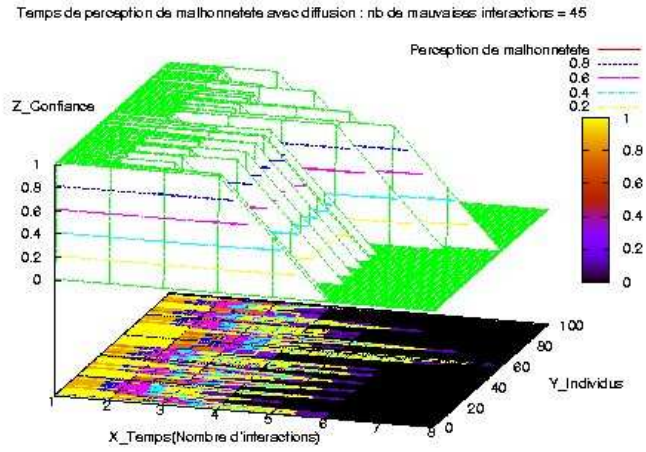
Simulation	Type de réseau	Taille du système	Nœuds malhonnêtes (pourcentage de mauvais échanges)	Groupes	Cheval de Troie	Bombe logique (déclenchement)
A	clique	100 nœuds	1 (80%)	-	-	-
B	clique	100 nœuds	1 (20%)	-	-	-
C	social	100 nœuds	1 (20%)	-	-	-
D	social	100 nœuds	100 (20%)	-	-	-
E	clique	100 nœuds	10, 20, 30, 40, 50, 60, 70, 80, 90, 100 (20%)	10, 20, 30, 40, 50, 60, 70, 80, 90, 100	-	-
F	social	100 nœuds	5, 15, 25, 35, 45, 55, 65, 75, 85, 95 (20%)	5, 15, 25, 35, 45, 55, 65, 75, 85, 95	-	-
G	clique	100 nœuds	41, 42, 43, ..., 99, 100 (20%)	41, 42, 43, ..., 99, 100	-	-
H	clique	100 nœuds	31, 32, 33, ..., 99, 100 (20%)	31, 32, 33, ..., 99, 100	-	-
I	social	100 nœuds	11, 12, 13, ..., 99, 100 (20%)	11, 12, 13, ..., 99, 100	-	-
J	clique	100 nœuds	1 (60%)	1, 20, 30, 40, 50, 60, 70, 80, 90, 100	1	-
K	clique	100 nœuds	1 (20%)	1, 20, 30, 40, 50, 60, 70, 80, 90, 100	1	-
L	social	100 nœuds	5 (20%)	5, 15, 25, 35, 45, 55, 65, 75, 85, 95	5	-
M	social	100 nœuds	1 (20%)	1, 20, 30, 40, 50, 60, 70, 80, 90, 100	1	-
N	clique	100 nœuds	1 (20%)	1, 20, 30, 40, 50, 60, 70, 80, 90, 100	-	1 (500000)
O	clique	100 nœuds	1 (20%)	1, 20, 30, 40, 50, 60, 70, 80, 90, 100	-	1 (100000)
P	social	100 nœuds	1 (20%)	1, 20, 30, 40, 50, 60, 70, 80, 90, 100	-	1 (100000)

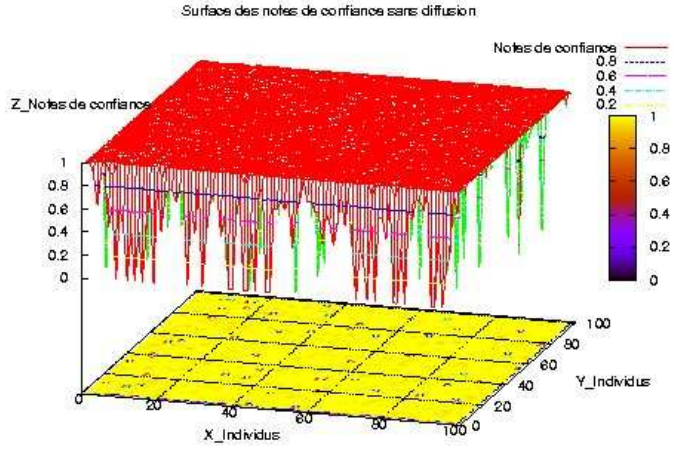
C La cohérence du modèle

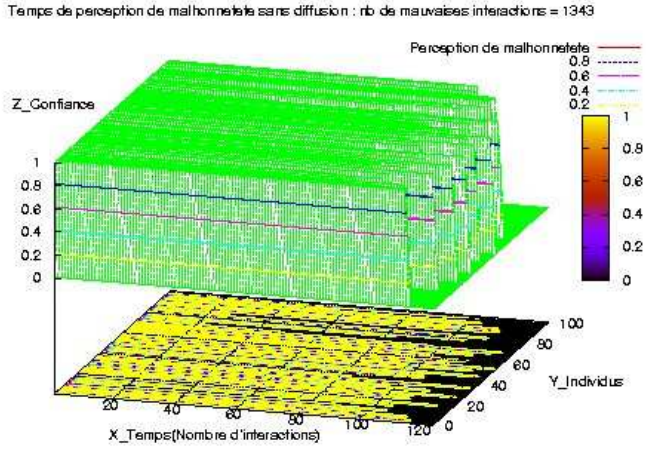
Simulation B

Simulations réalisées sur un graphe complet de 100 nœuds. Seul l'individu 1 est malhonnête. Il possède 20% de chance de faire une mauvaise interaction.



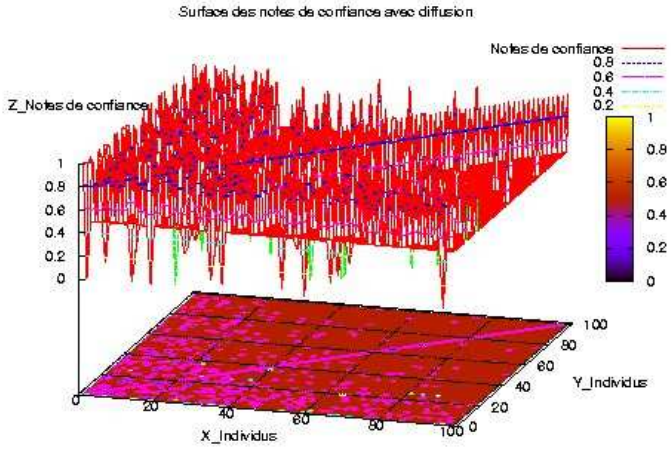


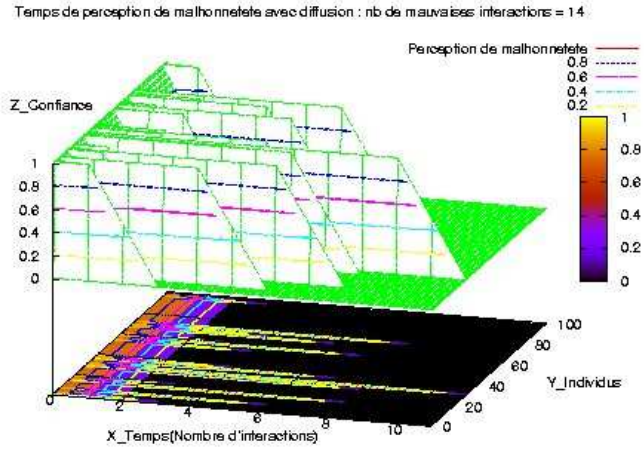


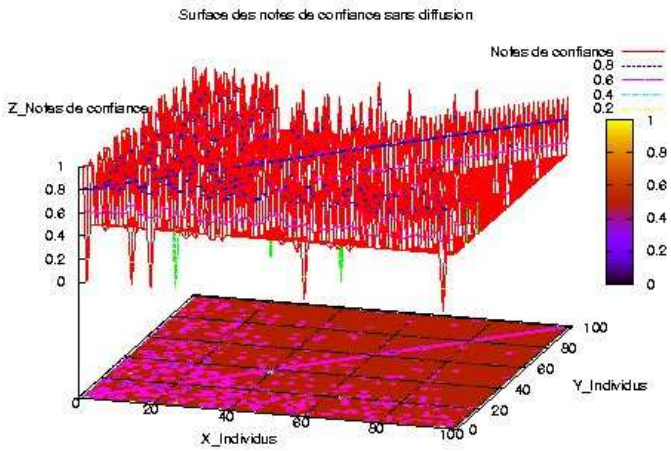


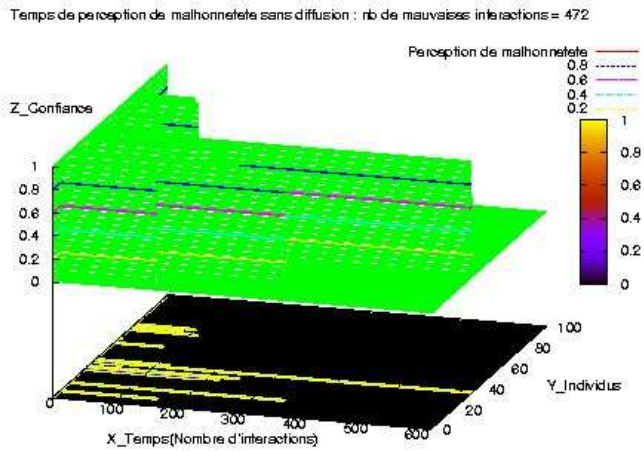
Simulation C

Simulations réalisées sur un graphe à distribution de degrés fixée suivant une loi de puissance ayant subi une augmentation du coefficient de *clustering* de 100 nœuds. Seul l'individu 1 est malhonnête. Il possède 20% de chance de faire une mauvaise interaction et correspond au nœud de plus haut degré du graphe.



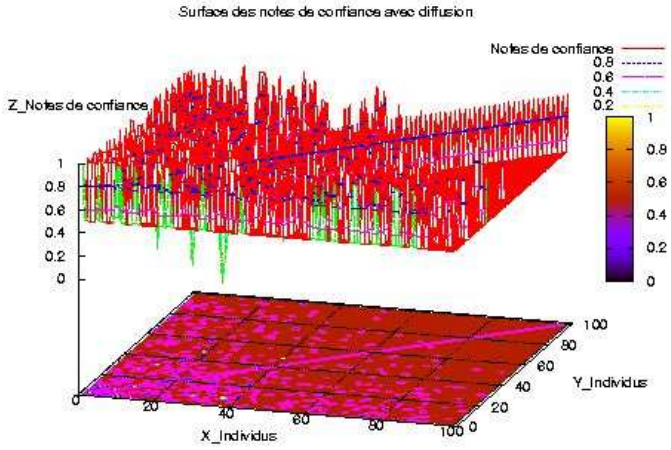


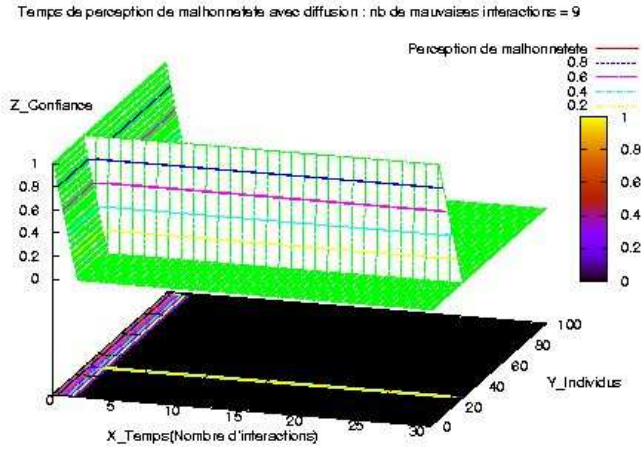


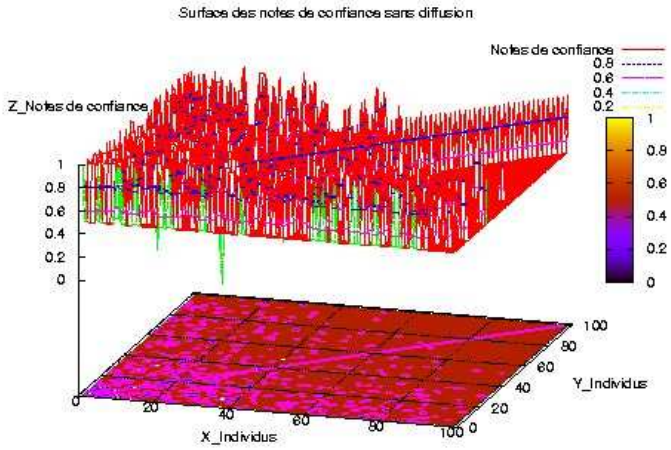


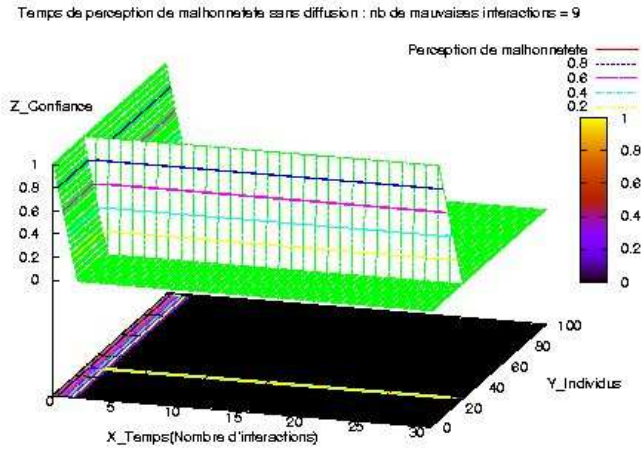
Simulation D

Simulations réalisées sur un graphe à distribution de degrés fixée suivant une loi de puissance ayant subi une augmentation du coefficient de *clustering* de 100 nœuds. Seul l'individu 100 est malhonnête. Il possède 20% de chance de faire une mauvaise interaction et correspond au nœud de plus bas degré du graphe.





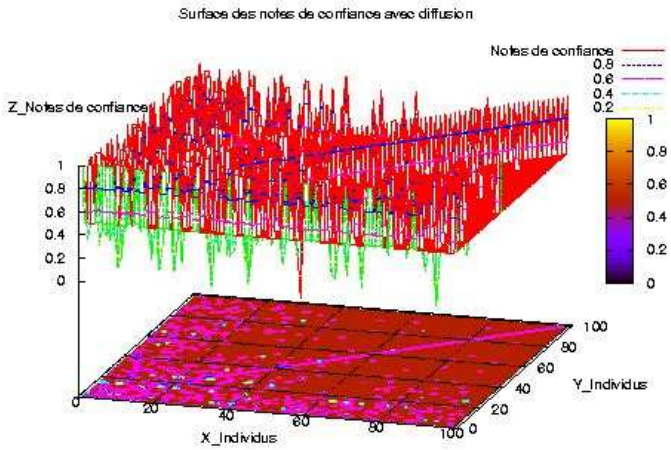


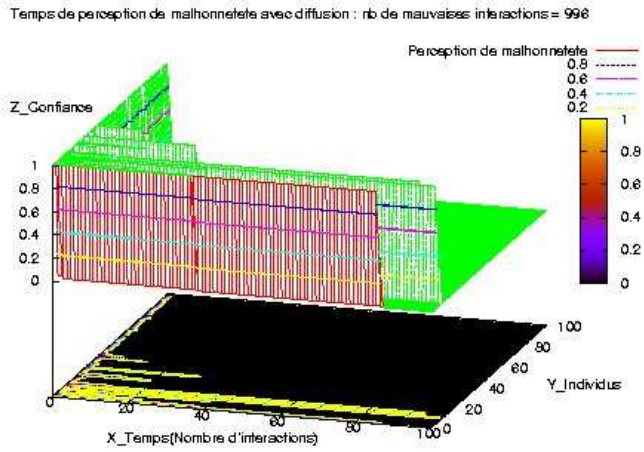


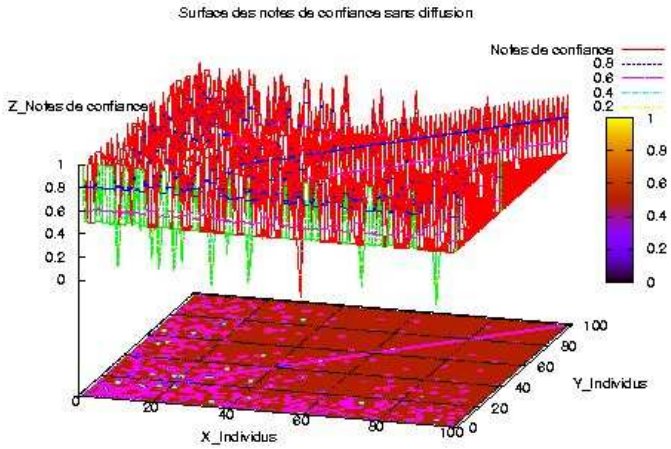
D Les coalitions “simples”

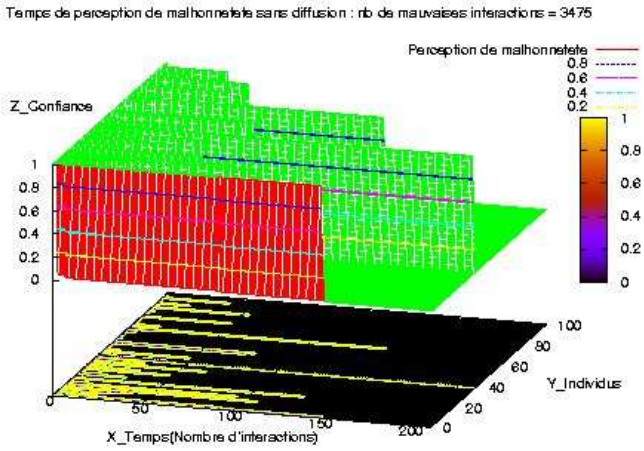
Simulation F

Simulations réalisées sur un graphe à distribution de degrés fixéee suivant une loi de puissance ayant subi une augmentation du coefficient de *clustering* de 100 nœuds mettant en œuvre un groupe de 10 individus malhonnêtes possédant 20% de chance de faire une mauvaise interaction. Ce groupe est composé des individus 5, 15, 25, 35, 45, 55, 65, 75, 85 et 95. Cette distribution des nœuds malhonnêtes permet de répartir “la malhonnêteté” par rapport aux degrés des individus du réseau.



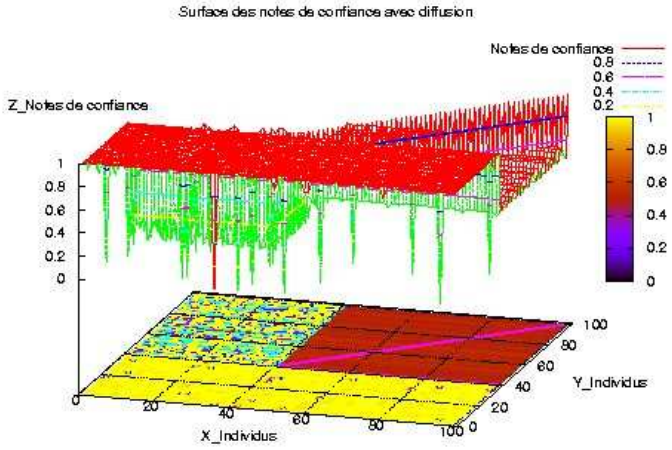


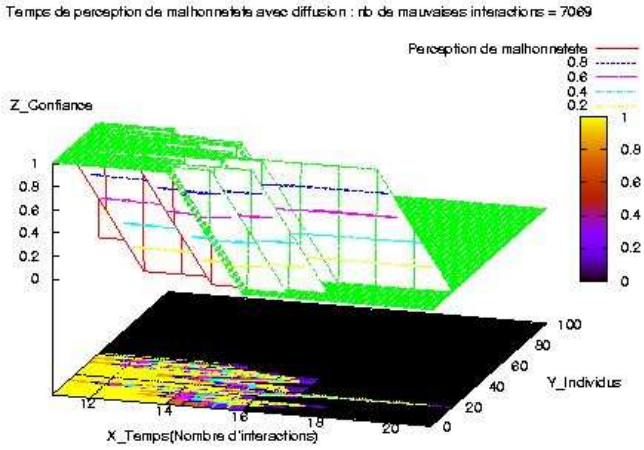


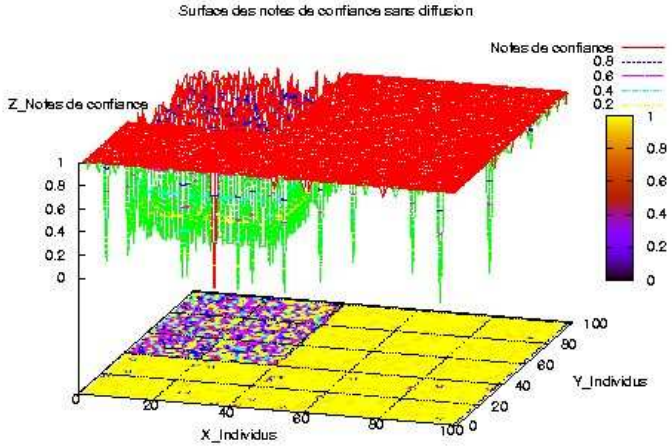


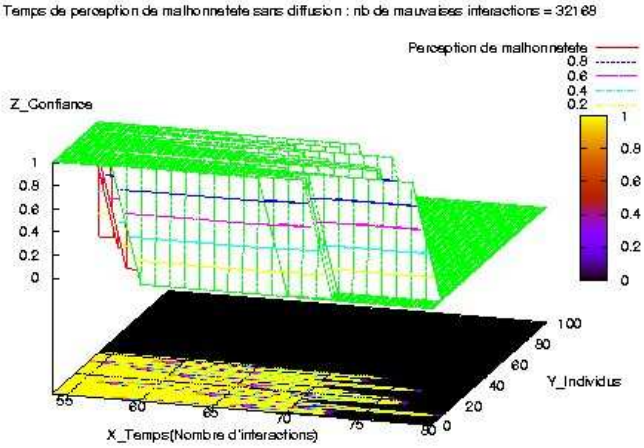
Simulation G

Simulations réalisées sur un graphe complet de 100 nœuds mettant en œuvre un groupe de 60 individus malhonnêtes possédant 20% de chance de faire une mauvaise interaction. Ce groupe est composé des individus 41 . . . 100.



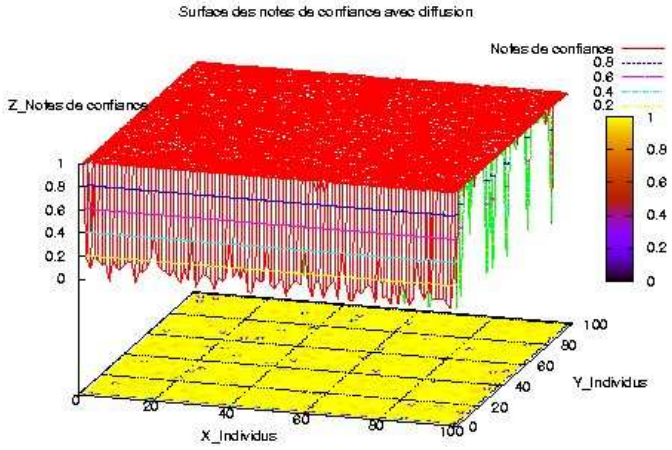


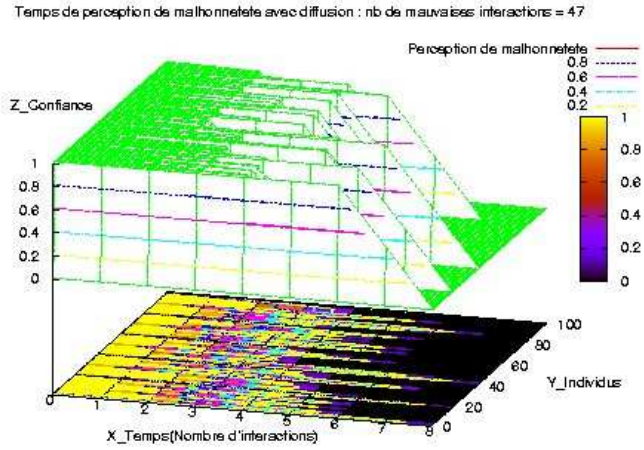


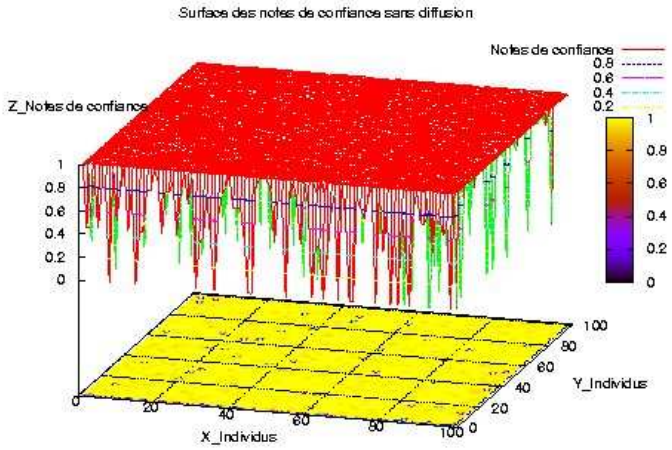


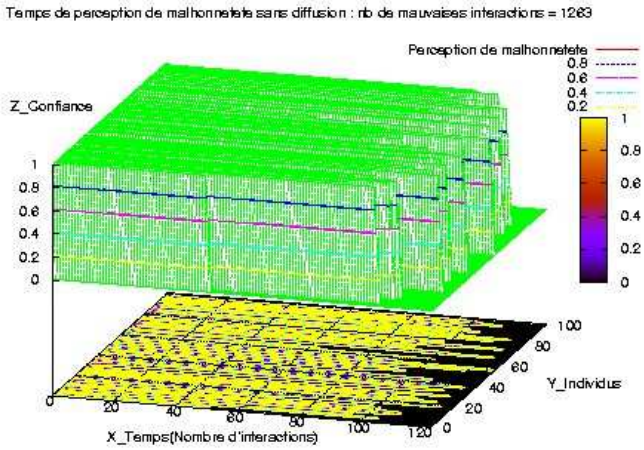
Simulation H

Simulations réalisées sur un graphe complet de 100 nœuds mettant en œuvre un groupe de 60 individus malhonnêtes possédant 20% de chance de faire une mauvaise interaction. Ce groupe est composé des individus 31 . . . 100.



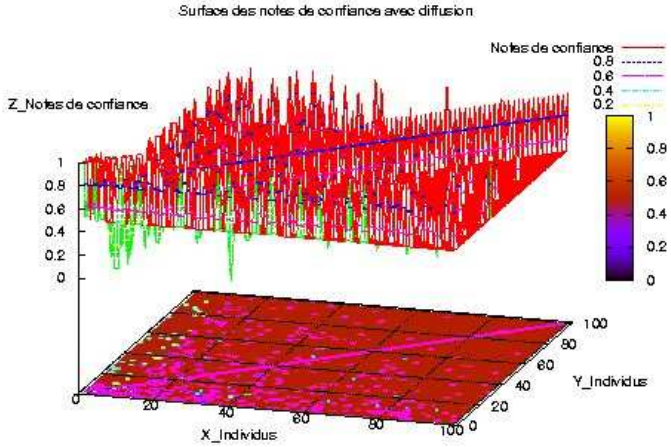


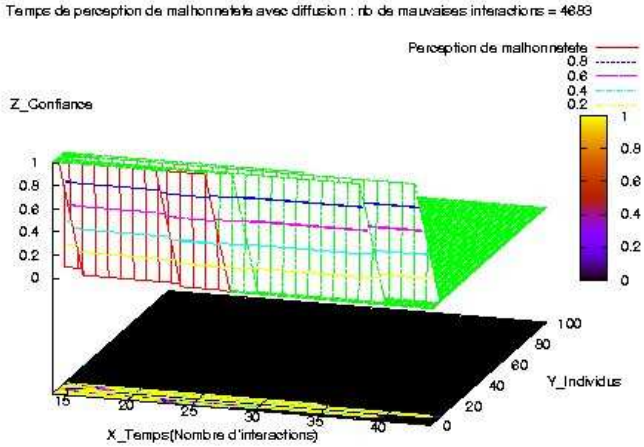


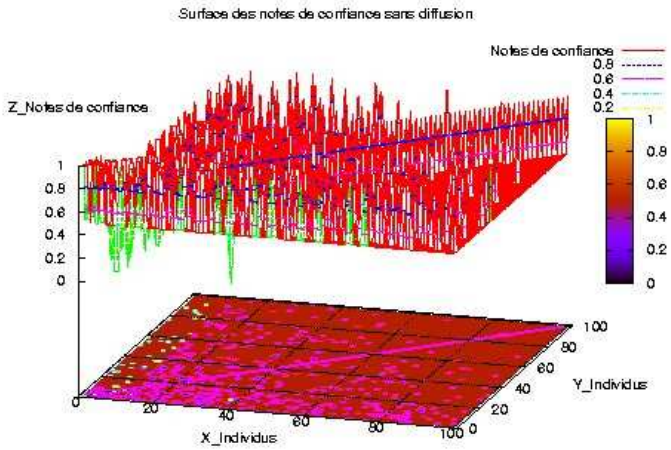


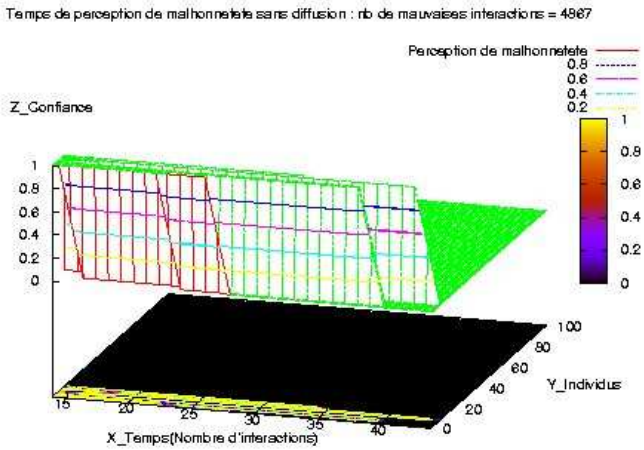
Simulation I

Simulations réalisées sur un graphe à distribution de degrés fixéee suivant une loi de puissance ayant subi une augmentation du coefficient de *clustering* de 100 nœuds. Elles mettent en œuvre un groupe de 90 individus malhonnêtes sur les nœuds de plus bas degrés (i.e. 11 . . . 100) possédant chacun 20% de chance de faire une mauvaise interaction.





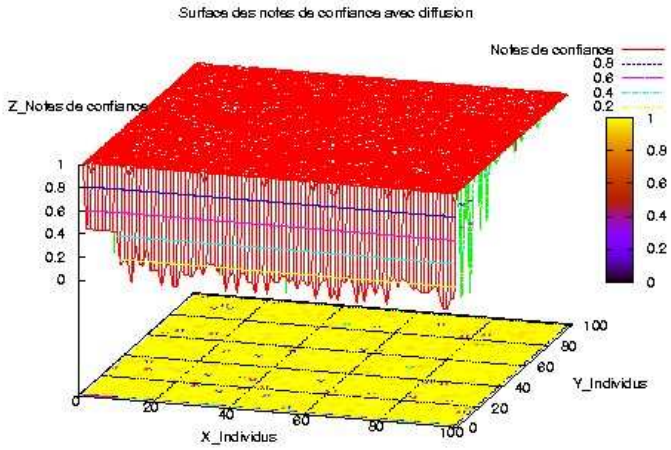


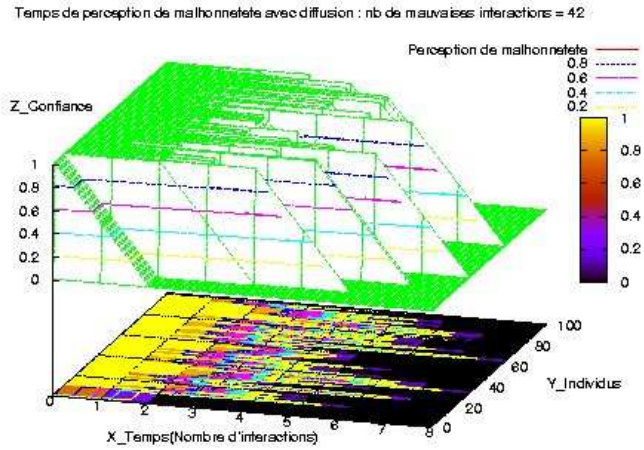


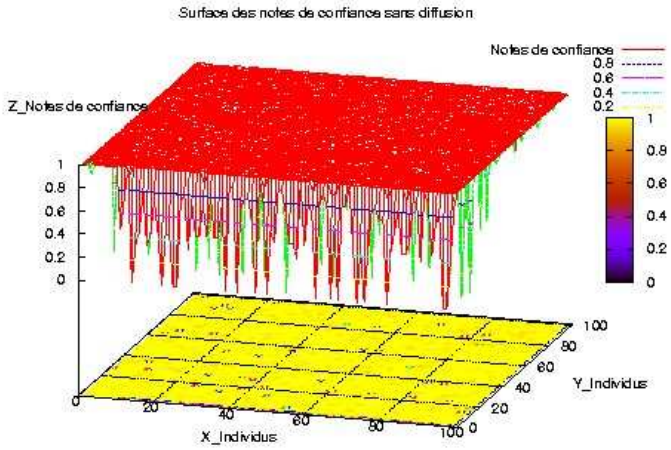
E La perception d'un "cheval de Troie"

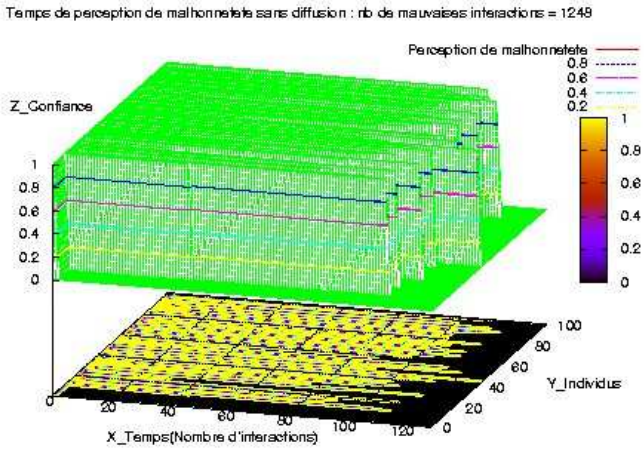
Simulation K

Simulations réalisées sur un graphe complet de 100 nœuds regroupant un cheval de Troie et 9 individus le protégeant en diffusant aux individus honnêtes de mauvaises informations à son sujet. Le cheval de Troie est positionné sur l'individu 1, appartient au groupe composé des personnes $20, 30, 40 \ldots 100$ et possède 20% de chances de faire une mauvaise interaction.



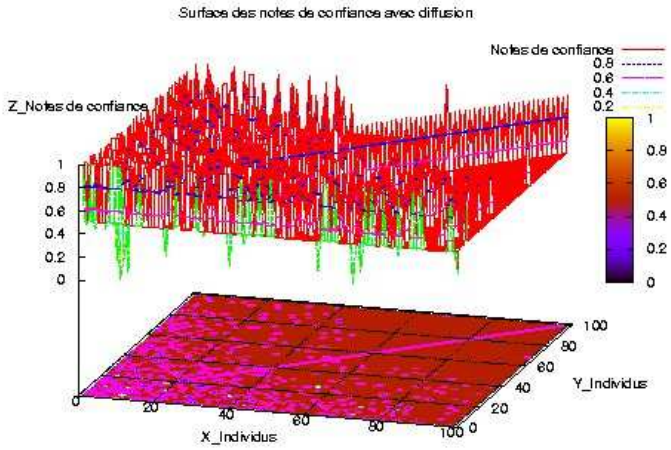


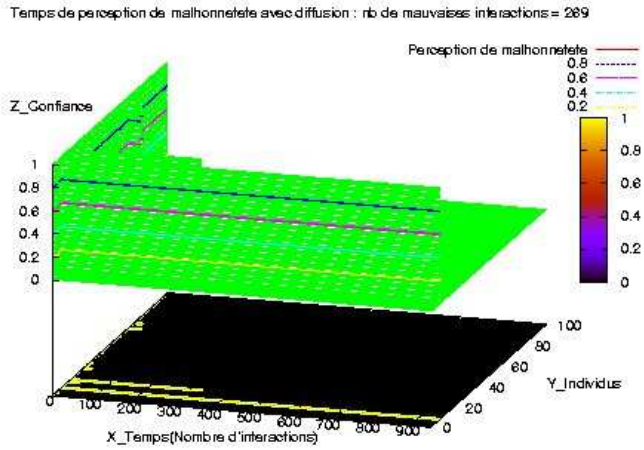


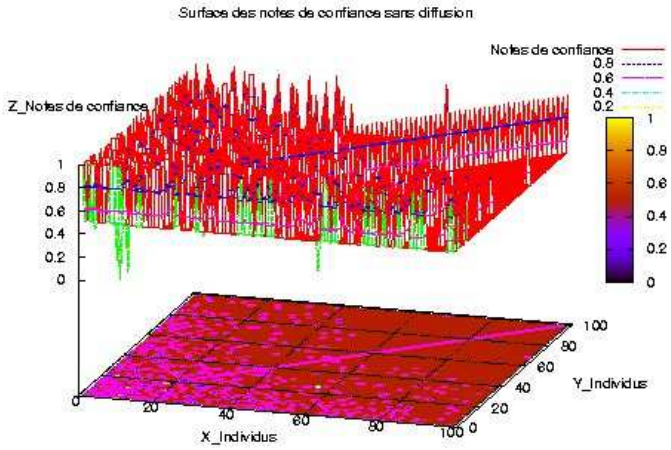


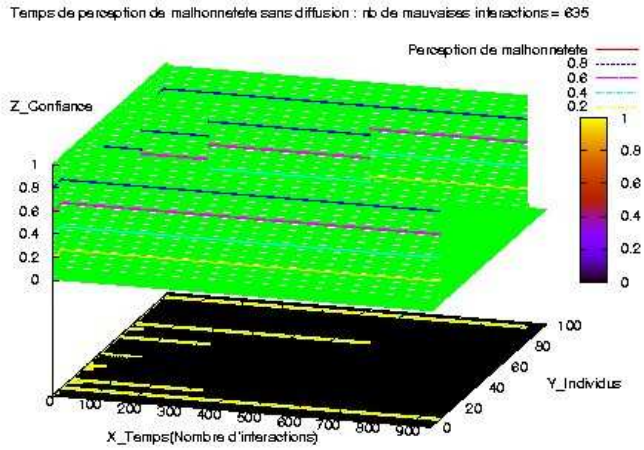
Simulation L

Simulations réalisées sur un graphe à distribution de degrés fixée suivant une loi de puissance ayant subi une augmentation du coefficient de *clustering* de 100 nœuds. Elles regroupent un cheval de Troie et 9 individus le protégeant en diffusant aux individus honnêtes de mauvaises informations à son sujet. Le cheval de Troie est positionné sur l'individu 5, appartient au groupe composé des personnes $15, 25, 35 \ldots 95$ et possède 20% de chances de faire une mauvaise interaction.



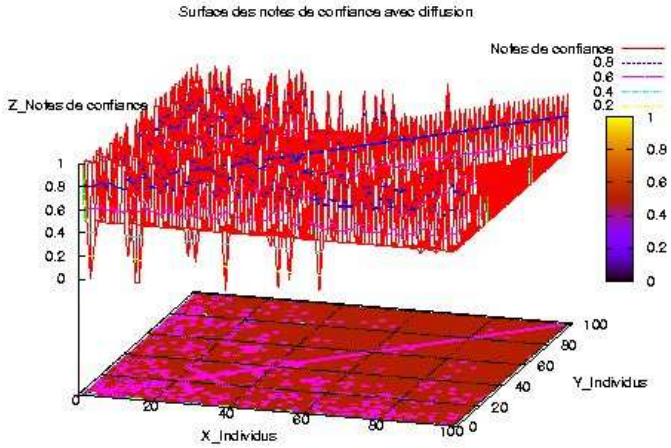


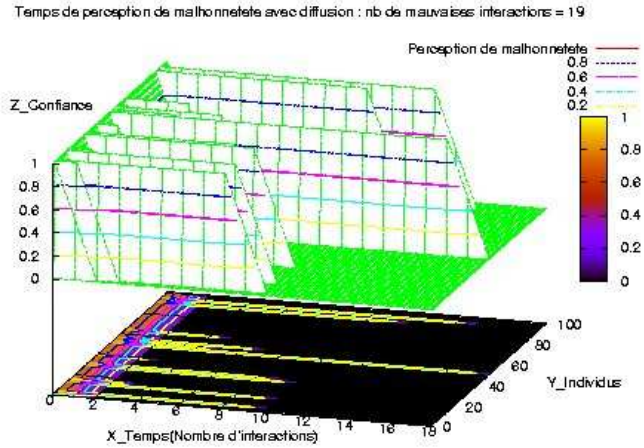


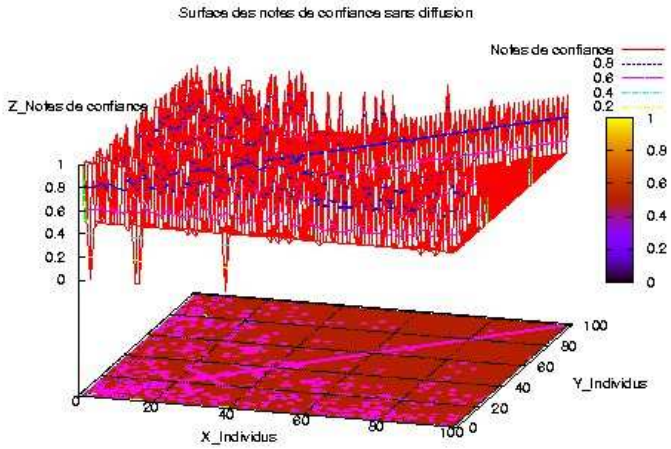


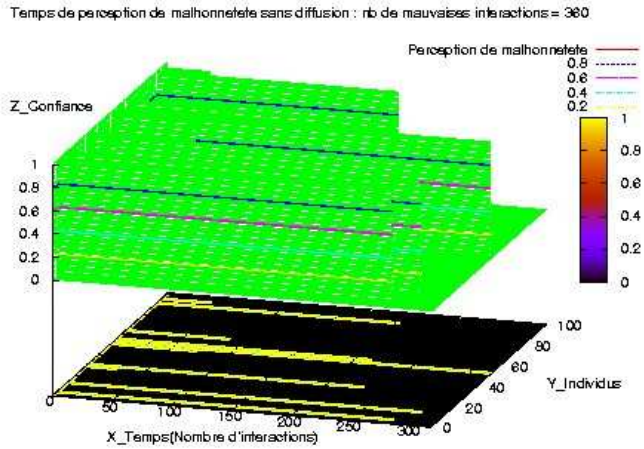
Simulation M

Simulations réalisées sur un graphe à distribution de degrés fixée suivant une loi de puissance ayant subi une augmentation du coefficient de *clustering* de 100 nœuds. Elles regroupent un cheval de Troie et 9 individus le protégeant en diffusant aux individus honnêtes de mauvaises informations à son sujet. Le cheval de Troie est positionné sur l'individu 1 ayant le plus grand nombre de voisins, appartient au groupe composé des personnes $20, 30, 40 \ldots 100$ et possède 20% de chances de faire une mauvaise interaction.





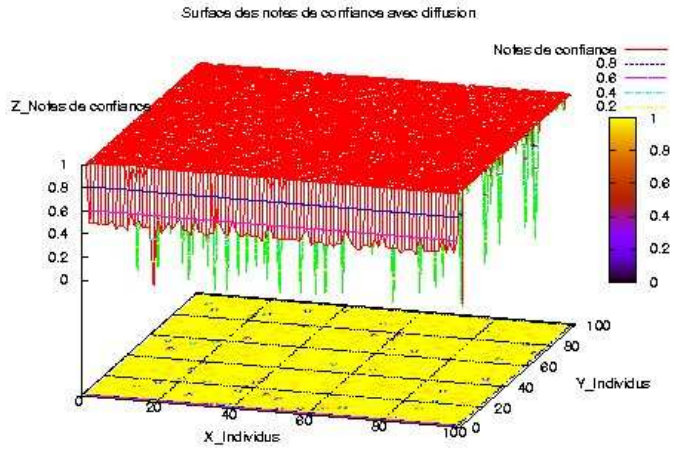


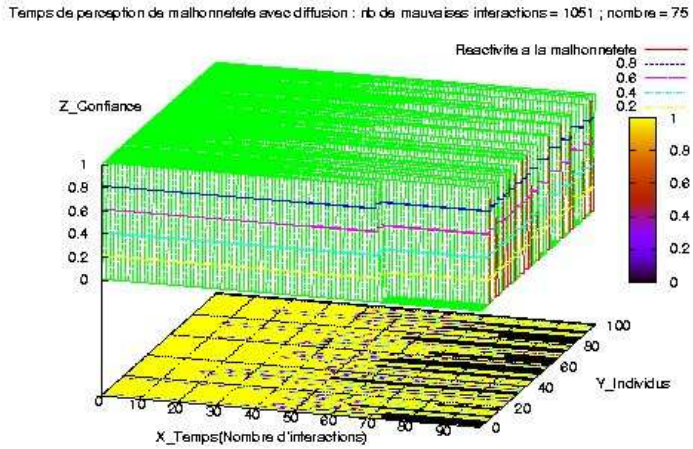


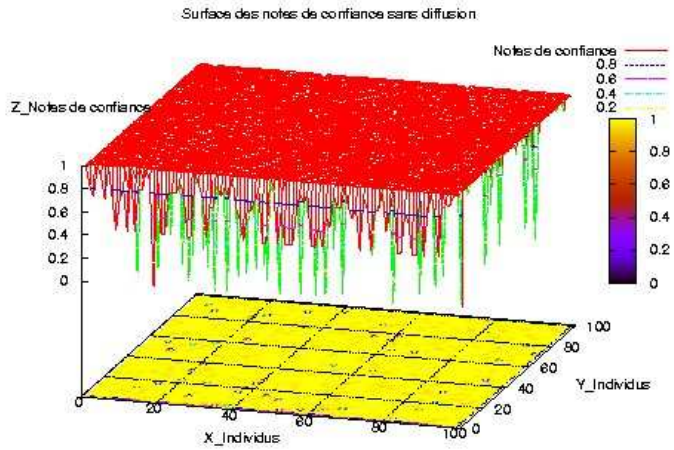
F La réactivité du système face à une "bombe logique"

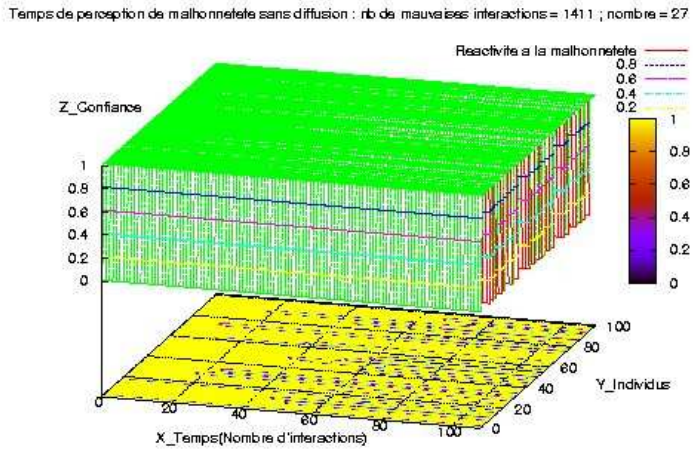
Simulation O

Simulations réalisées sur un graphe complet de 100 nœuds regroupant une bombe logique et 9 individus la protégeant en diffusant aux individus honnêtes de mauvaises informations à son sujet. La bombe logique est positionnée sur l'individu 1 et appartient au groupe composé des personnes $20, 30, 40 \ldots 100$. Elle est déclenchée à la 100000^e interaction et possède 20% de chances de faire une mauvaise interaction.









Simulation P

Simulations réalisées sur un graphe à distribution de degrés fixée suivant une loi de puissance ayant subi une augmentation du coefficient de *clustering* de 100 nœuds. Elles regroupent une bombe logique et 9 individus le protégeant en diffusant aux individus honnêtes de mauvaises informations à son sujet. La bombe logique est positionnée sur l'individu 1 ayant le plus grand nombre de voisins et appartient au groupe composé des personnes $20, 30, 40 \ldots 100$. Elle est déclenchée à la 100000^{e} interaction et possède 20% de chances de faire une mauvaise interaction.