Mémoire présenté devant l'Université de Paris-Dauphine pour l'obtention du Certificat d'Actuaire de Paris-Dauphine et l'admission à l'Institut des Actuaires

le

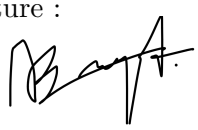
Par : Thomas Peyrat
Titre : Risque cyber, un modèle épidémiologique sur réseaux pour le risque d'accumulation du cyber silencieux.

Confidentialité : ☑ Non ☐ Oui (Durée : ☐ 1 an ☐ 2 ans)

Les signataires s'engagent à respecter la confidentialité ci-dessus

Membres présents du jury de l'Institut des Actuaires :

Entreprise :
Nom : Milliman SAS
Signature :



Directeur de Mémoire en entreprise :
Nom : Yousra Cherkaoui Tangi
Signature :

Membres présents du Jury du Certificat d'Actuaire de Paris-Dauphine :

Autorisation de publication et de mise en ligne sur un site de diffusion de documents actuariels (après expiration de l'éventuel délai de confidentialité)

Secrétariat :

Signature du responsable entreprise

Bibliothèque :

Signature du candidat

Résumé

PRA (PRA, 2016), ACPR (ACPR, 2019), EIOPA (EIOPA, 2020) et la Direction Générale du Trésor (Direction générale du Trésor, 2022) entre autres, ont souligné l'importance d'une évaluation silencieuse du cyber risque. De plus, le risque d'accumulation cyber est un enjeu de modélisation important comme le montre l'évènement Wannacry de 2017. Nous présentons dans cet article un portefeuille fictif où les interactions entre les assurés sont modélisées par un réseau qui se base sur les secteurs présentés lors d'une étude de l'OCDE (OCDE, 2018). Dans notre application, nous commençons par définir l'exposition silencieuse au sein d'un portefeuille comme la part des contrats n'excluant ou n'affirmant pas de manière explicite le risque cyber, et ce, pour chaque garantie de la police souscrite. Ensuite, en utilisant un scénario de perte d'exploitation, nous évaluons le nombre d'assurés infectés et la perte probable en utilisant des modèles épidémiologiques afin de modéliser un évènement d'accumulation. Il en ressort que l'ajustement de certains paramètres (la capacité d'intervention de l'assureur ou les secteurs du portefeuille) peut réduire la perte globale.

Mots-clés : *Cyber Silencieux*, *Risque d'accumulation* ; *Modèles épidémiologiques* ; *Réseaux*.

Abstract

PRA (PRA, 2016), ACPR (ACPR, 2019), EIOPA (EIOPA, 2020) and Direction Générale du Trésor (Direction générale du Trésor, 2022) among others, have highlighted the importance of silent cyber risk assessment. We present in this paper a toy portfolio where the interactions between the insured are modeled through a network inspired by OECD sectors data (OCDE, 2018). We first define the portfolio's silent exposure as the part of the contracts that don't explicitly exclude or affirm cyber risks, and this, for each coverage in the policy. Then, using a business interruption scenario, we evaluate the number of infected policyholders and the probable loss using epidemiological models. From this, we see that adjustments to certain parameters (insurer's intervention capacity or the sectors in the portfolio) could reduce the overall global loss.

Keywords : *Silent Cyber*; *Accumulation Risk*; *Epidemiological models*; *Networks.*

Note de Synthèse

Contexte

Comme présenté dans Hillairet et Lopez, 2022, le cyber risque, qu'il soit affirmatif ou non, présente plusieurs caractéristiques qui rendent sa modélisation spécifique. Le risque d'accumulation des pertes est l'une d'entre elles. Ce risque provient d'évènements majeurs qui touchent simultanément une grande proportion d'assurés, ce qui est contraire au principe même de mutualisation des risques. Nous parlons de cyber-risque silencieux (ou non affirmatif) lorsque les polices non cyber n'incluent ou n'excluent pas explicitement le cyber-risque dans leurs couvertures. Par exemple, nous pouvons citer le cas de Mondelez. Cette multinationale américaine de l'agroalimentaire a été victime de l'importante attaque par *ransomware* NotPetya en 2017, causant ainsi d'importantes difficultés opérationnelles. Lors de cette attaque, les montants des sinistres s'élevaient à 100m$ pour un cyber-événement sur une police d'assurance de biens, voir Cartagena et al., 2020.

Nous implémentons dans cet article un modèle épidémiologique sur un réseau qui modélise les interactions potentielles des assurés souscrivant une police d'assurance non-vie, non dédiée à la couverture du risque cyber. Nous utilisons une approche granulaire qui nous permet de modéliser chaque assuré. Le modèle peut ainsi être facilement complété avec les informations internes des assureurs.

Modélisation du risque d'accumulation cyber

Notre objectif est d'évaluer le nombre potentiel d'infectés et les pertes associées lors d'un événement d'accumulation cyber dans un portefeuille non cyber. Pour cela, nous utilisons un modèle épidémiologique stochastique se propageant dans une structure de réseau.

Du modèle épidémiologique déterministe au stochastique

De la même manière qu'un virus biologique se propage dans une population, pouvant conduire à une épidémie, les malwares peuvent générer des épisodes d'accumulation comme Wannacry ou NotPetya en 2017. Les modèles épidémiologiques compartimentaux sont adaptés pour décrire la propagation d'un virus au sein d'une population globale. L'un des plus célèbres est le modèle SIR, qui signifie Susceptible, Infecté et Rétabli (ou guéri). La façon la plus simple de représenter l'évolution de la population à travers les trois états est d'utiliser un système d'équations différentielles ordinaires (le modèle déterministe) où S, I et R comptent le nombre d'individus dans chaque état, et N est le nombre total d'individus :

$$\begin{aligned}
\frac{dS(t)}{dt} &= -\beta I(t)\frac{S(t)}{N}, \\
\frac{dI(t)}{dt} &= \beta I(t)\frac{S(t)}{N} - \gamma I(t), \\
\frac{dR(t)}{dt} &= \gamma I(t).
\end{aligned}$$

Les paramètres β et γ représentent respectivement les taux d'infection et de rétablissement. Comme expliqué dans Kiss et al., 2017, β est le taux avec lequel les individus infectés établissent des contacts infectieux (infection potentielle), ainsi le nombre βI représente le nombre total de contacts infectieux. Mais, parmi tous les contacts établis, seule une fraction $\frac{S}{N}$ est constituée d'individus sensibles. Un abus de notation est souvent réalisé lorsque nous traitons les modèles compartimentaux, en effet S, I et R représentent le nombre d'individus mais également les états compartimentaux S, I et R.

Comme expliqué dans Fahrenwaldt et al., 2018, les structures de réseau, utilisées pour modéliser les interactions au sein d'une population population, ont un impact direct sur la diffusion du virus. L'ajout d'un graphe (ou réseau) permet d'ajouter une hétérogénéité dans la diffusion des virus (dans notre cas des malwares) puisque leur vitesse de propagation peut varier en fonction de différents facteurs qui sont propres à chaque individu. Ceci est réalisé en utilisant des poids sur les arcs du graphe.

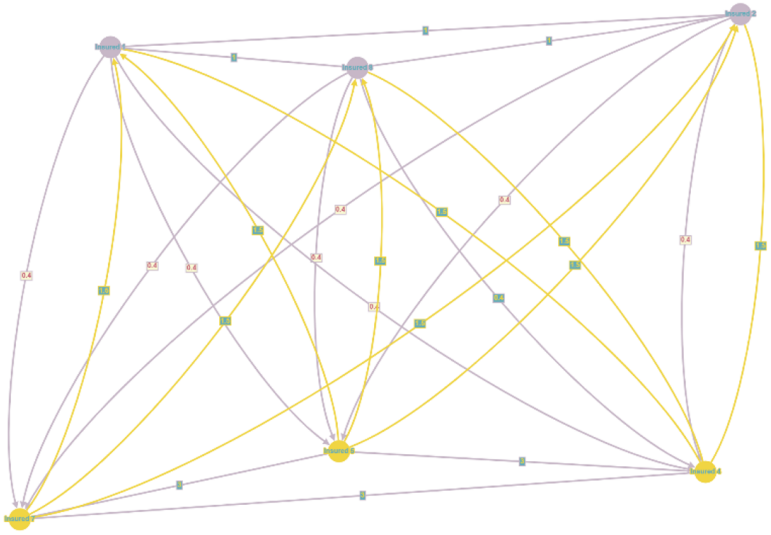


Figure 1 : Exemple d'un graphe orienté avec des poids sur les arêtes.

Comme nous pouvons le voir ci-dessus, les arcs jaunes représentent les poids auxquels les nœuds de la classe jaune infecteront les nœuds de la classe grise. Les arêtes pondérées au sein d'une même classe représentent les poids d'infection entre les nœuds de la même classe.

Pour modéliser la diffusion du virus dans un réseau, nous utilisons le modèle à processus de Markov continu. Ainsi, pour N nœuds, nous désignons par $E_s(t)$ l'état dans lequel se trouve le nœud s à l'instant t. Par conséquent, nous avons $E_s(t) \in \{S, I, R\}$ et,

$$E_{s_i}(t) : S \to I \quad \text{with rate} \quad \beta \sum_{s_j \in S} a_{ij} \mathbb{1}_{E_{s_j}(t)=I},$$

$$E_{s_i}(t) : I \to R \quad \text{with rate} \quad \gamma.$$

Où a_{ij} prend généralement des valeurs dans l'intervalle $\{0, 1\}$, 1 représentant le cas où un contact existe entre les nœuds s_i et s_j et 0 le cas où il n'y a pas de contact possible entre les deux nœuds. De plus, comme nous voulons ajouter des poids aux arêtes, nous permettons à a_{ij} de prendre d'autres valeurs que $\{0, 1\}$. Comme pour le modèle déterministe, nous constatons que la guérison d'un individu infecté ne dépend que de la valeur de γ.

L'algorithme utilisé pour simuler le modèle est l'*Event-driven fast SIR* décrit dans l'annexe A.1.2 de Kiss et al., 2017.

Quel réseau devons-nous utiliser ?

L'idée derrière l'ajout d'une structure de réseau entre les assurés est de mieux refléter l'environnement dans lequel le malware va se propager. En outre, les assureurs seront en mesure de déterminer quelles classes (secteurs par exemple) sont plus susceptibles d'être infectées ou non. Pour mieux illustrer cela, nous considérons le réseau sectoriel introduit dans Hillairet et al., 2021. Il est construit à partir des volumes échangés entre secteurs à l'issu d'une étude de l'OCDE. Nous le proportionnons en fonction d'un secteur, que nous appelons secteur de référence. Dans notre modélisation, ce dernier est le secteur *Mining*, ce qui signifie que le poids a_{ij} est égal à 1 si l'assuré s_i et l'assuré s_j appartiennent à ce même secteur, voir figure (2).

Nous devons garder à l'esprit que la structure du réseau peut être calibrée à l'aide d'informations disponibles à la souscription telles que les partenariats, les relations commerciales ou toute autre information susceptible de représenter une interaction. De plus, les secteurs pourraient être remplacés par des classes qui représentent de manière plus complexe la connectivité entre les assurés.

Sectors	Mining	Manufacturing	Energy	Construction	Services
Mining	1	4,61672	0,7082	2,25079	1,9795
Manufacturing	0,0994	0,83123	0,04259	0,17035	0,55363
Energy	0,21293	0,58359	0,90063	0,23659	0,71293
Construction	0,02997	0,10726	0,01104	0,22239	0,14353
Services	0,00473	0,06624	0,00631	0,02681	0,25394

Figure 2 : Matrice des poids du réseau selon le secteur d'activité des assurés.

Dans la figure (2), les coefficients diagonaux représentent la connectivité au sein des assurés d'un même secteur. Les autres coefficients illustrent la façon dont les assurés de différents secteurs sont connectés. De plus, ces coefficients seront utilisés pour construire la matrice d'adjacence du réseau (c'est-à-dire les valeurs de a_{ij}). Ainsi, chaque assuré est connecté aux autres mais avec des poids différents.

La matrice n'est pas symétrique, cela signifie que certains secteurs sont mieux défendus ou, au contraire, que certains pourraient être utilisés comme vecteur de propagation du virus.

Le portefeuille fictif

Afin d'étudier l'impact de la menace d'accumulation cybernétique dans un cadre non cyber, nous considérons un portefeuille d'assurance multirisques professionnels. Ce portefeuille contient une hétérogénéité dans le libellé des contrats en raison, par exemple, de l'année de souscription.

Chaque assuré appartient à un secteur d'activité

Au sein du portefeuille, les 1 000 assurés sont répartis de manière égale dans chacun des secteurs, voir figure (3).

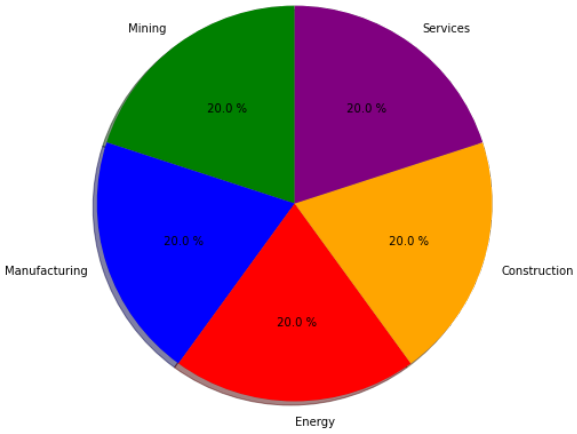


FIGURE 3 : Proportion d'assurés dans chacun des secteurs du portefeuilles.

Utilisation de l'exposition de chaque assuré

Dans le portefeuille fictif, au-delà de l'information du secteur de l'assuré, nous disposons également de l'exposition pour chaque garantie incluse dans sa police. Par exposition, nous entendons la différence entre (si elle est présente) la limite d'indemnisation et la franchise. L'exposition est censée représenter le montant réel indemnisable par l'assureur pour chacune des garanties de l'assuré. Un objectif important de notre modélisation est de pouvoir déduire l'exposition cyber silencieuse d'un portefeuille non cyber.

Évaluation des expositions silencieuses

De manière générale nous voulons évaluer le risque d'accumulation cyber porté par un portefeuille non cyber. Pour ce faire, nous commençons par déterminer les expositions silencieuses selon les quatre étapes décrites dans la figure (4). Ces étapes s'inspirent du cadre d'évaluation du risque cyber silencieux de l'IFoA, voir CARTAGENA et al., 2020.

FIGURE 4 : Étapes d'évaluation du risque cyber silencieux.

Les quatre étapes nous fournissent uniquement l'exposition silencieuse

Les quatre étapes du cadre illustré dans la figure ci-dessus nous fournissent uniquement l'exposition silencieuse. Or, dans notre cas, nous cherchons à avoir la distribution du nombre d'assurés infectés et celle des coûts au cours du temps. C'est la raison pour laquelle nous fusionnons les étapes 3 et 4 avec le modèle épidémiologique présenté dans la première section. Pour ce faire, chaque assuré infecté va activer une garantie en fonction d'un taux silencieux. Ce dernier représente la fréquence avec laquelle une garantie ne contient pas des clauses cyber affirmatives ou exclusives. Il est déterminé à l'étape 2 de la figure (4). Par exemple, dans un portefeuille d'assurance professionnelle où la perte d'exploitation est couverte, un taux silencieux de 20 % signifie que sur 100 assurés infectés dans notre portefeuille, 20 peuvent déclencher un indemnisation à partir d'un évènement cyber.

L'évaluation du taux silencieux est l'une des parties les plus critiques de la modélisation. En effet, une mauvaise évaluation entraînera une mauvaise évaluation de la perte probable finale. Pour mieux estimer ce paramètre, il est nécessaire de travailler avec les services de souscription et les services juridiques. De plus, certaines caractéristiques de la modélisation NLP pourraient être utilisées pour faciliter et accélérer l'estimation.

Dans notre portefeuille fictif, le taux silencieux varie pour chaque couverture, et il est déjà donné avec les autres informations du portefeuille.

Évaluation des scénarios

La génération de scénarios pour évaluer les pertes potentielles est une approche couramment utilisée en assurance. Elle est utilisée par exemple pour évaluer les pertes liées aux inondations et aux catastrophes naturelles. Dans la figure (5), nous avons quelques exemples des scénarios qui pourraient déclencher des couvertures silencieuses dans des polices spécifiques, voir Marsh, 2020.

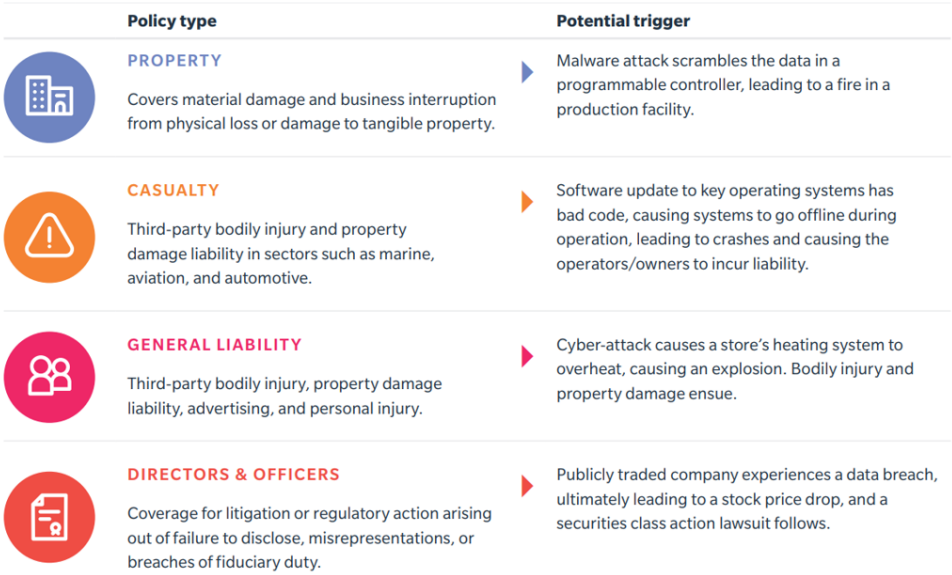


Figure 5 : Quelques scénarios pouvant déclencher des couvertures silencieuses (Marsh, 2020).

Dans notre modélisation, nous considérons que chaque assuré porte un risque silencieux sur chacune de ses garanties selon le taux silencieux précédemment évalué. Par conséquent, nous définissons un scénario comme une liste de couvertures déclenchables par l'évènement cyber. Afin d'ajouter de la variabilité aux scénarios, on pourrait associer chaque garantie à une probabilité de déclenchement.

Certains scénarios réalistes peuvent être trouvés dans Lloyd's, 2022.

Dans cet article, nous nous concentrons sur un scénario d'interruption d'activité qui déclenche uniquement la garantie perte d'exploitation de l'assuré.

Scénario perte d'exploitation

Description

Dans ce scénario, nous considérons qu'un *ransomware* provoque une interruption de l'activité depuis le moment de l'infection jusqu'à la restauration du système. Un *ransomware* est un type courant de logiciels malveillants qui chiffrent les données ou bloquent des systèmes entiers et restaurent tout après le paiement d'une rançon (généralement en bitcoins). Dans ce scénario, nous considérons que le *ransomware* bloque uniquement le système informatique et provoque ainsi l'activation de la garantie perte d'exploitation.

Le paramètre d'infection β (introduit dans la première section) sera fixé à 0,01 et γ, le paramètre de rétablissement, à 1. Un ordre de grandeur (très approximatif), est qu'avec ce taux de contagion un nœud infecté (assuré) contaminera environ 1% des nœuds sensibles auxquels il est lié et se rétablira en un jour environ. Ce pourcentage d'infection est susceptible de changer en raison du réseau pondéré que nous utilisons dans notre modèle, mais la récupération ne variera pas puisqu'elle ne dépend pas de la structure du réseau.

Modélisation des pertes

Nous considérons que la perte pour un assuré est une fonction croissante du temps étant paralysé par le *ransomware*. Pour chaque jour passé en perte d'exploitation, nous allons tirer une réalisation d'une loi gamma de paramètres a et b en fonction du secteur de l'assuré afin de simuler un montant indemnisable par l'assureur. Les paramètres a et b sont déterminés en fonction de l'espérance et de la variance que nous souhaitons donner à la distribution. Le montant indemnisable sera tronqué de l'exposition inscrite dans le portefeuille, ce qui signifie que l'indemnisation ne peut jamais aller au-delà de l'exposition.

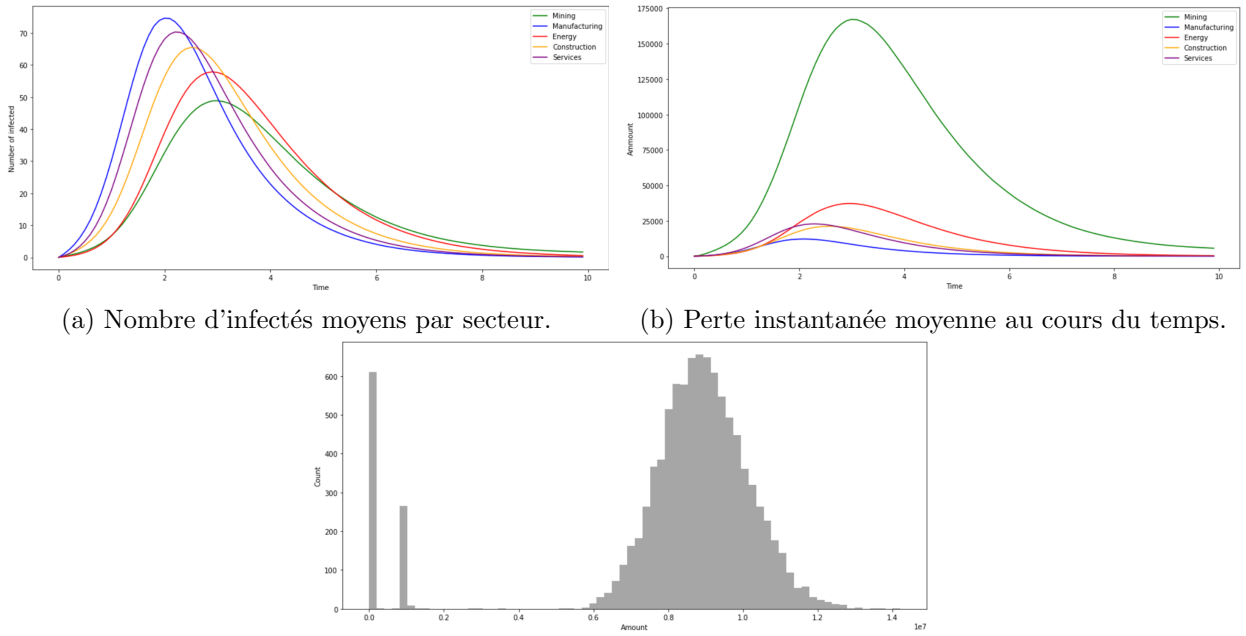
Table 1 : Paramètres pour des lois gamma modélisant le coût journalier selon le secteur de l'assuré.

Secteur	a	b	Espérance	Variance
Mining	200 000,00	0,5	100 000,00	50 000,00
Manufacturing	10 000,00	0,5	5 000,00	2 500,00
Energy	40 000,00	0,5	20 000,00	10 000,00
Construction	20 000,00	0,5	10 000,00	5 000,00
Services	20 000,00	0,5	10 000,00	5 000,00

Nous pouvons voir dans le tableau précédent que les secteurs de *Construction* et *Services* ont la même distribution des coûts journaliers. Rappelons que pour une variable aléatoire $G \sim \mathcal{G}(a, b)$ nous avons $\mathbb{E}[G] = a \times b$ et $\mathbb{V}ar[G] = a \times b^2$. De plus, la structure du réseau sera celle introduite dans la première section. Comme nous l'avons déjà mentionné, étant donné que nous traitons le cas du cyber silencieux, toutes les infections n'entraînent pas l'activation de la couverture des pertes d'exploitation. Dans cet exemple, le taux silencieux, c'est-à-dire le taux auquel une infection conduit à une indemnisation, est de 32 %.

Résultats sur le portefeuille fictif

Nous pouvons voir dans la figure ci-dessous comment le secteur *Manufacturing* est le premier à atteindre son pic d'infection. Ceci est directement lié au fait que le secteur *Manufacturing* a un poids a_{ij} très élevé provenant du secteur *Mining*. Ce dernier aura tendance à infecter tous les autres secteurs, voir la figure (2).



(a) Nombre d'infectés moyens par secteur.

(b) Perte instantanée moyenne au cours du temps.

(c) Distribution des pertes cumulées après 10 jours d'infection.

FIGURE 6 : Évolution du nombre d'infectés et de la perte instantanée par secteur au cours du temps.

De plus, nous remarquons dans la figure (6b), que le secteur qui génère les plus grosses indemnisations journalières est le secteur *Mining*. Cela montre qu'une faible infection du secteur, Figure (6a), ne conduit pas nécessairement à des coûts d'indemnisation réduits.

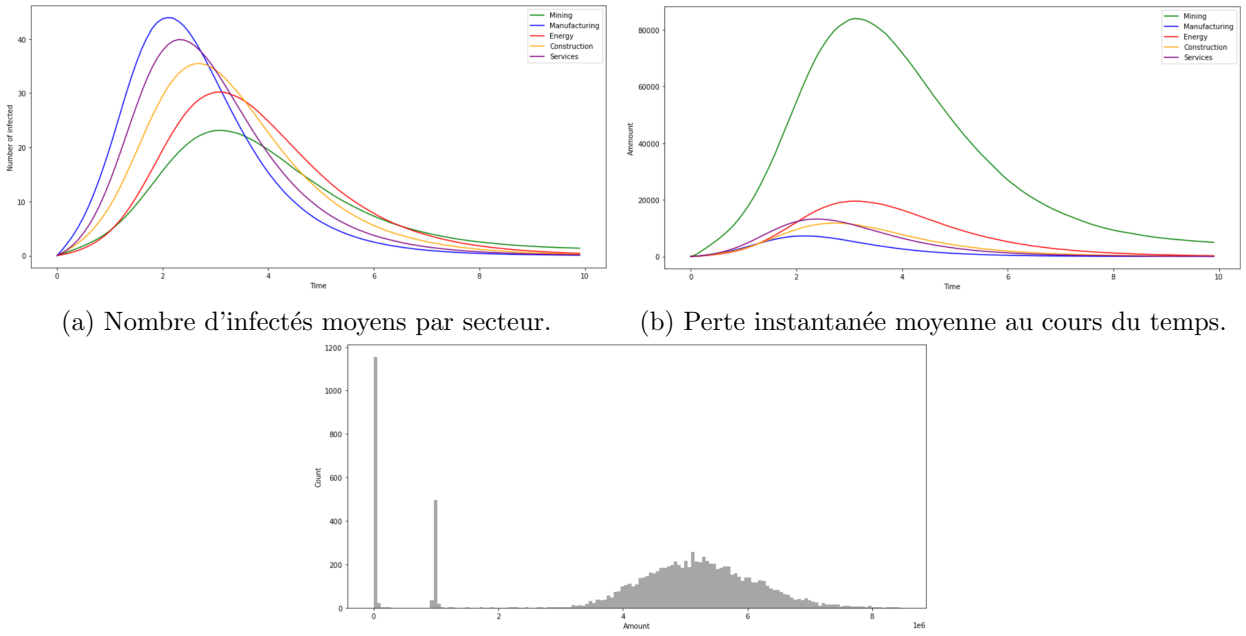
Il convient de noter qu'il existe un état stationnaire où le logiciel malveillant est éradiqué. Dans ce cas, il ne reste plus aucun infecté et donc plus aucune contamination ne peut être générée. Ceci conduit à une accumulation proche de 0, comme on peut le voir sur la figure (6c). Cette distribution est obtenue en réalisant 10 000 simulations jusqu'à 10 jours. Avec ces paramètres et ces caractéristiques de portefeuille, la valeur moyenne de la perte cumulée est de **8 204 785€**.

Augmenter la capacité d'intervention de l'assureur

L'augmentation de la capacité d'intervention est modélisée par l'augmentation du paramètre de récupération γ, ce qui accélère le rétablissement des assurés infectés. Lors de la cybercrise Wannacry en 2017 (MOHURLE et PATIL, 2017), bien que l'exploit de cyberattaque EternalBlue ait été patché (MS17-010) pour les utilisateurs de Windows en mars 2017, de nombreux utilisateurs étaient encore vulnérables pendant la crise, et certains même un an plus tard, voir VLCEK, 2018.

Certaines mesures de prévention pourraient être mises en œuvre pour améliorer l'efficacité de l'intervention de l'assureur, comme le patching des vulnérabilités et, la sensibilisation des assurés aux mises à jour du système. Notre modèle ne prend pas en compte les impacts de la prévention, comme

dans Hillairet et al., 2021, où un paramètre de réaction au cyber environnement est introduit. Ce paramètre rend les assurés plus prudents lorsque des attaques sont détectées.



(a) Nombre d'infectés moyens par secteur.

(b) Perte instantanée moyenne au cours du temps.

(c) Distribution des pertes cumulées après 10 jours d'infection.

Figure 7 : Évolution du nombre d'infectés et de la perte instantanée par secteur au cours du temps.

Sur la figure (7a), nous pouvons voir comment l'augmentation du paramètre de rétablissement diminue le pic d'infection de tous les secteurs. L'augmentation du paramètre de rétablissement entraîne une diminution des infections et donc des coûts plus faibles dans tous les secteurs, voir la figure (7b). Le calibrage de ce paramètre pourrait être effectué en utilisant le temps de récupération espéré, qui est égal à $1/\gamma$.

L'augmentation du taux de rétablissement pourrait avoir un coût pour l'assureur. En ajoutant cette information, nous pourrions comparer les avantages de l'augmentation de la capacité d'intervention et son coût. Augmenter le taux de rétablissement γ à 1,5 nous permet de diminuer de moitié la perte cumulée globale : **4 400 217 €**.

Modification de la répartition sectorielle des assurés

Dans la figure (3), le portefeuille fictif a une distribution homogène des assurés dans tous les secteurs. Mais comme nous pouvons le voir dans les résultats de la figure (6b), le secteur le plus cher à indemniser est le secteur *Mining*. Nous modifions maintenant cette répartition en réduisant la part du secteur minier, voir la figure (8). Nous analysons l'impact de cette modification sur la propagation du malware dans le temps pour chaque secteur, voir la figure (9a) et la perte associée, voir la figure (9c).

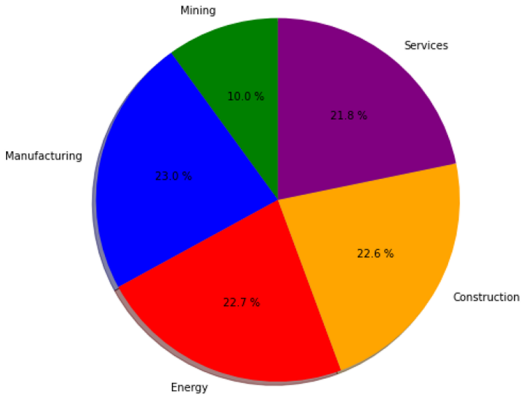
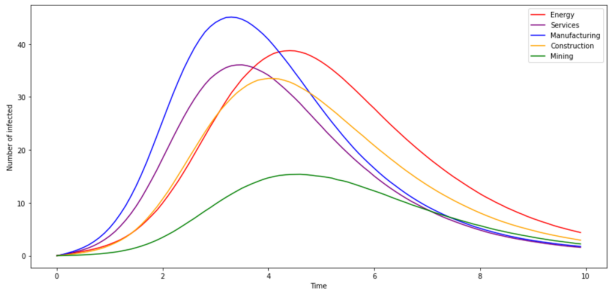
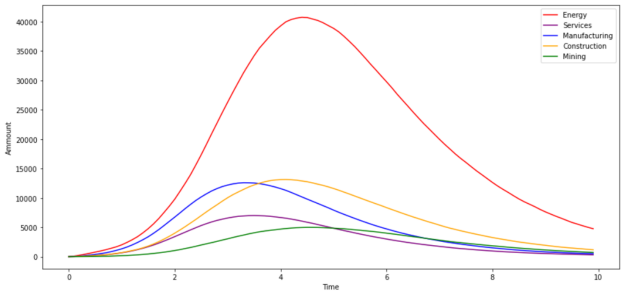


FIGURE 8 : Proportion d'assurés par secteur dans le nouveau portefeuille.

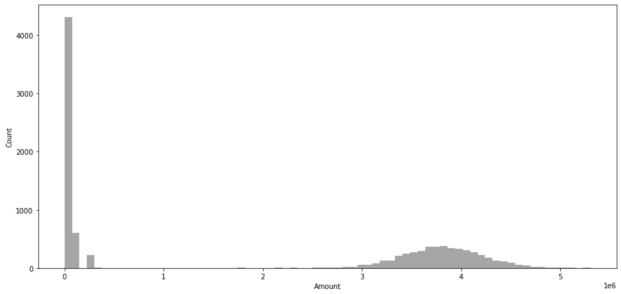
Comme nous pouvons le voir sur la figure (8), le secteur minier ne représente plus que 10% de l'ensemble des secteurs. Cela a un impact direct sur la propagation des logiciels malveillants. Dans la figure (9a), le secteur le plus rapidement contaminé est toujours le secteur manufacturier, et le secteur le plus lent est toujours le secteur minier. Cela est dû à la forte contagiosité entre le secteur *Mining* et *Manufacturing*, voir la figure (2) . Dans la figure (9c), le secteur le plus coûteux est *Energy*. La vitesse de diffusion est conditionnée par le secteur.



(a) Nombre d'infectés moyens par secteur.



(b) Perte instantanée moyenne au cours du temps.



(c) Distribution des pertes cumulées après 10 jours d'infection.

FIGURE 9 : Évolution du nombre d'infectés et de la perte instantanée par secteur au cours du temps.

Connaissant les liens de connectivité entre les secteur, nous avons pu établir une meilleure répartition des assurés dans le portefeuille. Nous sommes ainsi en mesure de déduire quelques lignes directrices pour limiter les pertes potentielles dans le cas d'un scénario de cyber accumulation.

En réduisant le nombre d'assurés du secteur *Mining* dans le portefeuille, nous diminuons la perte

globale à **1 872 643 €**. Cette forte diminution de la perte finale est due à la diminution du nombre d'assurés dans le secteur minier mais aussi au fait que les assurés de ce même secteur sont les plus chers à indemniser.

Conclusion

Ces dernières années, le cyber silencieux a été l'une des grandes préoccupations des services de souscription et des services juridiques en raison de la complexité de son évaluation.

Plus généralement, l'évaluation du risque d'accumulation cyber peut se faire à l'aide de modèles épidémiologiques. Dans cette étude nous avons introduit une structure de réseau qui permet de modéliser l'environnement dans lequel le virus se propage. La donnée permettant de construire ce réseau est le secteur des assurés, mais d'autres éléments peuvent être utilisées pour décrire les interactions entre les assurés.

Ces outils nous permettent non seulement de calculer les pertes potentielles mais peuvent également fournir des éléments pour déduire des lignes directrices chez les assureurs. Ainsi ils sont en mesure de mieux gérer leurs risques et de prendre les mesures en conséquences. En outre, des scénarios plus complexes pourraient être mis en œuvre, mais davantage d'informations sur la souscription seraient nécessaires pour conserver le réalisme de l'étude.

Synthesis note

Context

As presented in Hillairet and Lopez, 2022, cyber risk, whether it is affirmative or not, has several characteristics that make its modeling more specific. The accumulation of losses is one of them. This risk arises when a large number of insureds cause claims in a relatively short period of time. We refer to silent (or non-affirmative) cyber risk when non cyber policies don't explicitly include or exclude cyber risk in their coverages. The Mondelez case is one to cite. Mondelez is an American multinational food company that was victim of the major ransomware attack NotPetya in 2017 causing major operational difficulties. In this attack, claims amounts were 100m$ for a cyber event on a property policy, see Cartagena et al., 2020. We implement in this paper an epidemiological model on a network which models non cyber policyholder's potential interactions. We use a granular approach allowing us to model each policyholder. The model can thus be easily completed with the insurer's internal information.

Modeling cyber accumulation risk

We aim to assess the potential number of infected and the associated losses in a cyber accumulation event in a non cyber portfolio. For that, we use a stochastic epidemiological model spreading in a network structure.

From déterministic to stochastic épidemiological models

As the same way as a biological virus spreads through a population, potentially leading to an epidemic, malwares can generate accumulation episodes such as Wannacry or NotPetya in 2017. Compartmental epidemiological models are adapted to describe the spread of a virus among a global population. One of the most famous being the SIR model, standing for Susceptible, Infected and Recovered (or removed). The simplest way to represent the evolution of the population through the three states is using an ODE (Ordinary differential equation) system (the deterministic model) where S, I and R counts the number of individuals in each state, and N is the overall number of individuals:

$$\begin{aligned}
\frac{dS(t)}{dt} &= -\beta I(t)\frac{S(t)}{N}, \\
\frac{dI(t)}{dt} &= \beta I(t)\frac{S(t)}{N} - \gamma I(t), \\
\frac{dR(t)}{dt} &= \gamma I(t).
\end{aligned}$$

The parameters β and γ represent respectively the infection and the recovery rates. As explained in Kiss et al., 2017 β is the rate at which infected individuals make contacts (potential infection), so

the number βI represents the total number of infectious contacts. But, among all contacts made, only a fraction $\frac{S}{N}$ are susceptible individuals.

As explained in Fahrenwaldt et al., 2018 a network structure among the population has a direct impact on the diffusion of the virus. Adding a graph allows us to add heterogeneity in the diffusion of viruses (in our case malwares) since their spreading speed varies according to the different classes of the population. This is achieved by using weights on the graph's edges.

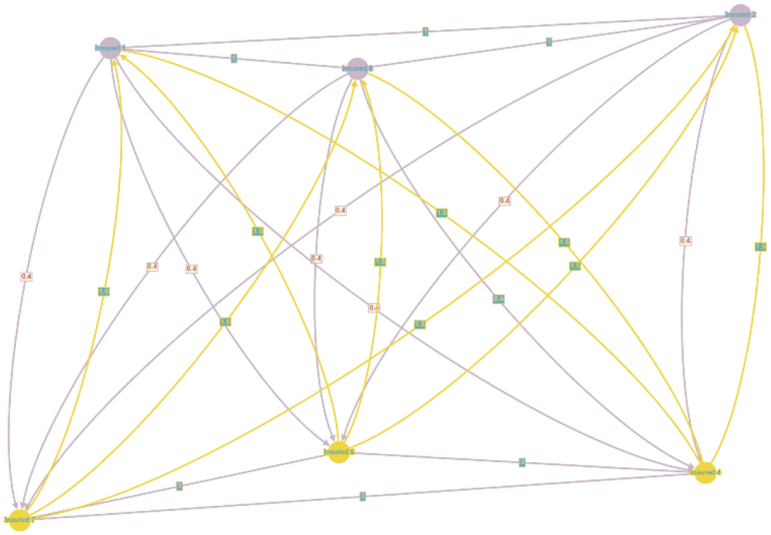


Figure 10 : Example of an homogeneous weighted graph for two classes.

As we can see above, yellow edges represent the weights at which nodes from the yellow class will infect nodes from the grey class. The weighted edges within the same class represent the infection weights between nodes within the same class.

For modeling the diffusion of the virus through a network we use the continuous Markov-process model. So, for N nodes, we denote by $E_s(t)$ the state at which the node s is at the time t. Hence, we have $E_s(t) \in \{S, I, R\}$ and,

$$E_{s_i}(t) : S \to I \quad \text{with rate} \quad \beta \sum_{s_j \in S} a_{ij} \mathbb{1}_{E_{s_j}(t)=I},$$

$$E_{s_i}(t) : I \to R \quad \text{with rate} \quad \gamma.$$

Where a_{ij} usually takes values in $\{0, 1\}$ with 1 representing the case where a contact exists between nodes s_i and s_j and 0 the case where there is no possible contact between the two nodes. Moreover, as we want to add weights to edges, we allow a_{ij} to take other values than $\{0, 1\}$. As for the deterministic model, we can see that the recovery of an infected individual only depends on the value of γ.

The algorithm used for computing the model is the Event-Driven fast SIR described in Appendix A.1.2 of Kiss et al., 2017.

Which network should be used ?

The idea behind adding a network structure among policyholders is to better reflect the environment in which the malware will spread. Furthermore, insurers will be able to determine which classes (sectors for example) are more likely to be infected or not. To better illustrate this, we consider the

sector network introduced in HILLAIRET et al., 2021. It is constructed using the exchanged volumes across sectors from an OCED study. We proportionate it according to one sector, it is what we call a reference sector. In our modeling, the latter is the mining sector meaning that the weights a_{ij} is equal to 1 if the policyholder s_i and the policyholder s_j belong to the same sector, see Figure (11).

We must keep in mind that the network structure can be calibrated using underwriting information such as partnerships, business relationships, or any other relevant information. In fact, sectors could be replaced by classes representing a more complex way of quantifying connectivity among policyholders.

Sectors	Mining	Manufacturing	Energy	Construction	Services
Mining	1	4,61672	0,7082	2,25079	1,9795
Manufacturing	0,0994	0,83123	0,04259	0,17035	0,55363
Energy	0,21293	0,58359	0,90063	0,23659	0,71293
Construction	0,02997	0,10726	0,01104	0,22239	0,14353
Services	0,00473	0,06624	0,00631	0,02681	0,25394

FIGURE 11 : Network weights according to the different sectors.

In Figure (11) the diagonal coefficients represent the connectivity within policyholders of the same sector. The rest, illustrate how members of different sectors are connected. Moreover the coefficients will be used for constructing the adjacency matrix of the network (i.e., the values of a_{ij}). So, each policyholder is connected to the others but with different weights.

The matrix is not symmetric, this means that some sectors are better defended or, on the contrary, some could be used as a vector to spread the virus.

The Toy Portfolio

In order to study the impact of cyber accumulation threat in non cyber framework, we consider a portfolio of professional multi-risk insurance having some heterogeneity in the wording due to, for example, the year of underwriting.

Each policyholder belongs to a sector

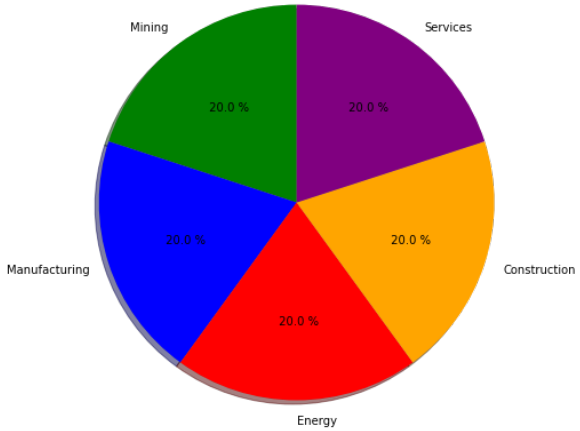


FIGURE 12 : Proportion of policiholders in each sector for the toy portfolio.

Moreover, the number of policyholders by sector is equally divided and described in Figure (12). The total number of policyholders in the toy portfolio is 1000.

Using the exposure of each policyholder

In the Toy Portfolio, beyond the information of the sector per insured, we also have the exposure for each coverage included in the policy. By exposure we mean the difference between (if present) the sublimit and the franchised. It is meant to represent the real compensable amount by the insurer for each coverage and each policyholder. A substantive aspect of our modeling is to be able to quantify the silent cyber exposure.

Assessing silent exposure

We want to evaluate cyber accumulation risk in a non cyber portfolio. To do so, we evaluate the silent exposure according to the four steps described in Figure (13). These steps are inspired by the IFoA's silent cyber assessment framework, see Cartagena et al., 2020.

Figure 13 : The four steps for assessing silent cyber.

The four steps "only" provide us silent exposure

The four steps of the framework illustrated in the above figure provide us only with the silent exposure. In our case, we aim to have the distribution of the number of the attacked nodes and that of costs over time. It is the reason why we mix steps 3 and 4 with the epidemiological model presented in the first section. To do so, each policyholder infected, will activate a guarantee according to a silent rate. The latter represents how often a coverage misses affirmative or exclusive cyber clauses. It is determined in step 2 of Figure (13). For instance, in a professional insurance portfolio where business interruption is covered, a 20% silent rate means that among 100 infected policyholders in our portfolio, 20 may trigger the compensation.

Evaluating the silent rate is one of the most critical parts of the modeling. Indeed, a wrong assessing will result in a wrong evaluation of the final potential loss. To better estimate this parameter, it is

necessary to work with underwriting and legal services. Moreover, some features of NLP modeling could be used to facilitate and accelerate the estimation.

In our Toy Portfolio, the silent rate varies for each coverage, and it is already given with the other information in the portfolio.

Evaluating scenerios

Generating scenarios for assessing potential losses is a commonly used approach in insurance. It is used for example to assess flooding natural disasters losses. In Figure (14), we have some examples of scenarios that could trigger silent coverages in specifics policies, see Marsh, 2020.

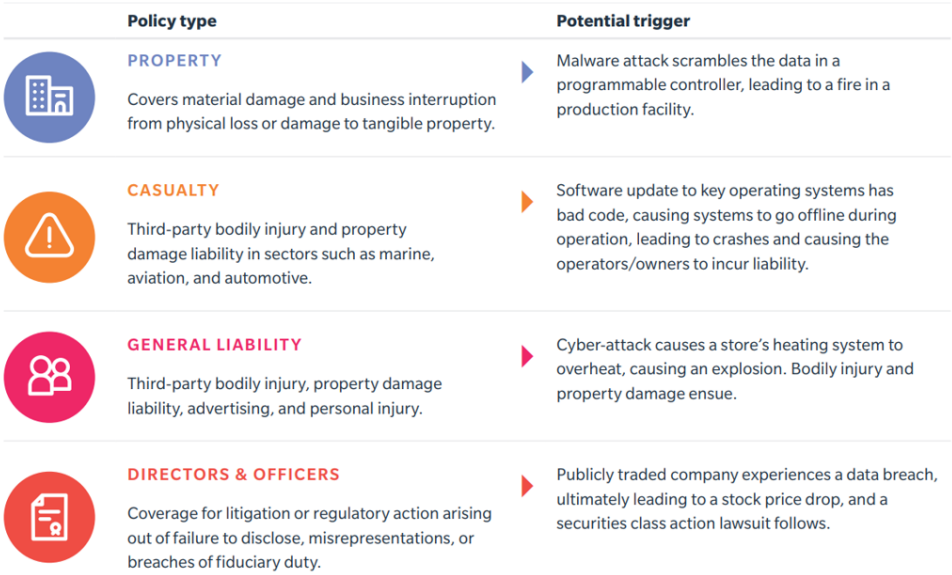


Figure 14 : Some scenarios that could trigger silent coverages (Marsh, 2020).

In our modeling we consider that each policyholder carries a silent risk on each coverage according to the silent rate previously assessed. Hence, we define a scenario as a list of triggerable coverages. Intending to add variability to scenarios, one could link each coverage to a trigger probability. Some realistic scenarios can be found in Lloyd's, 2022. We focus on this paper on a business interruption scenario.

Business interruption scenario

Description

In this scenario we consider a ransomware causing a business interruption from the time of infection until the system is restored. Ransomwares are a common type of malwares that encrypt data or block entire systems and restore everything once a ransom (usually in bitcoins) is paid. In this scenario we consider that the ransomware will only block the informatic system and thus cause the business interruption.

The infection parameter β (introduced in the first section) will be fixed at 0.01 and γ, the recovery parameter, at 1. An order of magnitude (very approximate), is that with this rate of contagion an infected (insured) node will contaminate about 1% of the susceptible nodes to which it is linked and

will recover in about a day. This infection percentage is subject to change due to the weighted graph we use in our model, but the recovery will not vary since it doesn't depend on the network structure.

Modeling the claims

We consider that the loss for one policyholder is an increasing function over the time spent paralyzed by the ransomware. For each day spent in business interruption, we'll generate a random gamma distribution depending on the policyholder's sector to simulate a compensable amount. The compensable amount is truncated by the exposure marked in the portfolio meaning that the compensation can never go beyond the exposure.

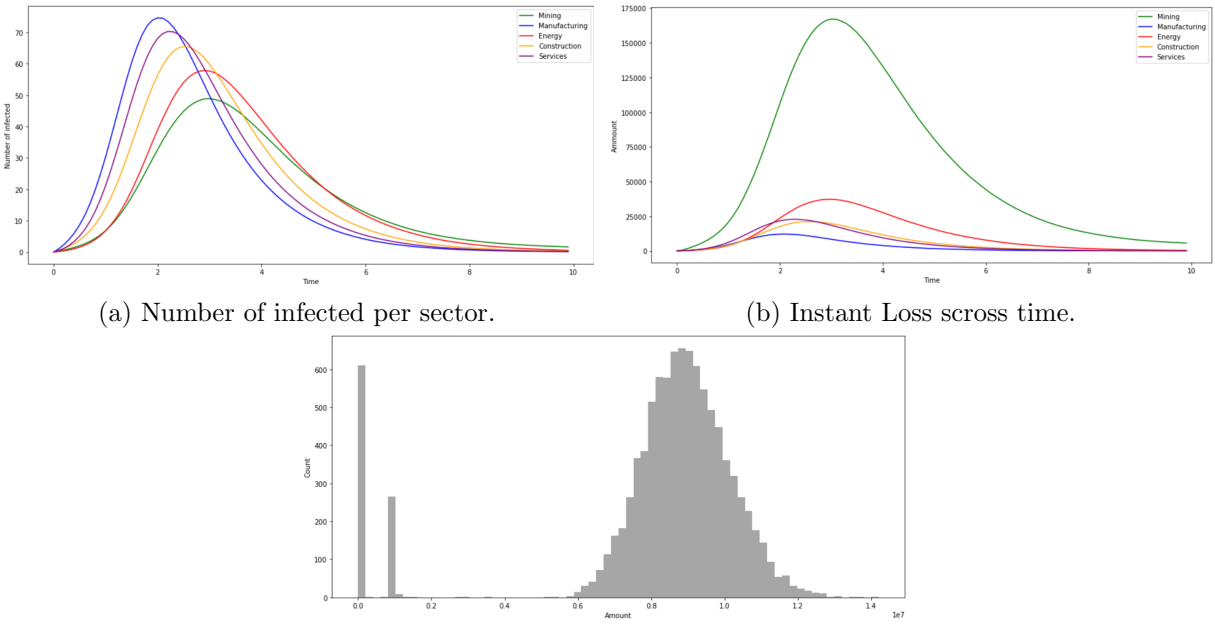
TABLE 2 : Parameter for the gamma distibutions modeling the daily cost according to policyholder sector.

Sector	a	b	Expectation	Variance
Mining	200 000,00	0,5	100 000,00	50 000,00
Manufacturing	10 000,00	0,5	5 000,00	2 500,00
Energy	40 000,00	0,5	20 000,00	10 000,00
Construction	20 000,00	0,5	10 000,00	5 000,00
Services	20 000,00	0,5	10 000,00	5 000,00

We can see in the previous table that the construction and services sectors have the same daily cost distribution. Recall that for a random variable $G \sim \mathcal{G}(a, b)$ we have $\mathbb{E}[G] = a \times b$ and $\mathbb{V}ar[G] = a \times b^2$. Furthermore, the network structure will be the one introduced in first section. As mentioned before, since we are dealing with silent cyber, not all infections lead to the activation of the business interruption coverage. In this example, the silent rate, the rate at which an infection leads to a compensation, is 32%.

Results on the toy portfolio

We can see in the figure below how the manufacturing sector is the first one to reach its infection peak. This is directly linked to the fact that manufacturing is the sector with the highest a_{ij} weight according to the mining sector. The latter infects all other sectors, see Figure (11).



(a) Number of infected per sector.

(b) Instant Loss scross time.

(c) Loss distribution 10 days after the infection.

FIGURE 15 : Evolution of the number of infected and the instantaneous loss by sector over time.

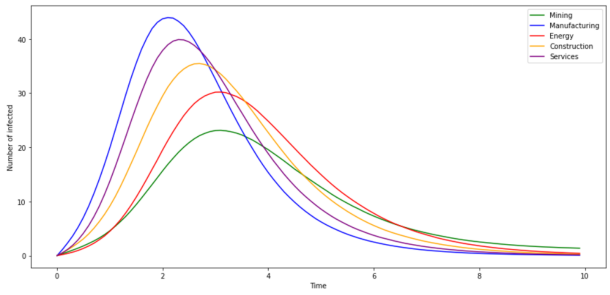
Also, we note in Figure (15c), that the most exposed sector to huge daily compensations is mining. This indicates that limited contaminations, Figure (15a), don't necessarily lead to reduced claims costs.

It is noteworthy that there exists a stationary state where the malware is eradicated. In this case, no infected remains and thus, no more contaminations can be generated, this leads to an accumulation near to 0, as we can see in Figure (15). The latter is obtained using 10 000 simulations up to 10 days. With these parameters and portfolio characteristics, the mean value of the cumulative loss is **8 204 785€**.

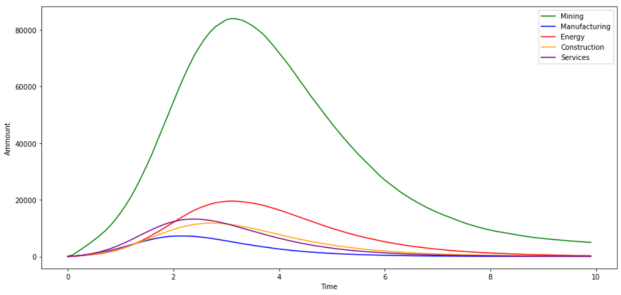
Increasing insurer's "intervention" capacity

Increasing the intervention capacity is modeled by increasing the recovery parameter γ. This will lead to a faster recovery of infected policyholders. During the Wannacry cyber crisis in 2017 (MOHURLE et PATIL, 2017), although the EternalBlue cyber-attack exploit was patched (MS17-010) for windows users in March 2017, many windows users were still vulnerable during the crisis, and some even one year later, see VLCEK, 2018.

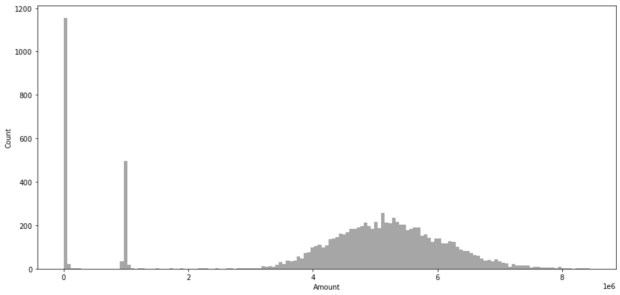
Some prevention measures could be implemented to enhance the insurer's intervention efficiency such as vulnerabilities patching and, the awareness of policyholders to the system updates. Our model does not consider prevention impacts, such as in HILLAIRET et al., 2021, where a parameter of reaction to cyber environment is introduced, making policyholders more cautious when attacks are detected.



(a) Number of infected per sector.



(b) Instant Loss scross time.



(c) Loss distribution 10 days after the infection.

FIGURE 16 : Evolution of the number of infected and the instantaneous loss by sector over time.

In Figure (16a), we can see how increasing the recovery parameter decreases the peak of the overall paths for all sectors. As it could be expected, increasing the recovery parameters leads to less infections and thus to less costs in all sectors, see Figure (16c). Calibrating this parameter could be done using the expected recovery time which is equal to $1/\gamma$.

Increasing the recovery parameter might have a cost for the insurer. By adding this information, we could compare the benefits of increasing the intervention capacity against its cost. Increasing the recovery parameter γ to 1.5 allows us to decrease the overall cumulative loss by half : **4 400 217 €**.

Modifying the policyholder's sector distribution

In Figure (12), the toy portfolio has a homogeneous distribution across all sectors. But as we can see in the results in Figure (15c), the most expensive sector is mining. We now modify this distribution by reducing the share of the mining sector, see Figure (17). We analyze how it impacts the spread of the malware across time for each sector, see Figure (18a) and the associated loss, see Figure (18c).

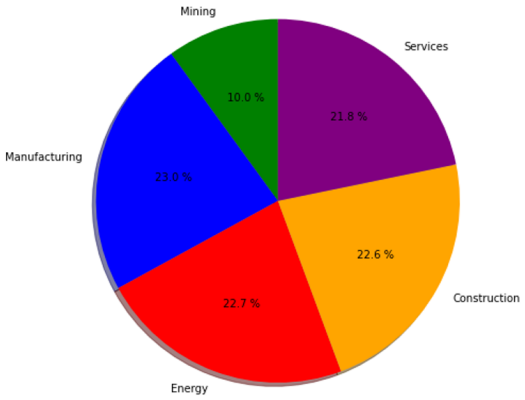
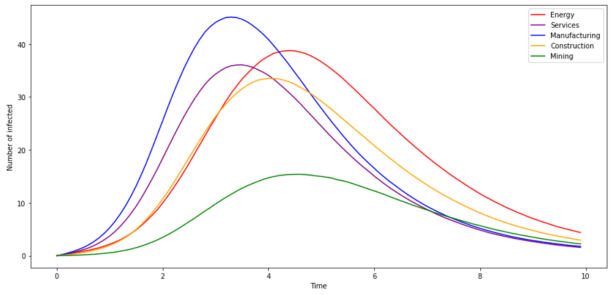
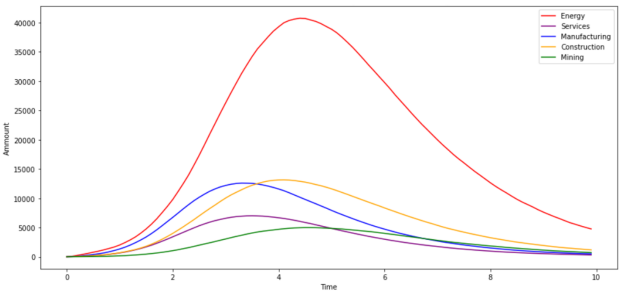


FIGURE 17 : Number of policyholders per sector in the new portfolio.

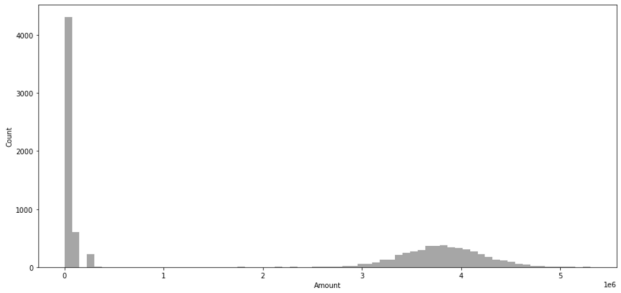
As we can see in Figure (17), the mining sector now represents only 10% of the overall sectors. This has direct impact on the spread of the malware. In Figure (18a), the fastest contaminated sector is still manufacturing, and the slowest sector is still mining. This is due to the high contagion between mining and manufacturing, see Figure (11) . In Figure (18c), the most expensive sector is energy. The diffusion's speed is conditioned by the sector.



(a) Number of infected per sector.



(b) Instant Loss scross time.



(c) Loss distribution 10 days after the infection.

FIGURE 18 : Evolution of the number of infected and the instantaneous loss by sector over time.

By knowing how sectors are linked between them, we could establish the best sectors distribution for the portfolio. We are thus able to deduce some guidelines to limit the potentials losses in case of a cyber accumulation scenario.

By reducing the number of mining policyholders in the portfolio, we decrease the overall loss to **1 872 643 €**. This strong decrease of the final loss is due to the decrease of the number of insured within the mining sector but also to the fact that the insured of this same sector are the most expensive to indemnify.

Concludying remarks

In the last few years, silent cyber has been one of the big concerns in underwriting and legal services due to the complexity of its assessment.

More generally, evaluating cyber accumulation risk can be done using epidemiological models. We present in this study a network structure that can be added to model the environment in which the virus will spread. It is the sector in our case, but other classifications can be used to describe interactions among policyholders.

These tools don't only allow us to calculate potential losses but can also provide guidelines for insurers to better manage their provisions, portfolios, silent assessment's priorities. Moreover, more complex scenarios could be implemented, but more underwriting information would be necessary to keep the realism of the study.

Remerciements

La rédaction de ce mémoire clôture mon cycle de Master et donc également une étape importante de ma vie. De ce fait, je tiens à remercier toutes les personnes qui, ont contribué (directement ou indirectement) au succès de mon stage, qui m'ont aidé lors de la rédaction, et plus généralement, qui m'ont soutenu ces dernières années.

Je souhaiterais dans un premier temps remercier, ma tutrice d'entreprise Mme Yousra Cherkaoui-Tangi, pour sa patience, sa bienveillance, sa disponibilité mais surtout, pour ses conseils et remarques qui ont toujours été d'une pertinence exemplaire.

Je tiens à remercier l'équipe R&D de Milliman, en particulier M. Alexandre Boumezoued, pour sa bienveillance, son professionnalisme, et les précieux conseils qui m'ont guidé tout le long de mon stage. De la même façon, je remercie toute l'équipe pédagogique de l'université Paris Dauphine, en particulier M. Christophe Dutang, M. Quentin Guibert et M. Adrien Suru, pour la qualité des enseignements et de l'accompagnement académique, sans lesquels, la rédaction de ce mémoire n'aurait pas été possible.

Je remercier également l'équipe VEGA de Milliman, pour leur bonne humeur, leur bienveillance, leur accueil chaleureux et leurs précieux conseils.

Je tiens à témoigner toute ma reconnaissance aux personnes suivantes, pour leur aide dans la réalisation de ce mémoire :

Mme. Eve Titon, qui a su nourrir notre réflexion autour des modèles épidémiologiques pour la modélisation du risque d'accumulation cyber.

Mme. Hinarii Pichevin, qui à plusieurs reprises m'a apporté les éléments nécessaires à la construction de portefeuilles fictifs.

M. Mohamed Benkhalfa, qui au-delà de s'intéresser à nos travaux, a su nous orienter pour l'évaluation du risque cyber silencieux.

Je profite de l'occasion pour remercier les personnes et institutions qui ont indirectement contribué à la réalisation de ce mémoire :

Les équipes pédagogiques de l'INSA de Toulouse et de Lyon qui m'ont permis d'acquérir les compétences nécessaires afin d'accéder au double diplôme avec l'université Paris Dauphine.

M. Anthony Réveillac (INSA Toulouse) qui, m'a donné l'opportunité de découvrir les mathématiques actuarielles, a éveillé mon goût pour la recherche, et m'a soutenu dans mes recherches de stage puis plus récemment de doctorat.

Mme. Caroline Hillairet, pour sa confiance, son soutien permanent et ses précieux conseils, sans lesquels je n'aurais pas su orienter mes projets professionnels et académiques.

M. Eric Dumont (INSA Lyon), sans qui je n'aurais certainement pas la confiance et le raisonnement que j'ai aujourd'hui.

Finalement, les pages viennent à manquer pour partager à quel point je remercie ma famille, mes proches et mes amis qui, m'ont toujours apporté leur soutien inconditionnel et m'ont fait comprendre que "la gratitude c'est la mémoire du cœur".

Table des matières

Résumé **3**

Abstract **4**

Note de Synthèse **5**

Synthesis note **15**

Remerciements **25**

Table des matières **27**

Introduction **29**

1 Le risque cyber **31**

1.1 Introduction 31

1.2 Le défi de la modélisation du risque cyber 37

1.3 Assurer le risque cyber 41

2 Modélisation de la composante systémique **47**

2.1 Les modèles de pandémie avec structure de réseau 47

2.2 Etat de l'art du risque d'accumulation appliqué au risque cyber 63

2.3 Étude pratique du SIR 69

3 Application au silent cyber **81**

3.1 Évaluation de l'exposition cyber silencieuse 81

3.2 Modélisation du risque d'accumulation 95

3.3 Pour aller plus loin 106

Conclusion **111**

Bibliographie **112**

A Compléments relatifs aux éléments présentés dans le mémoire. **119**

A.1 Auto-corrélation des évènements Cyber 119

A.2 Compléments pour la représentation des graphes 123

A.3 Brève analyse du modèle SIS sur réseau 125

A.4 Analyse des pics des trajectoires au cours du temps 128

A.5 Pseudo-codes des algorithmes de Gillespie et event drive 133

A.6 Annexe 1 de la directive du Parlement Européen et Conseil de l'Union Européenne, 2009 135

A.7 Exemples de clauses d'exclusions 139

A.8 Etapes détaillées du cadre d'évaluation de l'IFoA 141

Introduction

Engrené dans la révolution numérique, le développement d'internet et les progrès technologiques ont bouleversé notre société depuis les années 90. Aujourd'hui, nos achats (e-commerce), notre travail (télétravail), notre vie sociale (réseaux sociaux) et bien d'autres aspects qui régissent le fonctionnement de notre société dépendent du numérique. Avec ces transformations, de nouveaux risques ont également vu le jour : retard du développement intellectuel chez les enfants, cyberharcèlement, cybercriminalité, cyberguerre... Ainsi les risques cyber menacent la santé des individus comme la pérennité des entreprises ou des états.

Du grec *kubernân* signifiant gouverner, le préfixe "cyber" désigne l'utilisation du réseau internet. Ainsi le risque cyber désigne plus amplement les risques provenant des systèmes informatiques et du réseau internet.

Face à ces nouveaux risques, les premiers contrats d'assurance cyber dédiés aux entreprises sont apparus à la fin des années 90 aux Etats-Unis. Depuis le marché n'a cessé de se développer afin de couvrir ces risques croissants. Cela fait plusieurs années, que le risque cyber se hisse à la première place des risques émergents chez les assureurs.

L'une des difficultés que pose la modélisation actuarielle du risque cyber est la prise en compte du phénomène d'accumulation. En effet, en 2017 la plus grande cyberattaque jamais observée a eu lieu. Avec plus de 300 000 ordinateurs infectés les dégâts économiques s'estiment à plusieurs milliards de dollars, répartis dans plus de 150 pays (Hillairet et Lopez, 2022). Les victimes n'ont pas de profils particuliers, entreprises, particuliers, hôpitaux... le rançongiciel infecte, attaque et se propage sur tous les ordinateurs Windows vulnérables. Le risque d'accumulation est important à prendre en compte puisqu'il s'oppose au principe fondateur de l'assurance : la mutualisation des risques.

Au delà du phénomène d'accumulation, le risque des polices silencieuses est également à prendre en compte pour les contrats non cyber. En effet, avec l'essor des nouvelles technologies, certaines garanties proposées dans les polices non cyber peuvent désormais être activées par des évènements cyber (Incendies, Pertes d'Exploitations, Responsabilité Civile ...). Ainsi, un déséquilibre se crée entre les risques que couvrent réellement la garantie et la prime perçue par l'assureur.

Ce mémoire s'intéresse au risque d'accumulation cyber appliqué au cas du cyber silencieux. En conséquence, nous cherchons à savoir comment une prise en compte du risque d'accumulation cyber permet une meilleure résilience des portefeuilles non-cyber.

Afin de répondre à cette question, nous commencerons dans le chapitre 1 par présenter plus amplement le risque cyber, de la réglementation aux problématiques de modélisation actuarielle. Puis dans le chapitre 2, en se basant sur des modèles d'épidémiologie et de réseaux, nous montrerons en quoi cette modélisation permet de générer des phénomènes d'accumulation. Finalement, en adaptant notre modélisation au cyber silencieux, nous illustrerons dans le chapitre 3 comment évoluent les pertes lors d'un évènement d'accumulation cyber sur un portefeuille d'assurance.

Chapitre 1

Le risque cyber

Que ce soit à travers un article de faits divers ou par la réception d'un mail frauduleux, les risques liés à la cybercriminalité (dénommé risque cyber dans la suite) évoquent une vague notion de danger à la grande majorité de la population. Dans ce premier chapitre, nous éclaircirons les différentes notions autour du risque cyber et illustrerons l'environnement assurantiel dans lequel il évolue. Nous commencerons donc par présenter sa nature, comment il agit, quelles sont les cibles et la réglementation autour de ce risque. Puis nous présenterons ses caractéristiques techniques d'un point de vue assurantiel. Finalement, nous introduirons le cyber silencieux par un exemple concret.

1.1 Introduction

D'après le site du gouvernement sur la prévention des risques majeurs (Prévention des risques majeurs, 2022) "une cyber-attaque est une atteinte à des systèmes informatiques réalisée dans un but malveillant. Elle cible différents dispositifs informatiques : des ordinateurs ou des serveurs, isolés ou en réseaux, reliés ou non à Internet, des équipements périphériques tels que les imprimantes, ou encore des appareils communicants comme les téléphones mobiles, les « smartphones » ou les tablettes"

Selon la cartographie prospective des risques (Fédération française des assureurs, 2022), les cyberattaques représentent pour la 5ème année consécutive le premier risque pour les assureurs. Cette étude qualitative est soutenue par le rapport annuel de l'Agence Nationale de la Sécurité des Systèmes Informatiques ANSSI, 2020 qui conclut par une forte augmentation des cyberattaques. En effet les signalements par rançongiciel ont plus que triplé entre 2019 et 2020. Plus récemment, à la suite du premier conseil des ministres de la rentrée 2022, le président de la république Emmanuel Macron, n'a pas manqué de souligner l'importance des "nouveaux risques", "comme le risque cyber" (BFMTV, 2022).

Dans la section suivante, nous allons présenter le mode opératoire du risque cyber, quels sont les acteurs, leurs motivations et les cibles.

1.1.1 Les types d'attaques

Selon le site du gouvernement sur la prévention des risques majeurs (Prévention des risques majeurs, 2022), le risque cyber est décomposé selon quatre types de risques : la cybercriminalité, l'atteinte à l'image, l'espionnage et le sabotage. Le contenu suivant est basé sur les informations

disponibles sur ce même site.

1.1.1.1 La cybercriminalité

L'objectif de l'attaque est d'obtenir des informations personnelles pour les revendre ou les exploiter. On distingue notamment :

- Le *phishing* ou hameçonnage : cette attaque consiste à usurper l'identité d'un tiers (personne ou entreprise) dans le but d'obtenir des renseignements personnels comme des identifiants bancaires.
- Le *ransomware* ou rançongiciel : il s'agit des logiciels informatiques malveillants les plus en vogues ces dernières années. L'objectif est de chiffrer les données puis de demander un envoi d'argent en échange de leur déchiffrement.

1.1.1.2 L'atteinte à l'image

L'objectif est de déstabiliser en portant une atteinte directe à l'image de l'entreprise, administration ou autre. Les attaques se réalisent via :

- Le déni de service (ddos) : cette attaque sature une ressource particulière du système informatique de la cible jusqu'à sa défaillance. Ainsi, le site et les services qu'ils proposent deviennent inaccessibles.
- Le *defacement* ou défiguration : souvent lié à des fins idéologiques ou politiques, l'objectif est de modifier un site internet tant dans son contenu que dans son apparence. Ce qui porte atteinte directement à l'intégrité du site.

1.1.1.3 L'espionnage

Comme son nom l'indique, l'objectif est de tirer des renseignements de la cible pour des fins économiques, scientifiques ou politiques. Les techniques employées sont généralement assez sophistiquées et ciblées, comme :

- Le *watering hole* ou point d'eau : les attaquants commencent par choisir un site "appât" sur lequel les victimes (généralement d'une même organisation) vont se rendre. Ainsi les cybercriminels déposent sur ce site un virus qui va infecter les cibles une fois qu'elles s'y seront rendues.
- Le *spearphishing* ou hameçonnage ciblé : de la même façon que le *phishing*, les attaquants usurpent l'identité d'une personne physique ou morale. Cependant, le *sparephising* cherche à contaminer le système d'une personne en particulier pour lui soutirer des informations sans qu'elle s'en aperçoive.

1.1.1.4 Le sabotage

Le dernier type de risque relié au cyber est le sabotage. Les moyens employés par les attaquants sont très nombreux et englobent souvent plusieurs méthodes. L'objectif derrière un sabotage peut avoir

diverses raisons mais les conséquences d'un tel acte peuvent facilement mettre en péril l'économie d'une organisation, la santé des personnes et même le fonctionnement d'une nation si le secteur touché est d'une importance majeure.

1.1.2 Motivations, cibles et points d'entrées

1.1.2.1 Les acteurs et leurs motivations

Comme nous l'avons vu précédemment, les différentes attaques n'ont pas toutes les mêmes finalités. Ainsi, les acteurs à l'origine de ces cybercrimes n'ont pas tous les mêmes intentions. Comme l'indique le rapport Verizon, 2022 qui fait un état des infractions cyber, les principales motivations des cybercriminels restent financières ou personnelles, c'est à dire, la recherche d'un enrichissement financier ou personnel (informations ou autres types gains). Cependant d'autres raisons peuvent motiver ces individus à passer à l'acte. En effet la société et son fonctionnement étant de plus en plus dépendante des systèmes informatiques, des militants ou organisations étatiques n'hésitent pas user de ces méthodes pour la déstabiliser, par exemple lors des élections présidentielles. Le fort impact que peuvent avoir les attaques cyber motive également certains états à en faire un usage militaire. Finalement, le *Hacking* véhicule parmi les personnes qui le pratiquent, un certain challenge au sein de la communauté, et les attaques peuvent être motivées par le prestige que leur succès engendrerait.

Comme nous le verrons plus tard dans ce chapitre, les dommages que peuvent générer les cyber attaques sont très importants et peuvent directement toucher à la vie des citoyens. Les décisions gouvernementales sur le plan militaire doivent donc permettre de garantir la souveraineté des systèmes informatiques publics comme privés. Ainsi en 2009 est créé l'Agence Nationale de la Sécurité des Systèmes d'Information (ANSSI) qui est directement rattachée au Secrétariat Général de la Défense et de la Sécurité Nationale (SGDSN).

L'image du *hacker* solitaire qui avec un simple ordinateur infiltre le système informatique d'une entreprise depuis un café est loin de représenter la réalité. D'après un article qui dresse le panorama de la menace informatique de l'année 2021 (ANSSI, 2022), l'agence constate une forte spécialisation et professionnalisation des cybercriminels. Ainsi s'est créé un vrai réseau économique associant fournisseurs de services spécialisés dont les membres collaborent plus ou moins étroitement en fonction des opportunités et des objectifs du moment. Toujours d'après le même rapport, l'ANSSI constate la vente de rançongiciels en tant que services (*Ransomware-as-a-Service - RaaS*). En effet ils font intervenir plusieurs acteurs spécialisés dans des domaines particuliers allant de l'analyse des vulnérabilités des entreprises, à l'expertise des protocoles de réseaux.

Les acteurs à l'origine des cyber attaques ont donc des profils très variés et disposent de plusieurs méthodes pour atteindre leurs objectifs. Cependant, selon un autre rapport (ANSSI, 2021), l'agence a constaté en 2020 une hausse de signalements de 255% par rapport à l'année 2019. Avec trois tendances notables : le *RaaS* présenté en amont, le *Big Game Hunting* qui consiste à sélectionner des cibles (ou des données particulières) de haute valeur, et la double extorsion qui extrait les données en plus de leur chiffrement préalable.

1.1.2.2 Les cibles

Selon un article qui donne l'image commune de la situation des cyber risques entre l'Allemagne et la France, représentés par la BSI (équivalent à l'ANSSI) et l'ANSSI respectivement, aucun secteur ne serait épargné par le risque cyber (ANSSI et BSI, 2021).

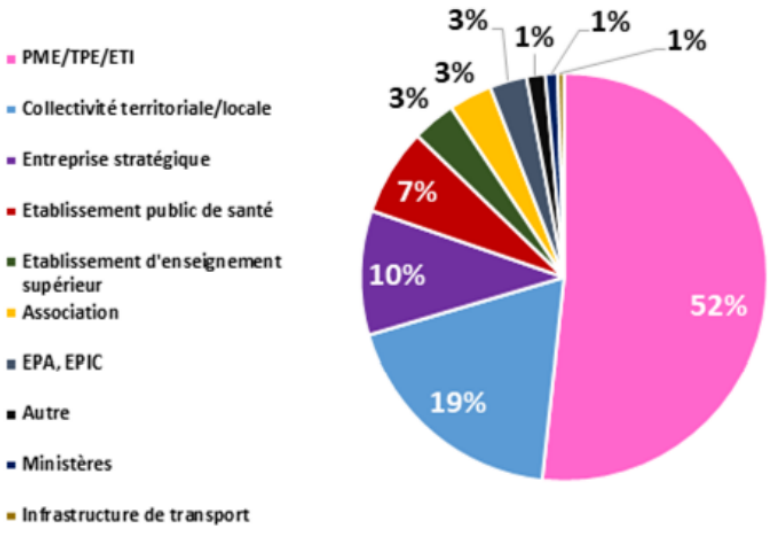


FIGURE 1.1 : Répartition des entités victimes d'attaques par rançongiciel dans le cadre des incidents traités par l'ANSSI en 2021 (ANSSI, 2022).

Nous observons sur la figure (1.1) que les PME/TPE/ETI représentent la catégorie de victimes les plus à risque vis à vis des attaques par raçongiciels.

1.1.2.3 Les points d'entrées

En 2022, un groupe de travail mis en place par le Forum des Compétences (FORUM DES COMPÉTENCES et EGERIE, 2022), distingue trois catégories de points d'entrées dans les entreprises. La catégorie **physique/humain** concerne les attaques qui nécessitent, un accès physique au système d'information, ou passer par l'intermédiaire d'un individu ayant accès au système. Les attaques appartenant à la deuxième catégorie, **l'indirecte**, exploitent la chaîne de sous-traitance ou un utilisateur situé hors des locaux de l'entreprise. Finalement, pour les attaques de la catégorie **internet**, elles exploitent des failles et des vulnérabilités des systèmes d'informations qui sont directement exploitables depuis internet.

1.1.3 Des enjeux et des réponses législatives par plusieurs organismes

Nous allons à présent traiter deux aspects législatifs autour du cyber. Le premier, concerne des obligations importantes pour les entreprises en lien avec la cybersécurité. Le deuxième aspect traite les réponses législatives apportées par les états pour limiter la menace. Nous commencerons donc par présenter quelques points réglementaires importants puis nous présenterons les différentes institutions qui jouent un rôle à différents niveaux.

1.1.3.1 Quelques points de réglementation

Plus communément connu sous règlement général sur la protection des données (**RGPD**) mais Officiellement appelé règlement UE 2016/679 du Parlement européen et du Conseil du 27 avril 2016

relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données et abrogeant la directive 95/46/CE, est un règlement de l'Union Européenne qui s'applique depuis mai 2018. Malgré les quatre années de négociations législatives qui ont été nécessaires à sa rédaction les objectifs de ce règlement sont simples : protéger les personnes lors de traitements de leurs données personnelles tout en responsabilisant les acteurs derrière ces traitements. Ainsi, les citoyens maintiennent le contrôle de leurs données personnelles tout en disposant d'un cadre législatif pour sanctionner ceux qui leurs en empêcheraient. Une entreprise victime d'un rançongiciel s'expose donc à des sanctions par l'application du RGPD si des données personnelles sont touchées. La commission nationale de l'informatique et des libertés (CNIL) peut donc sanctionner l'entreprise pour divers motifs, par exemple, des dommages et intérêts.

D'après une enquête réalisée en 2016 par Bitdefender (Bitdefender, 2016), 32% des entreprises françaises seraient prêtes à payer les rançons à la suite d'une attaque cyber. Selon un autre rapport publié en 2021 (Hiscox, 2021), 49 % des entreprises françaises auraient subi une cyber attaque, dont 34% l'auraient déjà subie en 2020. Toujours d'après la même étude, 19% des entreprises françaises auraient payé la demande en rançon, faisant de la France, le troisième pays après l'Allemagne (21%) et les Etats-Unis (21%) payant le plus de rançons à la suite d'une attaque cyber. L'ANSSI expliquait ce haut chiffre du fait qu'un grand nombre d'assureurs remboursaient les frais de rançon dans les contrats cyber, incitant ainsi le paiement des rançons par les entreprises victimes. En mai 2021 Axa France (suivi par Generali début 2022) avait suspendu la commercialisation l'option "cyber-rançonnage" (Le Parisien, 2022). Le 7 septembre 2022, la Direction Générale du Trésor a publié, à la suite d'une concertation nationale sur l'assurance du risque cyber (5 juillet 2021), un article dont un axe vise à "clarifier le cadre juridique de l'assurance du risque cyber" et propose notamment le remboursement des rançons sous condition d'un dépôt de plainte dans les 48h (Direction générale du Trésor, 2022). Ce même jour, le ministre de l'intérieur et des outre-mer, Gérald Darmanin, présente au Conseil des ministres un projet de loi d'orientation et de programmation du ministère de l'intérieur (LOPMI). Cette loi tient compte de la recommandation faite par le ministère de l'économie et va donc encadrer le remboursement des cyber-rançons en exigeant notamment le dépôt de plainte de la victime sous 48h. Ces informations sont disponibles sur le site du gouvernement dédié à l'avancement du projet de loi (Vie Publique, 2022).

D'un autre côté, en juillet 2016 a été publiée la directive (UE) 2016/1148 du Parlement européen et du Conseil. Cette directive s'applique aux entreprises des secteurs de l'énergie, du transport, de la banque, des infrastructures des marchés financiers, de la santé, de l'eau et des infrastructures numériques. L'objectif de la directive est d'établir des exigences communes en matière de sécurité des systèmes de réseau et d'information au niveau de l'union européenne. Ainsi, la directive Network and Information Sécurity (**NIS**) a été adoptée par le Parlement et inscrite au journal officiel en février 2018 et établit donc certaines exigences de sécurité et des règles pour notifier les incidents numériques.

Avec l'essor de l'IoT (*Internet of Things* ou internet des objets), la migration vers le cloud, la forte augmentation du télétravail, de nouveaux enjeux en cybersécurité apparaissent. Ainsi le Parlement européen travaille en ce moment sur une nouvelle directive, NIS2, dans la continuité des principes établis par NIS. De nouveaux secteurs seraient impliqués par NIS2, avec de nouvelles obligations, comme par exemple, le signalement des incidents de sécurité dans les 24h (Negreiro et Del Mar, 2022).

Le cybersécurité est au cœur des débats et va continuer de faire bouger les lignes dans les années qui viennent. Mais ces transformations sont entre autres poussées par des organismes nationaux ou des états, qui interviennent à plusieurs niveaux : national, européen ou international.

1.1.3.2 Des acteurs important à la gestion du risque

L'*International Telecommunication Union* (**ITU**), est une agence spécialisée au sein de l'Organisation des Nations unies (ONU) qui parmi ses diverses fonctions, traite des problématiques de cybersécurité. Ils ont notamment construit en 2015, l'indice global de cybersécurité, *Global Cybersecurity Index* (GCU), qui apporte aux 193 membres de l'ITU une mesure de leurs engagements en cybersécurité. Ainsi cet indice permet aux états d'avoir une indication des actions à prendre pour poursuivre leurs avancées dans ce domaine.

En novembre 2001 a été signé le premier traité international visant à luter contre la cybercriminalité. Élaboré par le Conseil de l'Europe, il aborde les crimes commis sur internet comme le droit d'auteur, la fraude informatique, la pédopornographie, mais également les violations de la sécurité des réseaux. Son objectif principal est de "poursuivre une politique pénale commune visant à protéger la société contre la cybercriminalité". La convention prévoit donc plusieurs dispositions à cet égard comme par exemple "la mise en place d'un réseau 24h/24 et 7j/7 pour assurer une rapide assistance" entre les pays signataires. Aujourd'hui de nombreux pays externes à l'Union Européenne ont ratifié l'accord, cependant, des pays importants comme le Brésil, l'Inde ou encore la Russie, refusent sa ratification pour diverses raisons.

Nous avons présenté plus haut différents textes européens (RGPD, NIS et NIS2) qui traitent le thème de la cybersécurité. Cependant, il existe l'Agence de l'Union Européenne pour la cybersécurité (**ENISA** en anglais) créé en 2004, qui porte plusieurs rôles au niveau européen. Entre autres, elle conseille et assiste la Commission et les États sur le sujet, recueille et analyse les données d'incidents, suit l'élaboration de normes en matière de sécurité des réseaux et de l'information. Son rôle au sein de l'Union Européenne a été renforcé en 2019 par le règlement *Cybersécurity Act* qui établit un cadre européen de certification en cybersécurité dont l'ENISA est au cœur. Plus d'informations sur la politique européenne en matière de cybersécurité sont disponibles sur le site internet dédié (Commission européenne, 2022).

D'après le site de la diplomatie française (France Diplomatie, 2022) "au sein de l'Union européenne (UE), la France défend une vision ambitieuse et le concept d'autonomie stratégique numérique de l'UE ". Ce qui se décline en trois axes : l'axe technologique, qui passe notamment par le soutien de la recherche et du développement de pointe, l'axe réglementaire, qui vise à maintenir le juste milieu entre compétitivité, développement numérique et protection des citoyens, des entreprises et des Etats, et finalement, l'axe capacitaire, qui soutient les capacités de cyberdéfense des différents acteurs européens publics comme privés.

A l'échelle nationale, nous avons déjà présenté l'ANSSI qui "apporte son expertise et son assistance technique aux administrations et aux entreprises", avec un point d'attention particulier porté sur les opérateurs d'importance vitale (OIV). Un OIV est une organisation dont l'activité est jugée par l'Etat comme indispensable à la survie de la nation. L'ANSSI participe également à des évènements qui cherchent à confronter les points de vue entre académiciens et professionnels autour du risque cyber. Ce fut le cas lors de la conférence "Cyber-risque et assurance" de 2017, mais plus récemment lors du printemps de l'assurance, organisé par l'Université Paris-Dauphine, durant les tables rondes du risque cyber.

Concernant le secteur de la Banque, de la Finance et de l'Assurance, le Forum des Compétences regroupe des experts en sécurité des systèmes d'information (SSI). Leur objectif est de construire une compétence globale en SSI au sein de ces secteurs. Ils publient régulièrement des articles et des livrables, portés par plusieurs groupes de travail.

1.2 Le défi de la modélisation du risque cyber

Chez les assureurs, le risque cyber s'est associé aux difficultés de modélisation qui l'entourent : idée principalement véhiculée par le manque de données nécessaires à sa modélisation. Dans cette section nous allons présenter les principaux défis de modélisation que ce risque pose aux assureurs. Nous commencerons par rappeler les fondements de modélisation de l'assurance, puis nous présenterons les difficultés que présente cette modélisation sur le risque cyber.

1.2.1 Rappels de la modélisation en assurance

1.2.1.1 Le principe de mutualisation

L'activité d'assurance s'est construite sur un principe fondamental, celui de la mutualisation des risques. C'est ce principe qui distingue l'assurance, du pari. Cette section reprend de nombreux éléments présentés dans Hillairet et Lopez, 2022.

Mathématiquement, ce principe repose sur la loi (forte) des grands nombres (LGN) qui établit que la moyenne empirique d'une suite de n variables aléatoires indépendantes (X_i) d'espérance finie m, converge (presque sûrement) vers cette même espérance. Ainsi,

$$\frac{1}{n}\sum_{i=1}^{n} X_i \longrightarrow_{n\to\infty} \mathbb{E}[X_1] = m \quad (p.s.)$$

De cette façon, si nous possédons un grand nombre de sinistres (ou indemnités) aléatoires, lorsque les dommages sont indépendants et identiquement distribués (qu'ils suivent la même loi de probabilité), le montant moyen revenant à chaque assuré est (approximativement) égal à l'espérance mathématique d'un sinistre. Ce premier résultat permet de déduire la prime pure d'un contrat d'assurance.

Un deuxième résultat mathématique important permet de contrôler l'écart entre la charge totale $\sum_{i=1}^{n} X_i$ et la prime pure acquise nm. C'est le Théorème Central Limite (TCL) qui permet d'établir que la charge totale est donc asymptotiquement gaussienne. Il permet entre autres, de déduire des intervalles de confiance.

Découle de cette méthode l'approche traditionnelle fréquence-sévérité qui remplace le nombre de sinistres n par un processus de Poisson N_t qui va, au cours du temps, compter le nombre de sinistres survenus jusqu'à l'instant t. L'une des principales hypothèses de ce processus est l'indépendance des intertemps d'arrivés (notés W_i) des sinistres et de leur distribution exponentielle. Ce qui présuppose la survenance indépendante de chacun de ces évènements. Souvent la charge de sinistre totale se note donc $C_t = \sum_{i=1}^{N_t} X_i$ où donc X_i est le montant d'un sinistre survenu à l'instant T_i et les intertemps entre les sinistres sont notés $W_i = T_{i+1} - T_i$.

1.2.1.2 L'assurabilité

Selon le principe de mutualisation des risques énoncé précédemment, l'une des conditions nécessaires est que le risque doit être aléatoire. Dans le cas du cyber, ce caractère varie fortement d'un profil d'un assuré à un autre. De plus, il dépend également des moyens mis en œuvre par les cybercriminels. Le caractère aléatoire du risque doit donc être accompagné par le profil de l'assuré et de l'environnement cyber dans lequel il se trouve.

L'assurabilité d'un risque doit au-delà d'être mutualisable, présenter un aléa moral et une anti sélection faible (ou contrôlable). Rappelons que l'aléa moral est ce qui rend le divorce non assurable, en effet les assurés pourraient, au-delà des raisons conjugales, être poussés au divorce par la réalisation de la Police d'assurance. L'anti-sélection quant à elle réside dans le fait que l'assuré connaît mieux ses risques que l'assureur. Une façon de palier à l'aléa moral est d'exiger aux assurés le maintien des efforts en termes de cybersécurité, ce qui passe notamment par des partenariats entre Assureurs et professionnels de la cyber sécurité (comme Allianz et Thales), RIGAUD, 2022. Le problème de l'anti-sélection est lui plus dur à déceler puisque, indépendant du contexte assurantiel, il est très dur pour une entreprise de connaître par exemple la topologie de son réseau informatique. Cependant, une hygiène informatique minimale permet d'éliminer un grand nombre de risques et donc de garantir à l'assureur une certaine mesure du risque auquel s'expose l'assuré. Cette hygiène informatique peut par exemple passer par la sensibilisation des employés aux risques cyber et aux méthodes utilisées par les attaquants.

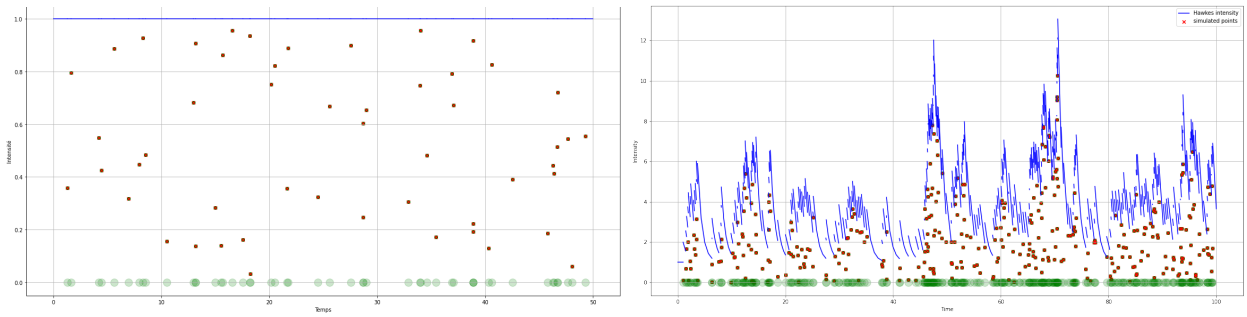
1.2.2 Une composante systémique

Comme nous l'avons vu précédemment il existe plusieurs types d'attaques qui composent les risques cyber. Cependant, certains cybercriminels exploitent des failles présentes dans les systèmes d'informations pour lancer leurs attaques. Or, ces systèmes étant interconnectés et employés par plusieurs utilisateurs, un phénomène de clustering et/ou d'accumulation peut apparaître.

1.2.2.1 Auto-correlation et clustering des évènements

La base de données Privacy Rights Clearinghouse (PRC) qui recense certains évènements cyber survenus aux Etats-Unis a permis de mettre en évidence l'auto-corrélation entre évènements cyber. Cependant, nous avons vu que l'approche traditionnelle fréquence-sévérité suppose l'indépendance entre les temps de survenance des sinistres. Ainsi, le processus de Poisson ne convient plus pour modéliser les évènements dont les temps de survenance ne sont plus indépendants. Une alternative efficace pour palier à ce problème est d'utiliser des processus de Hawkes, qui avec leur caractère auto-excitant, permettent de répliquer l'effet de clustering induit par l'auto-corrélation des évènements cyber. Sur les prochaines figures nous illustrerons l'intérêt de ces processus sans entrer dans le détail de comment est-ce qu'ils sont construits.

Nous pouvons voir en bleu sur la figure (1.2a) comment pour un processus de Poisson homogène l'intensité reste constante. Pour un processus de Hawkes, à noyau exponentiel sur la figure (1.2b), l'intensité dépend des évènements passés. Ainsi, le processus de Hawkes peut générer des évènements très proches les uns des autres comme le montrent les points verts sur la même figure.



(a) Poisson (b) Hawkes

FIGURE 1.2 : Intensité d'un processus de Poisson homogène et d'un processus de Hawkes à noyau exponentiel : En bleu est représenté l'intensité du processus au cours du temps, les points avec des croix rouges sont utilisés pour la simulation des processus, et les points verts translucides correspondent aux temps simulés par les processus.

Plusieurs articles scientifiques se sont intéressés à la modélisation du risque cyber à l'aide des processus de Hawkes comme BESSY-ROLAND et al., 2021 et un mémoire a amplement traité le sujet en 2019 (BESSY-ROLAND, 2019).

1.2.2.2 Un risque d'accumulation

De nos jours, les réseaux informatiques sont indispensables pour le fonctionnement de notre société, en particulier pour la communication des systèmes informatiques. En effet, lorsque nous utilisations par exemple une application de messagerie instantanée comme Teams, notre message subit plusieurs encapsulations qui permettent de répondre à divers besoins (intégrité, confidentialité, fiabilité, etc...). Plusieurs protocoles permettent de réaliser ces manipulations. Cependant ces systèmes peuvent contenir des failles et être exploités pour diverses finalités. Ces protocoles étant utilisés par plusieurs applications et par des systèmes informatiques tout autour du globe, un effet d'accumulation peut donc apparaître en cas d'attaque. Les failles peuvent concerner les protocoles de communication, mais comme tout produit conçu, d'autres failles pourraient être à la source de ces évènements d'accumulation.

Nous présenterons plus en détails dans la section (1.3.1.1) certains évènements historiques comme WannaCry qui en mai 2017, le rançongiciel de même nom à infecté environ 300 000 ordinateurs répartis dans plus de 150 pays. Pour Wannacry, le coût des rançons demandées restait faible (entre 300$ et 600$), mais l'accumulation de tels sinistres peut représenter un coût bien plus important pour l'assureur. De plus, les sinistres n'étant plus indépendants, le principe de mutualisation que nous avons présenté plus en amont s'écroule. Au-delà de mettre en péril la pérennité de l'assureur, ses capacités d'intervention auprès des assurés peuvent être dépassées par l'ampleur de l'évènement cyber. Car en effet, il est très commun que les contrats cyber incluent des prestations d'assistance afin d'assurer la reprise d'activité de l'entreprise touchée.

C'est ce risque qui est au cœur de notre étude. En effet nous chercherons à le modéliser et à l'appliquer à une catégorie particulière du risque cyber que nous présenterons en fin de chapitre. Le phénomène d'accumulation peut être modélisé en utilisant des processus d'épidémiologie. Dans le chapitre 2, nous présenterons le modèle le plus simple puis nous montrerons comment l'adapter pour inclure une structure de réseau.

1.2.3 D'autres défis

Jusqu'à présent, nous avons présenté les principaux défis que pose le risque cyber sur la survenance des évènements. Mais ce risque porte une autre contrainte de modélisation, celle des sinistres graves et celle du manque de données.

1.2.3.1 Des valeurs extrêmes

Nous avons vu que le principe de mutualisation suppose que la loi des sinistres est d'espérance finie. Pour pouvoir donc mettre en place une modélisation qui satisfasse ce principe, nous devons donc trouver une loi d'espérance finie qui représente bien les sinistres que l'on cherche à assurer. Au-delà du caractère technique pour assurer la mutualisation, si nous souhaitons créer un modèle économiquement viable, l'espérance a tout intérêt à rester la plus faible possible (car implique une prime attractive pour les assurés). Or, d'après AMRAE, 2022 qui dresse un aperçu des primes reçues et de la sinistralité du risque cyber en France, nous pouvons remarquer qu'un faible nombre d'évènement peuvent représenter un montant important sur la sinistralité globale. Ce qui complique la modélisation de la sinistralité.

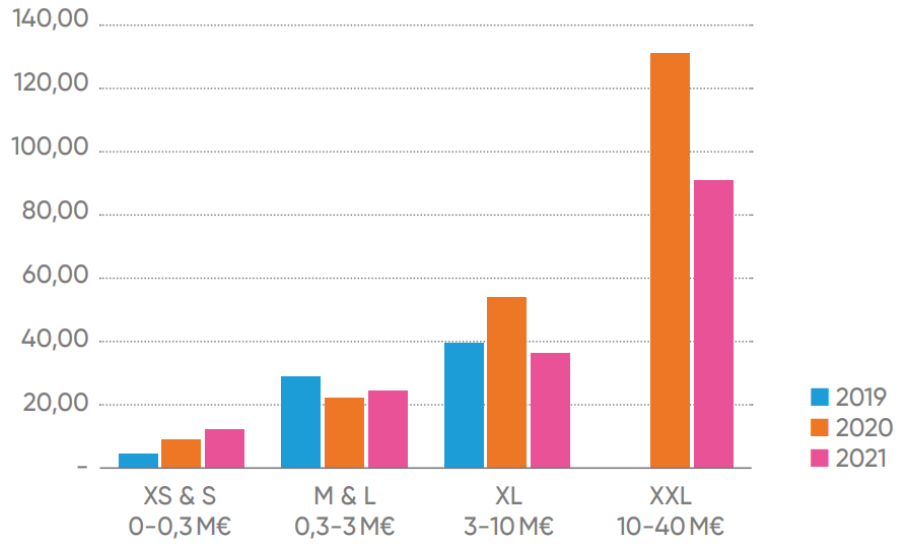


FIGURE 1.3 : Évolution de la distribution des sinistres, AMRAE, 2022.

Comme le précise l'étude, en 2021, "six sinistres XXL ont été indemnisés à hauteur de 90,6 M€, ce qui représente 55, 6% du volume total d'indemnisation".

Sur ce genre de données, la distribution de Pareto généralisée représente souvent une bonne alternative pour la modélisation. Cette loi dite à queue lourde, permet en effet de modéliser les sinistres graves. Nous renvoyons vers HILLAIRET et LOPEZ, 2022 pour plus d'informations sur les enjeux de la modélisation des sinistres extrêmes.

1.2.3.2 Le manque de données

Le 21 avril 2022, France Assureurs a publié un communiqué de presse (FRANCE ASSUREURS, 2022) pour faire de la lutte contre les menaces cyber une priorité nationale. L'une des propositions vise à

créer “un dispositif qui garantisse le respect de deux principes clés : le libre choix de l'utilisateur de partager ou non ses données ainsi que l'accès transparent et équitable pour tous les acteurs”.

En effet le risque cyber manque de maturité, que ce soit par l'absence d'information, l'imprécision des méthodes de collecte et leur stabilité dans le temps (voir, Hillairet et Lopez, 2022). De plus, le faible partage des données est entre autres dû au caractère sensible qu'impose le traitement de ce risque. Nul ne souhaiterait garder ses comptes dans une banque vulnérable aux attaques cyber. Ainsi les victimes sont très prudentes quand il s'agit de partager ce genre d'informations.

Précédemment nous avons brièvement présenté la base de données Privacy Rights Clearinghouse (PRC) qui rassemble des évènements de violation de données au sein de l'industrie américaine. Comme précisé dans Hillairet et Lopez, 2022, cette base de données publique se doit d'être manipulée avec précaution pour la modélisation puisque certains points de conception restent flous. En effet cette base n'est plus mise à jour depuis 2019.

Une deuxième base de données importante est la *Veris community database* qui collecte depuis 2010 des informations sur des incidents cyber selon une méthodologie particulière. Cette base de données publique cherche à établir un langage commun pour décrire les évènements cyber et de les répertorier. D'autres bases de données existent comme l'HHS, lIndiana mais également des base de données privées comme Ponemon, Advisen et Risk Based Security.

1.3 Assurer le risque cyber

Dans la première section de ce chapitre, nous avons présenté en guise d'introduction, ce qu'est le risque cyber et l'environnement dans lequel il évolue. Dans la deuxième section, nous avons présenté quelques freins à la modélisation de ce risque. Dans cette dernière section nous allons présenter plus généralement ce qu'est l'assurance cyber. Nous commencerons par illustrer quelques exemples d'évènement importants puis nous enchaînerons sur les principales garanties que propose le marché et les différents types de contrats. Finalement, nous conclurons en présentant le cas particulier du cyber silencieux et poser notre problématique.

1.3.1 Le marché de l'assurance cyber

De nos jours les assurances cyber sont essentiellement souscrites par les grandes entreprises. En effet comme le détaille l'Association pour le Management des Risques et des Assurances de l'Entreprise (AMRAE) dans son étude AMRAE, 2022, 82% du volume des primes sur le marché français est porté par les grandes entreprises, ce qui, confronté à la figure (1.1), montre un décalage entre les assurés et les plus exposés à ce risque (TPE/PME/ETI). De plus, après une année 2020 avec un ration Sinistres/ Primes défavorable de 167%, en 2021 ce ratio est repassé à 88%, en partie dû à la grande restriction des garanties proposées (clauses d'exclusions, limites d'indemnisations...). Cependant, selon Munich Re (Munich Re, 2022), le marché global de l'assurance cyber devrait représenter en 2025 approximativement 22,1 Milliards $USD en primes, contre 9,1 Milliards $USD en 2021. Ainsi, le marché de l'assurance a le vent en poupe pour encore au moins trois ans.

1.3.1.1 Exemples d'attaques

Nous allons présenter deux attaques de grande envergure qui ont pu marquer les esprits et montrer quelques conséquences que les cyberattaques peuvent engendrer. Rappelons qu'en 2015, d'après Dr. Jane LeClair (LeClair, 2015), le directeur général des opérations de l'Institut national de cybersécurité à Excelsior College, 60% des PME qui avaient subi une cyberattaque font faillite.

En mai 2017, le monde a connu la plus grande attaque par *ransomware* de l'histoire, WannaCry. Ce dernier a infecté plus de 300 000 ordinateurs dans plus de 150 pays, touchant essentiellement l'Inde, les Etats-Unis et la Russie. L'une des plus grandes victimes a été le système national de santé britannique (NHS) en affectant cinquante hôpitaux, soit 20% du réseau hospitalier du NHS.

WannaCry exploitait la faille EternalBlue qui rendait les utilisateurs Microsoft vulnérables. EternalBlue est un élément de programme développé par l'Agence de Sécurité Nationale américaine (NSA en anglais) qui exploite une faille de sécurité informatique présente dans le protocole SBMv1 (première version du protocole SMB). Le protocole SMB permet aux utilisateurs Windows, de partager des ressources sur des réseaux locaux. Ainsi le virus infectait les ordinateurs en chiffrant tous les fichiers, effectuait la demande de rançon et se propager en se connectant à d'autres ordinateurs. Pourtant, en mars 2017, Microsoft avait corrigé la faille via le bulletin MS17-010. Notons que le montant des rançons demandées (en bitcoin) n'était pas excessif, entre 300 et 600$USD. Cependant, certains estimaient que les pertes économiques pouvaient atteindre les 4 milliards de dollars (Staff Writer, 2017). Les acteurs derrière cette attaque restent inconnus mais les plus gros soupçons se tournent vers les organisations criminelles et les états.

Une deuxième attaque, plus récente, qui s'est produite en mai 2021 a directement visé le plus grand oléoduc d'hydrocarbures des Etats-Unis. En effet, environ 45% du carburant consommé par la côte Est du pays transite par le Colonial Pipeline. L'attaque par rançongiciel a directement visé les réseaux informatiques de l'entreprise. Par précaution et par crainte que les pirates aient réussi à atteindre des parties plus vulnérables du pipeline, la société a décidé de le fermer le temps de restaurer un niveau de sécurité suffisant. Cette mesure a directement impacté les acteurs dépendant de cet approvisionnement, par exemple, les aéroports internationaux de Charlotte et d'Atlanta ont dû soit trouver d'autres fournisseurs, soit modifier les horaires de certains vols. De plus, les états de la Caroline du nord, de la Floride et de l'Alabama ont signalé des pénuries, affectant entre autres le prix de l'essence. Ainsi, deux jours après l'attaque, le président Joe Biden a déclaré l'état d'urgence, permettant de lever les limites de transport routier de carburant et trois jours après l'attaque, c'est le gouverneur de Géorgie qui supprime temporairement la taxe sur l'essence. Concernant l'auteur de l'attaque, le principal suspect reste la Russie (Cyberattaque de Colonial Pipeline, 2022).

1.3.1.2 Exemple de garanties et de contrats cyber

Dans ce contexte les entreprises cherchent à se prémunir contre les attaques cyber. Cela peut passer par de l'investissement en cybersécurité mais la certitude d'une défense parfaite n'est jamais garantie. Ainsi les entreprises peuvent chercher à souscrire une assurance cyber. Les principales garanties que nous retrouvons pour les assurances cyber en Europe sont :

- La garantie réparation des dommages et rétablissement du système d'information : couvre les dommages matériels causés par l'attaque et garantit le rétablissement du système d'information de l'entreprise. Dans un évènement d'accumulation comme l'a pu être Wannacry, cette garantie pourrait être mise à mal si l'assureur n'est pas en mesure d'intervenir sur un grand nombre d'assurés dans un laps de temps relativement court.

- La garantie perte d'exploitation : couvre le manque à gagner de l'entreprise causé par l'interruption d'activité. Par exemple, dans le cas d'un rançongiciel, cette garantie peut être activée si le système d'information paralysé empêche l'activité commerciale de l'entreprise.

- La garantie responsabilité civile : couvre les dommages aux tiers. En effet, en cas de fuites de données, le RGPD peut engendrer des frais juridiques ou des amendes.

- La garanties frais de communication de crise : couvre les frais de communication de crise engendrés par l'attaque. En effet une attaque peut nuire à l'image de l'entreprise, l'obligeant à soigner sa communication ce qui engendre donc des frais exceptionnels.

Ces garanties peuvent se trouver dans des contrats classiques (multirisques professionnels, MRH...) avec une affirmation de la prise en compte d'évènements cyber, ce qui s'appelle du *Cyber Endorsement* mais peuvent également faire partie d'une police spécifique dite de *standalone cyber*. Cette dernière consiste donc à créer une police à part entière pour couvrir le risque cyber de l'intéresse. Le risque cyber évoluant très rapidement, il existe un risque chez les assureurs de mal gérer les évènements que les garanties peuvent couvrir, et ce, pour tout type de police confondu. En effet, un certain nombre de garanties (peu importe la police en question), n'excluent pas le risque cyber sans pour autant l'affirmer. Ainsi, l'engagement de l'assureur est déséquilibré vis à vis de la prime perçue. Concrètement, des garanties classiques qui n'ont pas été conçues pour couvrir des évènements d'origine cyber, peuvent en réalité être activées lors d'une cyber attaque. Ces garanties s'appellent souvent couvertures silencieuses mais sont sous entendues par cyber silencieux ou *silent cyber*.

1.3.2 Le cas particulier du cyber silencieux

Comme nous l'avons tout juste présenté, le cyber silencieux se produit lorsqu'une garantie n'affirme pas ou n'exclut pas la couverture en cas d'évènement cyber. Ce type de clause représente un grand enjeu pour les assureurs puisque la garantie n'est plus correctement tarifée vis-à-vis des risques qu'elle porte.

1.3.2.1 Le cas Merck & Co.

Un cas pratique très important est celui de l'entreprise pharmaceutique Merck & Co. qui en 2017 a subi (parmi plusieurs entreprises) de très lourdes pertes à la suite de l'attaque mondiale par ronçongiciel NotPetya. Avec plus de 40 000 ordinateurs endommagés par l'attaque, Merck demande ainsi une indemnisation de 1,4 milliards de dollars pour sa police d'assurance aux biens. L'assureur décide de ne pas indemniser Merck en invoquant l'exclusion par acte de guerre ou hostilité. En effet, le département de justice américain avait inculpé six ressortissant Russes avec des liens présumés avec les services de renseignements militaires russes.

En novembre 2019, Merck & Co. se retourne contre son assureur Ace American, en plaidant que l'attaque ne provenait pas d'un acte national officiel. En janvier 2022, la cour supérieure du New Jersey tranche en faveur de la multinationale en précisant que l'exclusion par risque de guerre et hostilité invoquée par l'assureur se référait à des conflits armés (Vittorio, 2022). Ainsi Ace American, s'est vu obligée d'indemniser la compagnie pour une garantie relevant du cyber silencieux.

Le cas de Merck & Co. est particulièrement marquant du fait des importantes sommes en jeu. Cependant, le débat sur l'exclusion du risque cyber par acte de guerre est toujours d'actualité. Ainsi,

Lloyd's Market Association (LMA) a publié quatre clauses d'exclusions pour les risques de guerre, cyber guerre et opérations cyber (LMA, 2021). Notons que ces clauses concernent le marché britannique ou américain et sont difficilement transposables au droit français, nous reviendrons sur ce point dans le Chapitre 3 lorsque nous traiterons plus en détail le cyber silencieux.

1.3.2.2 Problématique

Il est important de noter que l'attaque de rançongiciel qu'a subi Merck & Co. est semblable à celle de Wannacry. En effet NotPetya exploite la même faille, c'est à dire EternalBlue, et se propage de la même façon mais, au lieu de chiffrer les données il les efface tout en faisant une demande de rançon. Ce virus aurait tout de même infecté plus de 2 000 organisations à travers le monde, causant de graves dégâts sur son passage (Cloudflare, 2022). L'attaque de Wannacry s'est produite en mai 2017 tandis que NotPetya est apparu fin juin 2017.

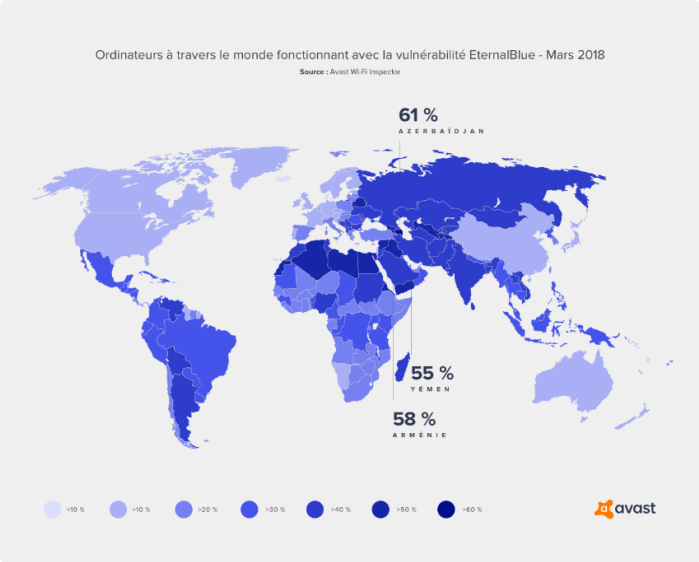


Figure 1.4 : Ordinateurs toujours vulnérables à des virus exploitant la faille Eternal Blue un an après l'évènement Wannacry (Vlcek, 2018).

La figure (1.4) est issue d'une analyse de la compagnie d'antivirus Avast et montre que malgré le *patch* de Microsoft contre la faille exploitée par EternalBlue, de nombreuses compagnies restaient vulnérables.

De plus, l'évaluation de l'exposition des assureurs au cyber silencieux n'est pas simple. En effet, les polices sont amenées à changer régulièrement et la rédaction d'une clause d'exclusion de risques pour une garantie est délicate.

Dans ce contexte, nous allons chercher à modéliser l'impact que pourrait avoir un évènement d'accumulation cyber sur des couvertures cyber silencieuses. Ainsi, nous cherchons à illustrer comment une prise en compte du risque d'accumulation cyber permet une meilleure résilience des portefeuilles non cyber.

La diversité des attaques et de ses auteurs entraînent une grande variabilité sur les impacts que peuvent avoir les cyber-attaques. De plus, aucun profil n'est épargné par la menace, et la faible maturité assurantielle du risque soulève de nouveaux défis de modélisation. L'un d'entre eux étant la modélisation du risque d'accumulation cyber. De plus, chez les assureurs, les couvertures cyber silencieuses (non-affirmatives) font porter à des garanties non-vies des risques cyber. Ainsi, dans ce mémoire nous cherchons à illustrer comment une prise en compte du risque d'accumulation cyber permet une meilleure résilience des portefeuilles non cyber.

Chapitre 2

Modélisation de la composante systémique

Nous avons vu dans le **Chapitre 1** (section 1.2) que le risque cyber est un risque particulier qui se caractérise par plusieurs problématiques à traiter pour aboutir à une modélisation la plus complète possible. Nous cherchons dans ce chapitre à apporter des éléments de modélisation du risque d'accumulation. Nous commencerons par introduire ces modèles. Puis nous ferons un état de l'art des différents mémoires et articles qui traitent le sujet. Finalement nous étudierons un modèle SIR sur plusieurs réseaux, tout en traitant les aspects algorithmiques.

2.1 Les modèles de pandémie avec structure de réseau

Dans le chapitre précédent, les différentes particularités des risques portés par le cyber ont été introduites. Ainsi, ce risque se confronte à des freins de modélisation comme peut l'être l'absence de données représentatives du marché français (voir section 1.2.3.2). Cependant, un des risques importants portés par le cyber est l'effet d'accumulation qui peut résulter de certains évènements particuliers. Nous pouvons citer les deux cyber-pandémies Wannacry (voir section 1.3.1.1) ou encore NotPetya (de 2017). Afin de modéliser ce phénomène d'accumulation, nous allons donc nous inspirer de modèles épidémiologiques qui sont par définition, capables de répliquer des pandémies. Dans cette section, nous allons partir du modèle épidémiologique le plus simple (le modèle déterministe), pour finalement arriver sur des modèles qui incluent une structure de réseaux pour modéliser la population.

2.1.1 Un peu d'histoire

Un poxivirus est un virus faisant partie de la famille des poxviridae. Rien de très évocateur pour les non spécialistes mais il s'agit notamment de la famille qui héberge la variole (ou petite vérole). Cette maladie virale serait apparue chez l'homme dans des villages du néolithique à la suite de contacts avec certains animaux. Comme pour la variole du singe dont on entend parler plus récemment. Depuis, plusieurs formes sont apparues, laissant place à des épidémies et pandémies tout au long de l'histoire. De nos jours, la variole est déclarée éradiquée par l'Organisation Mondiale de la Santé (OMS) grâce à la mise en place de plusieurs campagnes de vaccination, de surveillance et d'isolement.

En avril 1760, l'Académie Royale des Sciences de Paris présente un travail de Daniel Bernoulli

qui traite, entre autres, d'une nouvelle méthode d'analyse de la mortalité causée par la variole (de la Harpe et Gabriel, 2010). Il s'agirait du premier modèle mathématique appliqué à l'épidémiologie.

Avec son modèle, Bernoulli estime le nombre d'individus vivants x années après dans une cohorte où tous les individus sont inoculés dès la naissance. Ainsi, en exploitant une table de Halley (voir Duparquier, 1976 er Ciecka, 2008) pour définir sa cohorte, Bernoulli définit également par individu sensible (nous retiendrons susceptible pour la suite), un individu qui n'a pas été atteint par la variole et donc, non immunisé. De plus, il pose :

- n, le nombre indiquant qu'un individu susceptible a une chance sur n d'attraper la variole. Il fixe $n = 8$.
- m, le nombre indiquant qu'un individu infecté a une chance sur m de mourir de la variole. Il fixe $m = 8$.
- N, le nombre indiquant qu'un individu susceptible a une chance sur N, de mourir du fait de l'inoculation. Il fixe $N = 200$.

Ainsi en utilisant les probabilités précédentes, Bernoulli cherche à filtrer la table de Halley pour déduire le nombre d'individus qui n'ont jamais encore eu la variole (susceptible) notée $s(x)$. De plus, il est en mesure de déduire quelle serait l'évolution de la population s'il n'y avait pas eu de variole. Il pose,

$$\begin{aligned} s(x) &= \frac{m}{(m-1)e^{\frac{x}{n}}+1}\xi(x), \\ z(x) &= \frac{me^{\frac{x}{n}}}{(m-1)e^{\frac{x}{n}}+1}\xi(x). \end{aligned} \tag{2.1}$$

Où :

- $s(x)$ représente le nombre de susceptibles x années après présents dans la cohorte de Halley.
- $z(x)$ représente le nombre d'individus x années après présents dans une cohorte où la variole n'existerait pas du tout.
- $\xi(x)$, le nombre d'individus toujours en vie x années après dans la cohorte de Halley.

Finalement, il propose que la population vivante x années après dans une cohorte inoculée dès la naissance serait égale à

$$\frac{N-1}{N}z(x).$$

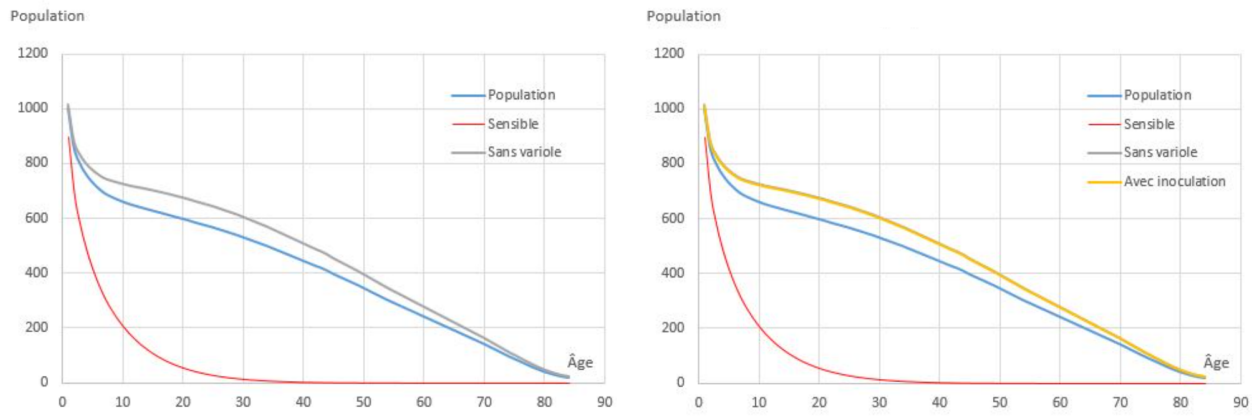


FIGURE 2.1 : Suivi des différentes populations au cours du temps (MONTEILH, 2020a).

Comme nous pouvons le voir sur la figure (2.1), le modèle de Daniel Bernoulli prévoit que l'évolution de la cohorte inoculée (courbe jaune) et celle sans variole (courbe grise) soient bien au-dessus de la cohorte de Halley (courbe bleu) dans laquelle la variole était présente. Ainsi, l'inoculation (traitement de la variole) permettrait malgré les risques qu'elle encoure, de diminuer considérablement les risques de décès par variole.

Pour obtenir les équations 2.1, Bernoulli se base entre autres sur l'existence de $\frac{d\xi}{dx}(x)$. Les références (MONTEILH, 2020a) et (de la HARPE et GABRIEL, 2010) proposent plus de détails sur l'histoire et la construction de ce modèle et dirigent vers des articles plus complets. Les démonstrations peuvent être trouvées dans DIETZ et HEESTERBEEK, 2002.

Quelques années après, d'autres modèles verront le jour comme le modèle de Hamer pour la Peste de 1906 (MONTEILH, 2020b), ou le modèle de Ross pour la Malaria de 1911 (MONTEILH, 2020c).

2.1.2 Du déterministe au stochastique

Malgré les modèles précédents, nombreux sont ceux qui considèrent que le début de la modélisation moderne en épidémiologie remonte au biochimiste Kermack et au médecin militaire McKendrick avec leur article de 1927 (KERMACK et MCKENDRICK, 1927). Commençons par étudier l'un de leurs modèles le plus connu : le modèle déterministe (ou à taux constants).

2.1.2.1 Le modèle déterministe

Comme présenté plus haut, Bernoulli introduit dans son modèle la notion d'individu sensible, que nous avons qualifié de susceptible. Si l'on étend cette idée, lors d'une infection, nous pourrions associer à un individu une unique étiquette qui permettrait de caractériser son état à chaque instant. Par exemple, avant que l'individu ne soit infecté, nous pouvons lui associer l'étiquette de sain, puis une fois infecté, celle d'infecté, puis celle d'hospitalisé, puis finalement celle de décédé ou de rétabli. Malheureusement, l'analogie prise fut celle des compartiments où, à chaque instant, les individus sont classés, selon leur état, dans différents compartiment prédéfinis. Ainsi, nous parlons donc de modèles épidémiologiques compartimentaux, le plus simple étant le SIR, pour Susceptible, Infecté et Rétabli.

Sur la figure (2.2), les compartiments sont représentés par les carrés, et les transitions possibles d'un compartiment à un autre par des flèches. Bien souvent, un abus de notation et fait dans la littérature. Ainsi les grandeurs S, I et R représentent le nombre d'individus dans les états également notés S, I et R. Nous ferons également cet abus lorsque nous traiterons ces modèles.

FIGURE 2.2 : Compartiments pour le modèle SIR.

Une fois les compartiments construits, la question qui se pose est celle de la dynamique. Comment faire transiter les individus d'un compartiment à un autre ? Pour illustrer cette question, imaginons que N individus se trouvent initialement dans le compartiment S de la figure (2.2), ceci serait le reflet d'une population saine, sans virus. Auquel cas, si l'on note S, I, R le nombre d'individus dans leurs compartiments respectifs, nous aurions $S = N$. A $t = 1$, un individu x_0 contracte un virus. Nous avons donc une population de N individus dont le nombre de susceptibles est de $S = N - 1$, et le nombre d'infectés, de $I = 1$. Supposons virus qui engendre un syndrome respiratoire aiguë sévère (SARS en Anglais), a la particularité de bien se propager dans les aérosols engendrés par les symptômes qu'il génère (toux, éternuements...). Ainsi, l'individu x_0 du compartiment I, commence à contaminer ses voisins. Après un jour ($t = 2$), nous reprenons une "photo" de la population et nous remarquons que nous avons $S = N - 5$ et $I = 5$. Ainsi x_0 a contaminé 4 personnes en une journée. Heureusement, 10 jours après l'infection, notre système immunitaire produit des anticorps contre ce virus, l'individu infecté sera ainsi rétabli. Si nous nous plaçons donc à $t = 11$, x_0 est rétabli et nous avons une population de N individus répartis en S susceptibles, I infectés et $R = 1$ rétablis. Nous observons donc comment l'individu x_0 a transité dans les différents compartiments qui l'ont caractérisé au cours du temps. Remarquons également que le nombre d'individus qui, à chaque instant, transite d'un état à un autre dépend des caractéristiques du virus. Ainsi pour décrire un modèle épidémiologique nous ajoutons à la figure (2.2) les taux de passages d'un compartiment à un autre comme illustré sur la figure (2.3). Ces taux correspondent au nombre d'individus qui vont transiter d'un compartiment à un autre.

FIGURE 2.3 : Compartiments et taux pour le modèle SIR.

L'article de 1927 de Kermack et McKendrick (KERMACK et MCKENDRICK, 1927) est souvent associé au modèle SIR déterministe, mais il s'agit en réalité que d'un cas particulier (p. 713) du modèle initialement présenté en début d'article. Cependant le SIR déterministe permet de bien comprendre le fonctionnement des modèles compartimentaux, leurs défauts et leurs potentielles applications. Ainsi

le modèle SIR déterministe pour une population constante de N individus est le suivant :

$$\begin{aligned} \frac{dS(t)}{dt} &= -\beta I(t)\frac{S(t)}{N}, \\ \frac{dI(t)}{dt} &= \beta I(t)\frac{S(t)}{N} - \gamma I(t), \\ \frac{dR(t)}{dt} &= \gamma I(t). \end{aligned} \tag{2.2}$$

Comme présenté dans Kiss et al., 2017, le paramètre β peut être vu comme le nombre de contacts infectieux moyens (contacts qui peuvent contaminer) qu'un individu infecté a au cours d'une unité de temps. Ainsi, le nombre total de contacts infectieux, par unité de temps, au sein de la population est de $\beta \times I$. Or, toute la population n'est pas n'est pas "infectable", en effet un individu déjà infecté ne peut pas se "ré-infecter", mais également, un individu rétabli aura développé des anticorps qui lui permettent d'éviter une nouvelle contamination. La part de la population véritablement susceptible d'être infectée est donc $\frac{S}{N}$, d'où, la variation d'individus susceptibles au cours d'une unité de temps réduit de $\beta \times I \times \frac{S}{N}$, ce qui correspond aux nouvelles infections. Sur la figure suivante nous proposons une simulation pour illustrer l'évolution des différentes populations dans chacun des compartiments.

Le paramètre γ quant à lui, représente la quantité $1/\tau$ où τ est la durée moyenne passée dans le stade infectieux. En effet l'idée est de dire que I infectés passent en moyenne un temps τ avant de décéder ou d'être immunisé, ce qui reviendrait au fait que I/τ infectés passent au compartiment R par unité de temps.

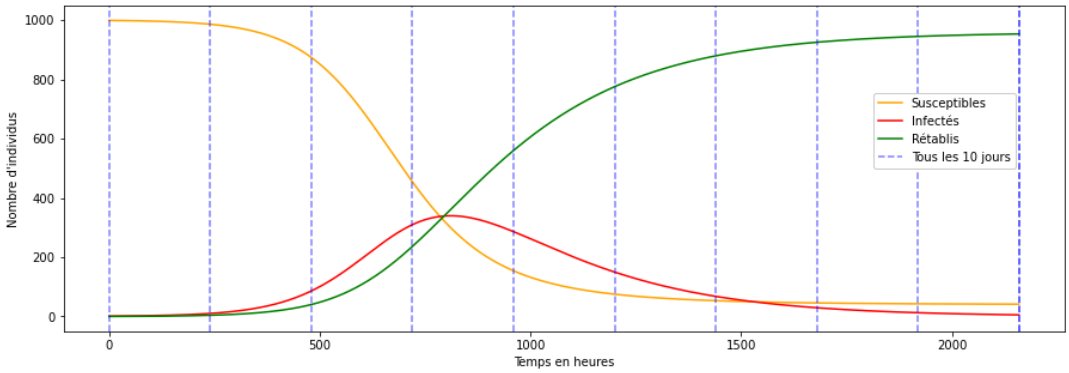


Figure 2.4 : SIR déterministe avec $\beta = 1/72$ et $\gamma = 1/240$ avec un seul infecté en $t = 0$.

Un autre type de modèle à compartiment très utilisé est le modèle SIS. Identique au SIR à la différence qu'un nœud rétabli peut de nouveau contracter l'infection, ce qui revient finalement à retourner dans le compartiment S.

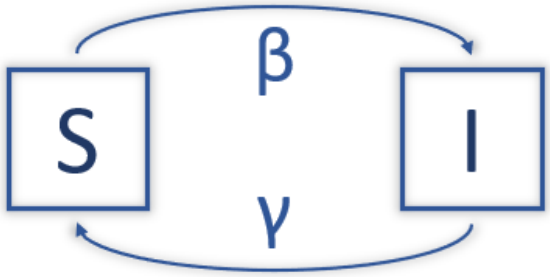


FIGURE 2.5 : Taux de transition entre les compartiments pour le modèle SIS.

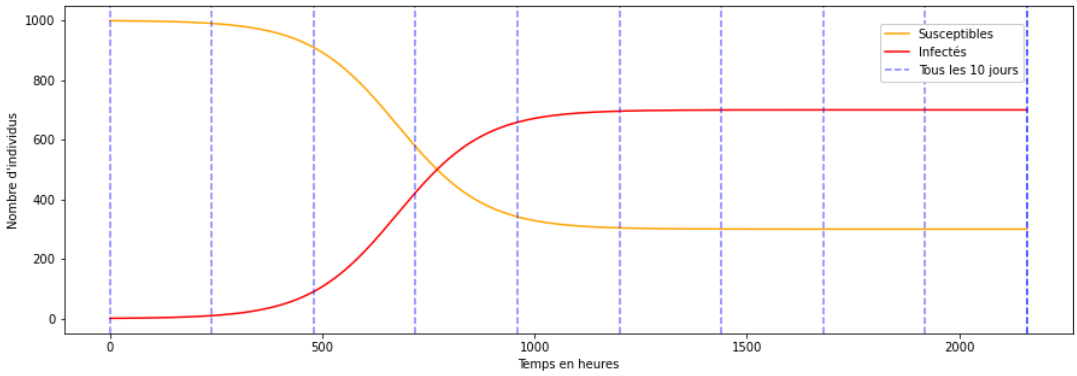


FIGURE 2.6 : SIS déterministe avec $\beta = 1/72$ et $\gamma = 1/240$ avec un infecté au départ.

Nous simulons sur la figure (2.6) le modèle SIS, présenté en figure (2.5), jusqu'au temps $t = 2160$ heures. Nous observons comment chaque compartiment (S et I) atteint un état stationnaire autour des 1200 heures après le début de l'épidémie. Ces états se caractérisent par des très faibles variations d'un intervalle de temps à un autre.

Une dernière analogie de ces modèles est celle des réservoirs d'eaux. En effet, reprenons le modèle SIR, et considérons chaque compartiment comme un bac capable de contenir de l'eau. A l'instant initial, toute l'eau est contenue dans le bac S, puis à l'ouverture de la vanne, de l'eau va s'écouler au fur et à mesure dans le bac I, puis après un certain temps, dans le bac R. Cette analogie permet de révéler une première limite du modèle, en effet si l'on regarde le système d'équations (2.2), nous nous apercevons que les quantités S, I et R sont supposées être continues au cours du temps, tout comme peut l'être un fluide (l'eau) dans un bac. Or, dans la modélisation que nous cherchons à faire, les quantités S, I et R prennent en réalité des valeurs dans $\mathbb{N}^+$ et non dans $\mathbb{R}^+$ car nous traitons des individus. Ainsi, lorsque $I \in]0; 1[$ le virus circule toujours alors qu'il n'y a peut-être plus d'infectés. Au-delà de cet inconvénient, la modélisation déterministe ne tient pas compte du temps passé par un individu dans l'état I. A chaque instant, une proportion constante d'individus va transiter d'un état à un autre selon leurs taux respectifs sans distinction de qui va transiter.

Ces modèles cherchent à montrer la dynamique globale du processus d'infection et de rétablissement au sein d'une très grande population et parfaitement homogène tant dans leurs caractéristiques que dans leurs contacts infectieux. En effet, l'une des principales hypothèses de ce modèle réside dans l'homogénéité des relations au sein de la population. Aucune distinction n'est faite entre les individus qui la composent. Le modèle tel que posé par le système d'équations (2.2) ne permet pas de prendre en compte une structure hétérogène des contacts infectieux entre les individus ou de leur inclure des

caractéristiques propres.

Une façon de palier à ces problèmes est d'introduire une structure de graphe au sein de la population. Ainsi, nous serions en mesure d'établir quels individus peuvent interagir entre eux et avec quelle intensité !

2.1.2.2 Le modèle stochastique

Comme précisé dans Allen, 2017, la modélisation stochastique est importante lorsque le nombre d'individus infectés est petit ou lorsque la variabilité dans les transmissions, rétablissements (dans le cas des SIR et SIS) ont un impact sur les résultats de l'épidémie. Une première façon d'ajouter de l'aléa dans la modélisation des modèles compartimentaux est de probabiliser les changements d'états des individus. Nous notons $S(t), I(t)$ et $R(t)$ le nombre d'individus dans chacun des compartiments à l'instant t et $\Delta S, \Delta I$ et ΔI la variation d'individus dans chacun des compartiments au cours d'un temps Δt. Ainsi nous posons les probabilités pour les évènements de la façon suivante :

$$\begin{aligned} \text{Infection} &: \quad (\Delta S = -1, \Delta I = +1, \Delta R = 0) \quad \text{avec probabilité} \quad \beta I(t)\frac{S(t)}{N}\Delta t + o(\Delta t), \\ \text{Rétablissement} &: \quad (\Delta S = 0, \Delta I = -1, \Delta R = +1) \quad \text{avec probabilité} \quad \gamma I(t)\Delta t + o(\Delta t). \end{aligned}$$

Sans rentrer dans les détails du modèle, l'écriture précédente satisfait la propriété de Markov, ce qui implique que les inter-temps T séparant deux évènements (infection ou rétablissement) surviennent selon une loi exponentielle de paramètre (noté λ) la somme des taux des évènements possibles : $T \sim \lambda e^{-\lambda t}$. Dans le cas du SIR nous avons $\lambda = \beta I(t)\frac{S(t)}{N} + \gamma I(t)$. C'est notamment sur cette propriété que se base l'algorithme de Gillespie pour le simuler (section 2.3.4.1).

Notons que par rapport au modèle déterministe présenté précédemment, le nombre d'individus dans chaque compartiment prend des valeurs dans $\mathbb{N}^+$ dans le modèle stochastique. Cependant il n'inclut toujours pas d'hétérogénéité dans les contacts entre individus.

Dans cette section nous avons présenté comment les modèles ont pu se transformer au cours du temps. De plus, les modèles compartimentaux permettent de faire transiter les individus selon diverses catégories qui les caractérisent au cours du temps. Nous verrons que dans la section 2.2 que les modèles SIS et SIR ont déjà été utilisés pour modéliser le risque d'accumulation cyber. Cependant nous cherchons à créer un modèle qui prenne en compte l'hétérogénéité des relations qui existent entre les individus.

2.1.3 Qu'est ce qu'un graphe ?

Nous cherchons à présent, à représenter les interactions que peuvent avoir certains individus (ou autre) au sein d'une population. Les réseaux internet, les réseaux sociaux et bien d'autres, permettent de décrire la façon dont sont connectés des outils informatiques, des individus ou tout autre objet que nous souhaitons modéliser.

2.1.3.1 Problème des sept ponts de Königsberg

Il y a maintenant quelques temps, dans ce qui est aujourd'hui Kaliningrad (cette petite enclave russe au coeur de l'Europe, au Nord de la Pologne), la ville comptait sept ponts afin de pouvoir

traverser la Pregolia comme illustré sur la figure (2.7). Ainsi, la question était de savoir s'il était possible de construire un parcours à travers la ville qui permette de traverser tous les ponts qu'une seule fois.

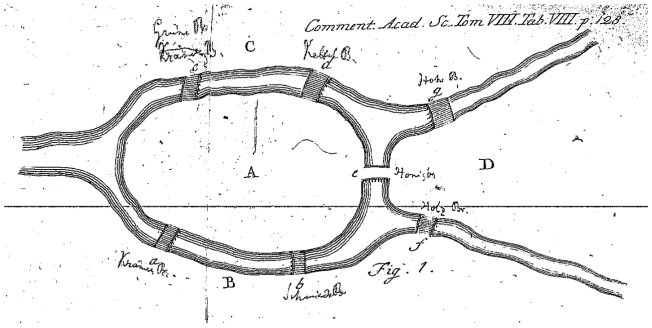


FIGURE 2.7 : Représentation du problème des septs pont par Euler (Source : EULER, 1741). Les ponts sont représentés par les lettres a, b, c, d, e et f, et les amas de terre ferme par les lettres A, B, C et D.

En apportant une représentation du problème à l'aide d'un graphe, Euler a démontre que ce type de parcours n'existait pas pour la ville de Königsberg. Ainsi, un parcours qui traverse toutes les arêtes d'un graphe une seule et unique fois s'appelle aujourd'hui un circuit eulérien. De plus, Euler a démontré que l'existence d'un tel parcours sur un graphe connexe (non-orienté et d'un seul tenant) dépend des degrés (nombre d'arêtes reliées au nœud) des nœuds. Aujourd'hui appelé théorème d'Euler-Hierholzer, il se décline en deux versions :

- Un graphe connexe admet un parcours eulérien si et seulement si ses sommets sont tous de degré pair sauf au plus deux.
- Un graphe connexe admet un circuit eulérien si et seulement si tous ses sommets sont de degré pair.

De cette façon Euler montre l'efficacité de la modélisation par graphe (et des réseaux) à résoudre des problèmes. Les notions sur les graphes apportées dans cet exemple sont présentées dans la suite.

Ainsi, afin d'apporter quelques éléments sur la théorie des graphes et de la construction de réseaux, nous compléterons les notions déjà introduites dans le mémoire de L.Kouakou (KOUAKOU, 2020), en les étendant aux graphes orientés et en introduisant des notions d'analyse spectrale. Pour cela nous nous appuierons sur le livre de Jørgen Bang-Jensen et Gregory Z. Gutin (BANG-JENSEN et GUTIN, 2009) et celui de Fan R. et K. Chung (FAN et CHUNG, 1997).

2.1.3.2 Les graphes orientés

Un **graphe orienté** est défini par la paire (S, A) où S est un ensemble dont les éléments sont appelés des sommets (ou nœuds lorsque l'on traite des réseaux). L'ensemble $A \subset S \times S$ est quant à lui un ensemble de paires de sommets dont les éléments sont appelés des arcs. Par exemple, si chaque nœud est associé à un chiffre, nous pourrions avoir $S = \{1, 2, 3\}$ et $A = \{(1, 2), (2, 1), (2, 3), (3, 2), (3, 1)\}$. De plus, nous appelons **ordre** du graphe G, le nombre d'éléments dans l'ensemble S. Pour éviter toute confusion, ce type de graphe est appelé *directed graph* ou *digraph* en anglais.

Soit un graphe orienté $G = (S, A)$.

- Le **demi-degré extérieur** k_s^+ d'un nœud $s \in S$ correspond au nombre d'arcs ayant ce nœud comme origine.
- Le **demi-degré intérieur** k_s^- d'un nœud $s \in S$ correspond au nombre d'arcs ayant ce nœud comme cible.
- Le **degré** k_s d'un nœud $s \in S$ correspond a la somme de ses deux deux demi degrés i.e. $k_s = k_s^+ + k_s^-$.
- Les arcs ayant pour origine et cible un seul nœud, nous avons donc $\sum_{s\in S} k_s^+ = \sum_{s\in S} k_s^-$.

En effet pour le dernier point, si l'on note $A \subset \{a = (s_1, s_2) | (s_1, s_2) \in S^2, s_1 \neq s_2\}$ alors,

$$k_s^+ = \sum_{a\in A} 1_{s_1=s},$$

d'où nous aurions,

$$\sum_{s\in S} k_s^+ = \sum_{s\in S}\sum_{a\in A} 1_{s_1=s} = \sum_{a\in A}\sum_{s\in S} 1_{s_1=s},$$

or,

$$\sum_{s\in S} 1_{s_1=s} = 1,$$

d'où finalement,

$$\sum_{a\in A}\sum_{s\in S} 1_{s_1=s} = \sum_{a\in A} 1 = Card(A).$$

De la même façon nous avons :

$$k_s^- = \sum_{a\in A} 1_{s_2=s} = \sum_{a\in A} 1 = Card(A).$$

Il est important de remarquer que l'écriture de k_s^+ et k_s^- peut se faire matriciellement. Ces matrices diagonales sont appelées matrice du demi-degré extérieur et matrice du demi-degré intérieur et contiennent le nombre de demi-degrés pour chaque nœud.

2.1.3.3 Les graphes non orientés

Il est dit que G est un **graphe non orienté** si pour chaque arc reliant deux nœuds il existe un arc les reliant en sens inverse, dans ce cas nous parlons d'arête reliant deux nœuds. Formellement l'écriture mathématique d'un graphe orienté diffère de celle d'un graphe non orienté, en effet l'ordre d'écriture des sommets dans les éléments de A n'a plus d'importance pour un graphe non orienté. Cependant, nous considérerons que le graphe $G = (S, A)$ avec $S = \{1, 2, 3\}$ et $A = \{(1, 2), (2, 1), (2, 3), (3, 2)\}$ est non orienté, puisque pour chaque nœud relié par un arc, il existe un arc reliant les nœuds en sens inverse.

Dans le cas des graphes non orientés $G = (S, A)$ nous avons :

- le **degré** du nœud $s \in S$, est noté k_s et correspond donc au nombre d'arêtes reliées au nœud s.
- le **degré moyen** du graphe $\langle K \rangle = \frac{1}{N}\sum_{s\in S} k_s$.

De la même façon que pour les graphes orientés, nous pouvons définir la matrice (diagonale) des degrés du graphe.

Si chaque sommet du graphe non orienté G possède le même degré , nous qualifierons G de **régulier**.

De plus, si l'on note $d_1, d_2, ..., d_L$ les différents degrés présents au sein du graphe et N_l le nombre de nœuds ayant pour degré d_l, alors **la densité du degré** correspond à $p_l = \frac{N_l}{N}$. Ainsi pour un graphe homogène, nous aurons $L = 1$ et $N_1 = N$.

2.1.3.4 Représentations d'un graphe

Nous avons vu que les graphes possèdent plusieurs caractéristiques qui permettent de les comparer entre eux. Comme nous le verrons dans cette partie, leur représentation peut également se faire de plusieurs façons.

La façon la plus simple de représenter un graphe reste la méthode graphique, par exemple, pour le graphe orienté $G_{ori} = (S, A_{ori})$ où $S = \{1, 2, 3, 4\}$ et $A_{ori} = \{(1, 4), (4, 1), (1, 2), (3, 1), (4, 3)\}$ et pour le graphe non orienté $G_{non} = (S, A_{non})$ où $A_{non} = \{(1, 3), (1, 2), (1, 4), (4, 3), \}$ nous proposons un représentation graphique sur la figure (2.8).

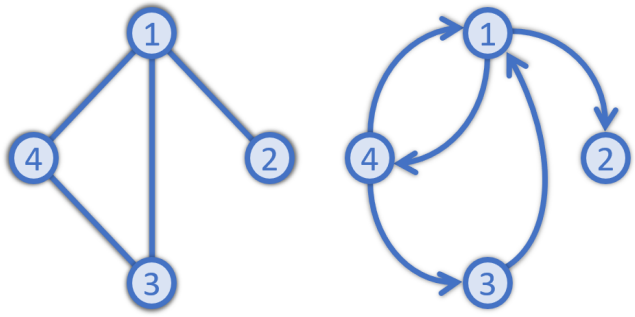


FIGURE 2.8 : Graphe à 4 nœuds, G_{non} non orienté à gauche et G_{ori} orienté à droite.

Les graphes peuvent être pleinement décrits par leur **matrice d'adjacence**, dans la suite du mémoire, nous reviendrons à plusieurs reprises sur cette représentation. Ainsi, pour un graphe $G = (S, A)$ où on note chaque nœud par un numéro, tel que $S = \{s_1 = 1, s_2 = 2, ..., s_N = N\}$, nous associons à G la matrice d'adjacence R définie par :

$$R_{ij} = \mathbb{1}_{(s_i, s_j) \in A} \quad \text{où,} \quad s_i \in S \quad \text{et} \quad s_j \in S. \tag{2.3}$$

Nous pouvons remarquer qu'un graphe non orienté aura donc une matrice symétrique. La matrice d'adjacence pour le graphe G_{non} décrit dans la figure (2.8) serait donc la matrice symétrique $R^n = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$ et pour le graphe G_{ori} nous aurions $R^o = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$. Il important de noter que la matrice d'adjacence est unique à un seul graphe, ainsi, deux graphes différents ne peuvent pas partager la même matrice d'adjacence.

Une deuxième représentation matricielle d'un graphe peut être obtenue par sa **matrice d'incidence**. Pour obtenir cette matrice, nous commençons par numéroter les arcs de la figure (2.8), ainsi nous pouvons écrire $A_{non} = \{a^1_{non} = (1, 2), ..., a^4_{non} = (4, 3)\}$ et $A_{ori} = \{a^1_{ori} = (1, 4), ..., a^5_{ori} = (4, 3)\}$.

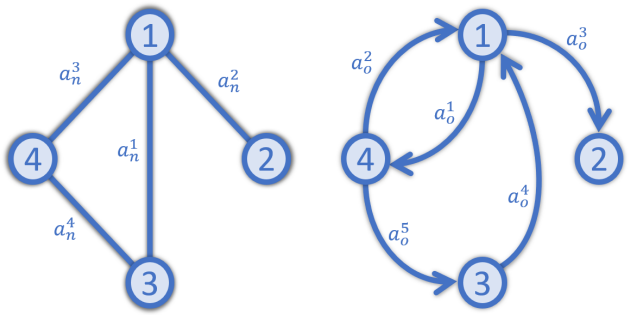


FIGURE 2.9 : Graphe à 4 nœuds avec libellés, G_{non} non orienté à gauche et G_{ori} orienté à droite.

Sur la figure (2.9), nous proposons une représentation en labellisant les arcs et arrêtes. Nous invitons le lecteur à consulter l'annexe (A.2) pour l'écriture de cette matrice et une courte ouverture sur l'analyse spectrale.

2.1.3.5 Les graphes pondérés

Une extension des graphes que nous avons pu présenter jusqu'à présent peut se faire en ajoutant des poids aux arêtes (ou arcs pour les graphes orientés) qui composent notre réseau. L'ajout de ces poids permet de rajouter de l'information contenue par le graphe. Nous le verrons plus tard (section 3.2.1.4), mais ces poids permettront de caractériser le milieu dans lequel le rançon logiciel se diffusera lors de notre modélisation épidémiologique.

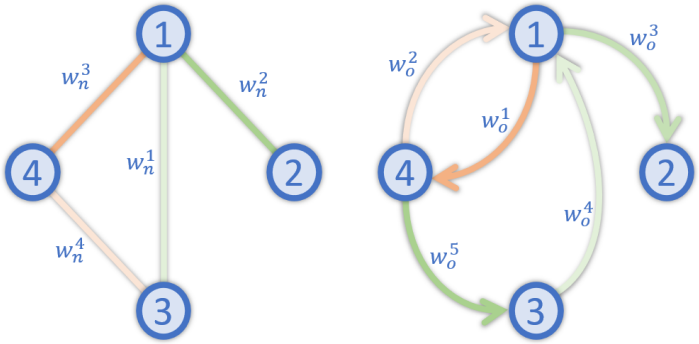


FIGURE 2.10 : Graphes ayant des poids sur chaque arête ou arc qui le compose, la couleur est ici proportionnelle au poids w_n^j et w_o^j.

Nous illustrons sur la figure 2.10 deux graphes, un orienté et l'autre non, contenant chacun des poids sur ses arcs ou nœuds.

2.1.3.6 Comment obtenir un graphe ?

A présent nous allons détailler les différentes méthodes de construction de graphe. Globalement nous retiendrons deux méthodes pour les générer. La première consiste à utiliser des données existantes,

tandis que la deuxième, va les générer aléatoirement selon une certaine “loi” que nous allons détailler dans la suite.

Jusqu'à présent nous n'avons pas vraiment fait de distinction entre les graphes et les réseaux mais il en existe bien une. La théorie des réseaux n'est en réalité que l'application de la théorie des graphes pour étudier des systèmes complexes qui interagissent entre eux. Nous pouvons donc construire un réseau sur une population en utilisant un critère représentatif des flux ou des liens que nous souhaitons représenter.

Le réseau le plus connu quand nous traitons ce sujet c'est le réseau internet. Conçu pour permettre l'interconnexion des systèmes informatiques, ce réseau est aujourd'hui indispensable au fonctionnement de la société. Ce qui a rendu ce réseau si unique, c'est son protocole de routage IP qui fait en sorte qu'un message arrive toujours chez son destinataire. Internet est aujourd'hui le plus grand réseau informatique et de nombreux projet ont tenté de dresser un schéma de ce dernier. Sur la figure (2.11), nous avons une image du projet OPTE qui en exploitant plusieurs méthodes (comme le *traceroute*) permet une cartographie des différents chemins de routage d'internet.

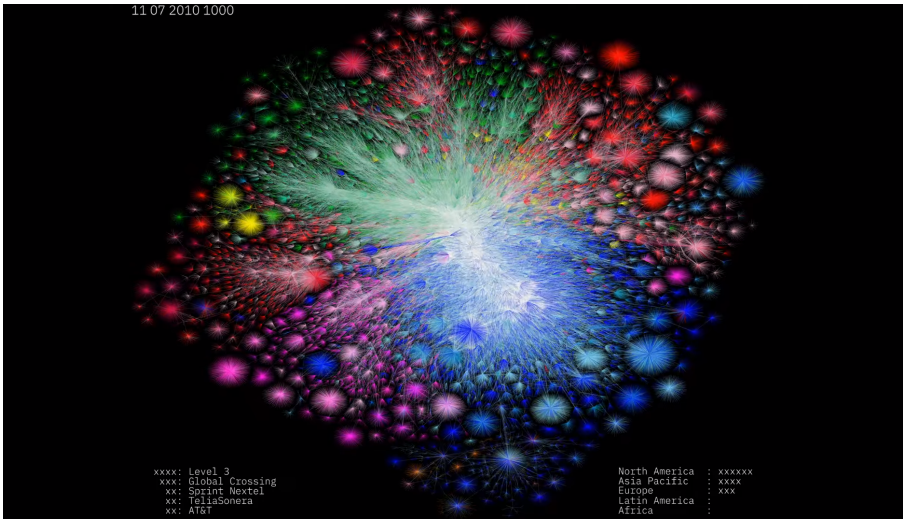


FIGURE 2.11 : Visualisation du projet OPTE représentant le réseau internet (LYON, 2022).

Des travaux plus encadrés sur la représentation du réseaux internet à l'aide de graphes sont menés par le Centre Appliqué à l'Analyse des Données d'Internet (Center for Applied Internet Data Analysis, CAIDA), voir CAIDA, 2021. CAIDA publie également des articles sur des sujets variées : par exemple une analyse de l'impact de la réglementation RGPD sur l'évolution de l'interconnexion des fournisseurs internet (ZHUO et al., 2020).

Finalement, il est possible d'obtenir un graphe en le construisant aléatoirement. L'un des graphes aléatoires le plus connu étant celui introduit par les mathématiciens hongrois Paul Erdõs et Alfréd Rényi en 1959. Communément appelé modèle de Erdõs-Renyi, il est souvent associé à deux méthodes pour obtenir des graphes aléatoires. La première méthode permet d'obtenir un graphe de n nœuds et M arêtes et consiste à choisir uniformément entre les différents graphes qui possèdent n nœuds et M arêtes. Par exemple, pour un graphe à 3 nœuds et 2 arêtes, il existe trois graphes possibles pour lesquels chacun a une probabilité $1/3$ d'être choisi. La deuxième méthode consiste à fixer un nombre n de nœuds et de fixer une probabilité p avec laquelle nous allons rajouter des arêtes entre les nœuds. De cette façon, un graphe de n nœuds et M arêtes a une probabilité $p^M(1-p)^{\binom{n}{2}-M}$ d'être généré.

En ce qui concerne la modélisation numérique des graphes, nous avons vu qu'un graphe peut être représenté à l'aide de sa matrice d'adjacence. Ainsi, la construction de graphes, qu'ils soient orientés ou non, peut se faire à l'aide d'une matrice. Nous pouvons citer entres autres le *package* python NetworkX, développé et maintenu par deux chercheurs du laboratoire national de Los Alamos (Los Alamos National Laboratory) et leur collègue de l'université Colgate (source du *package* HAGBERG et al., 2008). Ce package est riche en contenu, et il permet de générer plusieurs types de graphe mais également de les analyser. En se rapprochant de l'orienté objet, il permet également de poser des attributs aux nœuds et aux arêtes (ou arcs) qui composent le graphe, ce qui, comme nous le verrons dans le chapitre suivant, facilitera la modélisation des assurés par des nœuds.

Dans cette section nous avons montré comment les réseaux peuvent représenter une structure d'information. De plus, la matrice d'adjacence nous permet de caractériser un réseau, qu'il soit pondéré ou non. Nous utiliserons cette représentation par matrice d'adjacence dans le Chapitre 3 (section 3.2.1.4) pour modéliser le réseau entre assurés. Finalement, nous avons proposé plusieurs méthodes pour obtenir un graphe ou réseau.

2.1.4 Épidémiologie et réseaux

Jusqu'à présent, les modèles épidémiologiques présentés considèrent que les contacts entre individus sont homogènes : tous les individus sont en contact entre eux. Ainsi, ceux infectés peuvent diffuser le virus à un individu quelconque du compartiment récepteur (les susceptibles pour les modèles SIR et SIS). Cependant, dans la section précédente nous avons montré comment les réseaux sont en mesure de représenter les relations entre individus ou outils informatiques (ordinateurs de salariés, serveurs, routeurs, etc). Nous présentons dans cette section un modèle unifiant réseaux et modèles de pandémie.

2.1.4.1 Le modèle de Markov Continu

Dans cette section, nous allons présenter le modèle le plus simple utilisé pour diffuser un virus au sein d'un réseau, le modèle de Markov Continu.

L'une des principales différences avec le modèle déterministe c'est l'intérêt que nous allons porter aux individus infectés. En effet, dans les modèles précédents nous ne cherchions pas à savoir quels nœuds en particulier étaient infectés mais nous cherchions juste à avoir le nombre global d'infectés.

Afin d'obtenir une représentation graphique de tous les états possible, nous considérons un réseau $G = (S, A)$ où $S = \{s_1 = 1, s_2 = 2, s_3 = 3\}$ sur lequel les individus peuvent prendre deux états S ou I. Rappelons que ce modèle fait référence à un autre modèle compartimental, le SIS, qui à la différence du SIR, remplace la transition $I \to R$ par un retour à l'état susceptible $I \to S$. Ainsi, comme nous l'avons vu précédemment, dans le modèle SIS, les individus rétablis ne sont pas immunisés contre l'infection et peuvent être infecté de nouveau. Une façon de considérer le réseau est de le caractériser par les états dans lequel se trouvent ses nœuds. Les différents états possibles pour le réseau G sont donc $\{SSS, SSI, SIS, ISS, SII, ISI, IIS, III\}$ et sont représentés dans la figure (2.12).

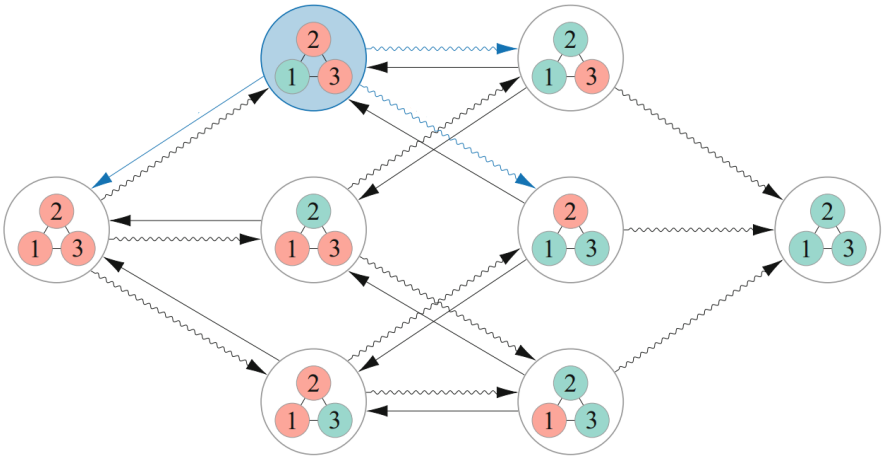


FIGURE 2.12 : Transitions possibles pour un réseau à trois nœuds (KISS et al., 2017).

Prenons l'état SII représenté en fond bleu sur la figure (2.12), les flèches continues représentent une transition d'état du réseau par infection, tandis que les flèches ondulées représentent des rétablissements. Remarquons que l'état SSS est un état absorbant, puisque n'ayant plus de nœuds infectés, aucune nouvelle infection ne peut avoir lieu.

L'idée principale du modèle de Markov est de pouvoir influencer le taux de propagation du virus selon le nombre d'infectés. Nous allons donc considérer que β représente le taux avec lequel un nœud contribue à la transmission du virus à un nœud susceptible. Une façon de l'écrire est la suivante. Pour tout nœud s, notons $E_s(t) \in \{S, I\}$ l'état dans lequel il se trouve à l'instant t, alors :

$$E_{s_i}(t) : S \to I \quad \text{avec un taux} \quad \beta \sum_{s_j \in S} a_{ij} \mathbb{1}_{E_{s_j}(t)=I},$$
$$E_{s_i}(t) : I \to S \quad \text{avec un taux} \quad \gamma.$$

Où a_{ij} est le coefficient de la matrice d'adjacence du réseau considéré.

Nous remarquons donc comment le taux d'infection, transition $S \to I$, dépend du nombre d'infectés voisins au nœud susceptible. Cependant, le taux de rétablissement, transition $I \to S$, ne dépend d'aucun élément extérieur en prenant la valeur γ.

De la même façon nous pouvons écrire les taux pour le modèle SIR où $E_s(t) \in \{S, I, R\}$ tel que,

$$E_{s_i}(t) : S \to I \quad \text{avec un taux} \quad \beta \sum_{s_j \in S} a_{ij} \mathbb{1}_{E_{s_j}(t)=I},$$
$$E_{s_i}(t) : I \to R \quad \text{avec un taux} \quad \gamma.$$

Il est important de noter que lorsque l'on traite les modèles épidémiologiques sur des graphes, le sens du paramètres β n'est pas le même que celui introduit dans le modèle déterministe. Dans le cas du modèle présenté en amont, le paramètre β représente le taux de contagion par arête, c'est donc "l'influence" qu'un seul individu infecté a sur un individu voisin susceptible.

Le modèle étant un processus de Markov en temps continu, les temps de changement d'état sont distribués selon des lois exponentielles.

Nous présenterons dans la dernière section de ce chapitre, deux algorithmes permettant de simuler ce genre de processus, l'algorithme de Gillespie et l'algorithme par génération d'évènements. Afin de visualiser comment se propage un virus sur un réseau, nous utilisons le package python (Miller et Ting, 2020) développé par deux chercheurs de l'institut de modélisation des maladies à Seattle. Sur la figure (2.13) nous proposons un graphe aléatoire sur lequel nous illustrerons quelques trajectoires du modèle SIR et SIS.

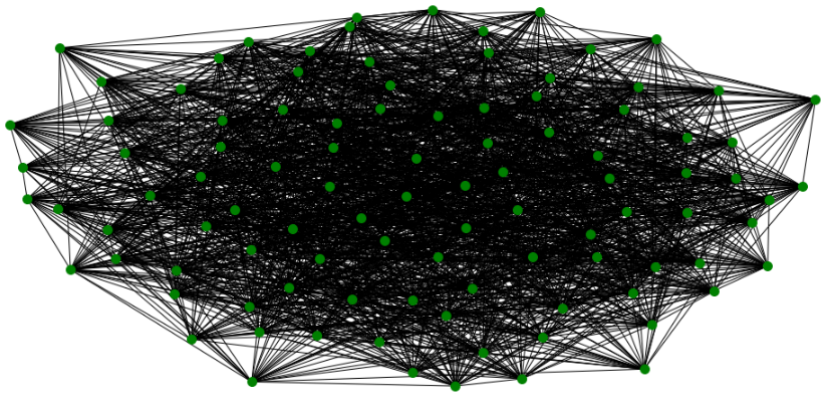
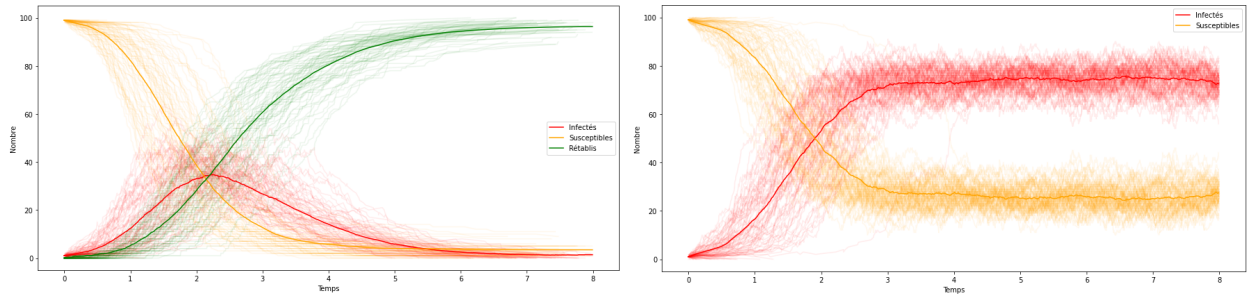


Figure 2.13 : Graphe d'Erdős-Rényi à 100 nœuds et de probabilité 0.4, utilisé pour les exemples de la figure (2.14).



(a) SIR (b) SIS

Figure 2.14 : Modèles SIR et SIS avec $\beta = 0.1$ et $\gamma = 1$.

Le modèle SIR de la figure (2.14) sera plus amplement étudié à la fin de ce chapitre. Cependant, nous pouvons remarquer les fortes ressemblances entre les courbes moyennes (les plus foncées) du modèle SIR par processus de Markov en figure (2.14a) et le SIR du modèle déterministe en figure (2.4), notons tout de même que la taille de la population étudiée n'est pas la même.

2.1.4.2 Première adaptation de la modélisation au cyber silencieux

Dans ce chapitre nous avons construit des modèles épidémiologiques qui modélisent la diffusion d'un virus au sein d'une population ayant une certaine structure. Plusieurs questions se posent pour

l'application de ces modèles au silent cyber, entre autres, celles du lien entre le modèle d'épidémiologie et l'activation d'une clause cyber.

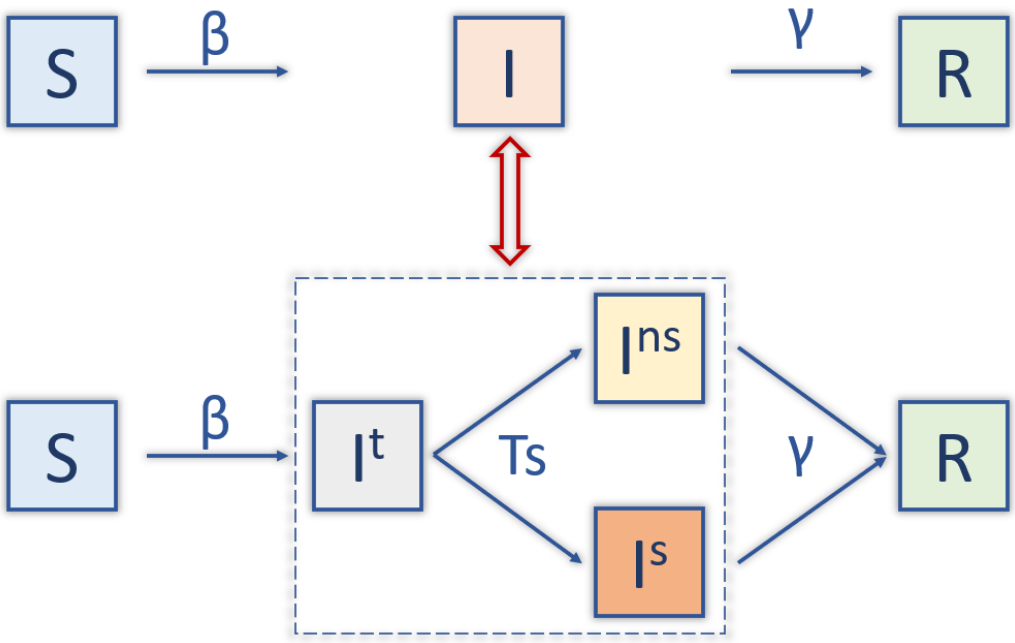


FIGURE 2.15 : Possible transposition du modèle SIR au cas du silent cyber.

Dans la figure (2.15), nous proposons une première transposition du modèle SIR présenté précédemment au cas du silent cyber. Les paramètre β et γ représentent toujours les mêmes que ceux du modèles SIR. Nous souhaitons donc propager une virus informatique sur une population d'assurés exposée au cyber silencieux. Pour réussir à mettre en place cette transposition nous partitionnons le compartiment I en trois sous compartiments : le premier état I^t est un état transitoire permettant de dater l'instant d'infection du nœud, puis un certain taux T_s va venir, indépendamment du processus de diffusion, discriminer quels nœuds correspondent à une infection silent I^s et lesquels non I^{ns}. Ainsi, seuls les individus infectés I^s, sont en mesure de générer des pertes qui soient imputables au cyber silencieux. Finalement, que l'infection soit I^s ou I^{ns}, l'individu se rétablit selon un taux γ.

Il est important de remarquer que les états I^t, I^s et I^{ns}, sont des états "fictifs" qui n'interviennent pas dans le processus de diffusion. Ainsi, identiquement au modèle de Markov SIR, dans lequel le taux de transition $S \to I$ dépend du nombre de voisins infectés au nœud susceptible, le taux de transition $S \to I^t$ en dépend tout autant. Également, les transitions $I^s \to R$ et $I^{ns} \to R$ sont indépendantes de la structure de réseau et ont pour taux de transition γ.

Nous avons présenté dans cette section un modèle de pandémie capable de prendre en compte une structure de réseau. De plus, nous avons proposé un premier modèle capable de distinguer deux catégories d'infectés. Appliqué au cyber silencieux, ce modèle permet de rappeler que la propagation du virus et l'activation du cyber silencieux sont deux processus indépendants. Nous développerons cette modélisation dans la section 3.2 pour que le modèle soit en mesure d'activer les garanties de l'assuré.

2.2 Etat de l'art du risque d'accumulation appliqué au risque cyber

La modélisation mathématique des risques est indispensable pour répondre aux diverses problématiques auxquelles sont confrontés les assureurs. Que ce soit à des fins de tarification, de provisionnement ou autre, la recherche mathématique joue un rôle important dans les modélisation des risques. Dans cette section, nous faisons l'état de l'art des différents modèles, plus ou moins proches les uns des autres, qui permettent de modéliser le risque d'accumulation cyber par des modèles épidémiologiques. Nous commencerons par présenter quelques mémoires puis nous poursuivrons la section par différents articles de recherche.

2.2.1 Mémoires de l'Institut des actuaires

Quittons le temps d'une paragraphe le monde de l'épidémiologie et reprenons le modèle traditionnel de la tarification en assurance, c'est à dire l'approche fréquence-sévérité. Nous renvoyons à la section 1.2.2.1 pour les problèmes que pose le cyber à la modélisation traditionnelle. L'un des premiers travaux innovants sur la modélisation du risque cyber sous cette approche fut celui de Yannick Bessy-Roland (Bessy-Roland, 2019). Dans son mémoire, le caractère d'auto-corrélation et du *clustering* d'évènement cyber présenté dans le chapitre 1 se fait par un processus de Hawkes. Ces processus auto-excitants, historiquement introduits pour modéliser les secousses sismiques sur terre, s'adaptent bien au cyber. Plusieurs applications de ce mémoire ont eu lieu, notamment en tarification ou en provisionnement. Cependant, rappelons que notre objectif est de modéliser le phénomène d'accumulation que ce risque peut produire.

2.2.1.1 Les probabilités d'infection d'un nœud

L'un des mémoires qui traite le risque d'accumulation cyber à l'aide de modèles d'épidémiologie est celui d'Armand Bonnac qui, en 2020, cherche à créer un modèle de tarification pour les collectivités locales (Bonnac, 2020). Dans son mémoire le modèle principal utilisé est un $\epsilon - SIS$ semblable à celui présenté dans la section (2.1.4.1) consacrée au modèle de Markov continu mais un taux ϵ est ajouté au taux d'infection afin de prendre en compte les contagions provenant de l'extérieur du réseau. Le modèle a initialement été introduit par Piet Van Mieghem et Eric Cator dans Van Mieghem et Cator, 2012 où ils le définissent par les taux suivants :

$$\begin{aligned} &E_{s_i}(t) : S \to I \quad \text{avec un taux} \quad \beta \sum_{s_j \in S} a_{ij} \mathbb{1}_{E_{s_j}(t)=I} \quad + \epsilon, \\ &E_{s_i}(t) : I \to S \quad \text{avec un taux} \quad \gamma. \end{aligned}$$

Puis en utilisant l'approximation NIMFA (N-intertwined mean field approximation, voir Van Mieghem et al., 2009), qui suppose que deux nœuds voisins sont statistiquement non corrélés, au modèle SIS, ceci permet d'obtenir la probabilité de contagion d'un nœud au sein d'un graphe. Il réalise ainsi une étude de sensibilité de l'influence de la topologie du réseau sur la probabilité d'infection d'un nœud, puis termine par une application sur un portefeuille de SMACL Assurances.

Le prochain mémoire se concentre sur le risque d'accumulation porté par le cyber.

2.2.1.2 Modélisation du risque d'accumulation

Dans le mémoire (RIGAUD, 2022) réalisé par G. Rigaud chez Axa, son objectif était de construire un modèle pour le risque d'accumulation du cyber affirmatif. En se basant sur des modèles d'épidémiologie stochastiques tel que présentés précédemment, G. Rigaud construit son modèle en se basant sur les trois étapes présentées dans la figure (2.16).

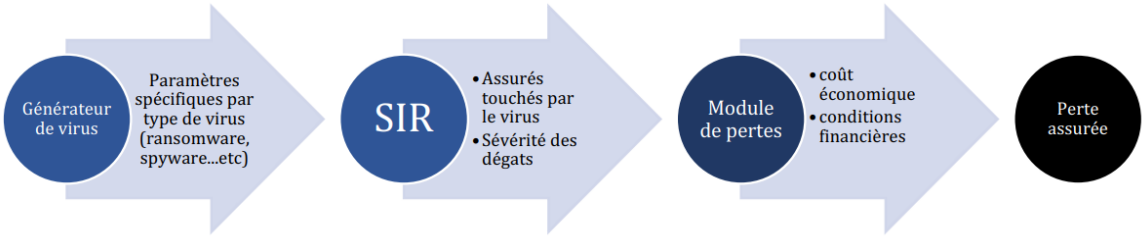


FIGURE 2.16 : Étapes du modèle Malware présenté dans (RIGAUD, 2022).

Il commence donc par définir un virus, puis en utilisant un modèle épidémiologique SIR stochastique il détermine les temps d'infections et de rétablissement de chaque assuré, finalement, il estime les pertes des assurés en associant aux résultats du modèle SIR des coûts économiques et des conditions financières.

Un apport important dans son mémoire est la méthode de calibration utilisée pour estimer certains paramètres de son modèle SIR. En effet, il consacre une bonne partie de son mémoire à recenser et à expliquer les données exploitables sur les évènements WannaCry et NotPetya pour calibrer son modèle selon l'approche bayésienne (méthode ABC, Approximate Bayesian Computation).

L'objectif de cette calibration est de déterminer la loi *a posteriori* des paramètres introduits dans le modèle permettant de reproduire des évènements proches à celui de NotPetya.

2.2.2 L'initiative de recherche Cyber risk : *actuarial modeling*

Depuis 2018, l'institut Louis Bachelier pilote un projet de recherche financé par le fond Axa pour la recherche autour de la modélisation du risque INSTITUT LOUIS BACHELIER, 2022. Ce projet centré sur un partenariat entre Sorbonnes Universités et l'ENSAE a permis, à travers de nombreux articles, d'apporter des résultats sur la modélisation de ce risque.

2.2.2.1 Première modélisation du risque d'accumulation

En 2021, Caroline Hillairet et Olivier Lopez publient un article traitant la propagation d'évènements cyber au sein d'un portefeuilles d'assurance (HILLAIRET et LOPEZ, 2021). La modélisation consiste à utiliser un modèle épidémiologique SIR pour simuler les états des assurés lors de la diffusion du virus. Des pertes peuvent ensuite être calculées selon les temps d'infection et de rétablissement.

Les auteurs introduisent trois fonctions de coûts qui correspondent aux différentes situations que l'on cherche à modéliser :

- $c_1 = \mathcal{C} \sup_{t>0} \mathcal{M}_t$,

- $c_2 = \mathbb{1}_{\sup_t \mathcal{T}_t \leq K}$,
- $c_3 = \int_0^{t_d} \phi\left(\frac{\mathcal{T}_t}{n}\right) d_t$.

Où $\mathcal{C}$ et K sont des constantes positives, t_d est la durée de l'attaque et ϕ une fonction positive. Ces trois fonctions se basent sur les trois quantités suivantes :

- $\mathcal{M}_t$ est le nombre cumulé d'infectés au temps t,
- $\mathcal{R}_t$ est le nombre d'infectés rétablis à l'instant t,
- $\mathcal{T}_t$ est le nombre d'infectés toujours en crise à l'instant t.

Ainsi nous pouvons utiliser :

- la fonction de coût c_1 si l'on cherche à modéliser un coût fixe par assuré infecté,
- la fonction de coût c_2 si l'on cherche à décrire les limites d'intervention de l'assureur face à l'évènement en question,
- la fonction de coût c_3 si l'on cherche à décrire la saturation des capacités de l'assureur face à l'évènement en question mais dont les coûts d'interventions seraient variables en fonction du nombres d'assurés infectés.

Au delà de simplement introduire ces quantités, les auteurs déduisent le comportement limite de $\mathcal{M}_t$ mais également des approximations Gaussiennes pour $\mathcal{M}_t$, $\mathcal{R}_t$ et $\mathcal{T}_t$.

Finalement, une application est faite sur un évènement de type WannaCry. Notons que la modélisation faite dans l'article considère que le portefeuille d'assurés représente un échantillon de la population globale dans laquelle se propage le virus. L'une des principales limites de ce modèle est celle présentée plus haut sur l'homogénéité des contacts entre les individus. C'est cette dernière limite qui a motivé la publication d'un deuxième article où le modèle considère une structure hétérogène au sein de la population.

2.2.2.2 Adaptation à l'hétérogénéité des contacts

Dans la continuité de l'article précédent, les mêmes auteurs s'intéressent à un nouveau modèle qui permet de prendre en compte une structure d'interconnexion au sein de la population (Hillairet et al., 2021). Autant dans le papier précédent, la partie diffusion du modèle n'était qu'un outil pour modéliser le risque d'accumulation, autant dans ce nouvel article, le modèle de diffusion est l'objet central de l'étude.

En s'inspirant du multi-group SIR présenté dans Magal et al., 2018, le modèle de pandémie utilisé prend en compte les caractéristiques des assurés qui pourraient influencer la transmission du virus. Sur la figure (2.17), nous observons les différents compartiments que peuvent prendre les individus. Scindés en deux groupes, les individus infectés du groupe 1 (respectivement 2) influencent la contagion dans son propre groupe mais également celle du groupe voisin avec des taux différents.

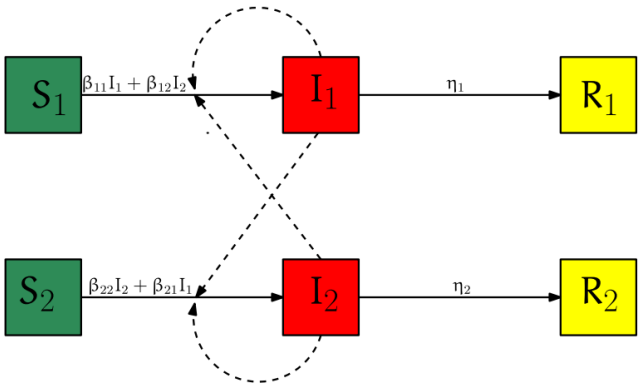


FIGURE 2.17 : Diagramme de transfert du flux individuel entre compartiments : les flèches pointillées représentent l'influence des nœuds infectés, qu'ils proviennent du même secteur ou du secteur voisin (MAGAL et al., 2018).

De la même façon, le nouveau modèle introduit dans HILLAIRET et al., 2021 scinde les assurés en différents groupes qui impactent les taux de contagions. Mais les auteurs vont plus loin en ajoutant aux taux de transmissions, certaines caractéristiques propres au risque cyber. Ainsi, un paramètre α va venir modéliser des attaques qui ne proviennent pas d'un effet contagieux mais d'un effet latent (peut être vu comme un risque minimal, toujours présent quel que soit la situation du voisin) ou externe à la population. Et un deuxième paramètre η va modéliser la protection du groupe, ce paramètre modélise le fait qu'une fois une menace détectée, les individus susceptibles seront plus méfiants et auront tendance à réduire leur risque de contamination.

Ainsi, si l'on note $s_j(t)$ (resp. $i_j(t)$, resp. $r_j(t)$) le nombre d'individus susceptibles (resp. infectés, resp. rétablis) et que j représentante le groupe auquel les individus appartiennent tel que $i = 1, ..., d$ nous pouvons décrire l'évolution de chaque compartiment par le système d'équations :

$$\begin{aligned}
\frac{ds_j(t)}{dt} &= -\eta_j(t)\left\{\alpha_j(t) + \sum_{k=1}^{d}\beta_{k,j}i_k\right\}s_j(t),\\
\frac{di_j(t)}{dt} &= \eta_j(t)\left\{\alpha_j(t) + \sum_{k=1}^{d}\beta_{k,j}i_k\right\}s_j(t) - \gamma_j i_j(t),\\
\frac{dr_j(t)}{dt} &= \gamma_j i_j(t).
\end{aligned}$$

Avec $\beta_{k,j}$ qui est le taux avec lequel la classe k contamine la classe j et γ_j la taux de rétablissement pour le groupe j.

La question qui se pose avec ce genre de modèle est celles des groupes et des valeurs des paramètres à utiliser. Dans l'application proposée par les auteurs, les individus sont classés selon cinq secteurs d'activités, et donc, les valeurs prises par les différents $\beta_{k,j}$ représentent la connectivité entre les différents secteurs. Les différents coefficients $\beta_{k,j}$ peuvent être représentés à l'aide d'une matrice $\mathbf{B}$, de cette façon, les auteurs proposent une matrice basée sur des données l'OCDE pour représenter une partie de ces coefficients. Les données utilisées pour sa construction sont les échanges de valeur ajoutée entre secteurs, ainsi une grande hypothèse est réalisée : celle que le volumes des flux économiques entre

secteurs soient proportionnels aux flux numériques de ces mêmes secteurs. Cette matrice est ensuite normalisée de sorte que la somme des coefficients soit égale à un.

	Mining	Manufacturing	Energy	Construction	Services	Total
Mining	0,0634	0,2927	0,0449	0,1427	0,1255	0,6692
Manufacturing	0,0063	0,0527	0,0027	0,0108	0,0351	0,1076
Energy	0,0135	0,0370	0,0571	0,0150	0,0452	0,1679
Construction	0,0019	0,0068	0,0007	0,0141	0,0091	0,0326
Services	0,0003	0,0042	0,0004	0,0017	0,0161	0,0227
Total	0,0855	0,3934	0,1057	0,1844	0,2309	1

FIGURE 2.18 : Matrice des interactions entre secteurs normalisée notée $\mathbf{B_0}$ (HILLAIRET et al., 2021).

Finalement, ils posent $\mathbf{B} = \beta\mathbf{B_0}$ afin d'inclure le facteur β propre aux caractéristiques du virus. Au-delà de l'apport fait à la modélisation entre groupes d'assurés, un théorème permet par une méthode de point fixe, d'obtenir le nombre d'individus infectés dans la population globale.

Notons tout de même que le SIR présenté dans l'article est stochastique et que la granularité du modèle s'arrête au groupe d'individus, ici représentés par les secteurs d'activités. Ce modèle suppose donc implicitement une structure de réseaux homogène entre les individus.

2.2.3 D'autres articles

Dans la suite nous allons présenter deux articles qui basent leur modélisation sur le modèles SIS par processus de Markov continu comme présenté plus en amont de chapitre. Une façon de décrire le modèle SIS est de considérer que l'état de chaque nœud au sein du graphe est représenté par la variable $X_i(t)$ qui prend comme valeur 1 si le nœud est infecté et 0 si le nœud est ne l'est pas. Ainsi nous avons pour N nœuds d'un graphe décrit par la matrice d'adjacence A dont les coefficients sont a_{ij} :

$$\begin{aligned} X_i(t) : 0 \to 1 \quad &\text{avec un taux} \quad \beta \sum_{j=1}^{N} a_{ij} X_i(t), \\ X_i(t) : 1 \to 0 \quad &\text{avec un taux} \quad \gamma. \end{aligned}$$

2.2.3.1 Une dépendance dans la structure de réseau

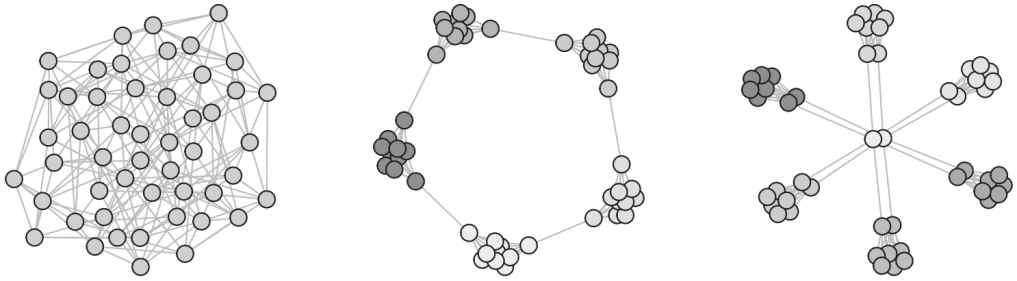
Dans FAHRENWALDT et al., 2018, les auteurs construisent une approche pour tarifer des contrats d'assurance cyber. Leur modélisation considère un processus SIS, comme illustré plus haut, pour diffuser le virus au sein d'un réseau.

Le processus de diffusion va permettre d'infecter les nœuds qui représentent des assurés. Sur ces nœuds infectés, un deuxième processus va permettre de modéliser les pertes engendrées.

Les auteurs consacrent une grande partie de leur article à étudier l'impact qu'ont les différentes approximations des probabilités d'infection pour chaque nœud. Ainsi certaines approximations comme

l'approximation des champs moyens d'Hilbert au premier ordre sous évaluent la probabilité agrégée d'infection, tandis que d'autres comme l'approximation des champs moyens indépendants au premier ordre la sur-estiment.

Finalement, une étude plus pratique est réalisée sur les trois graphes présentés dans la figure (2.19).



(A) **Homogeneous** (B) **Clustered** (C) **Star-shaped**

FIGURE 2.19 : Topologies des trois réseaux utilisés dans FAHRENWALDT et al., 2018

Parmi leurs conclusions, il est retenu que la structure des graphes modifie la façon dont se propagent les virus. Ainsi, la diffusion du virus ne dépend plus que des paramètres β et γ mais également de la structure du réseau considérée.

2.2.3.2 D'autres lois pour les temps d'infections et des pertes en fonction du temps

Dans XU et HUA, 2019, M. Xu et L. Hua construisent un cadre pour modéliser et tarifer le risque cyber. Le modèle considéré et le $\epsilon - SIS$ qui se différencie du SIS présenté précédemment par l'inclusion d'un facteur en plus dans le taux de contagion :

$$\begin{aligned} X_i(t) : 0 \to 1 \quad &\text{avec un taux} \quad \beta \sum_{j=1}^{N} a_{ij} X_i(t) + \epsilon_i, \\ X_i(t) : 1 \to 0 \quad &\text{avec un taux} \quad \gamma. \end{aligned}$$

Ce facteur ϵ_i joue le même rôle que l'ϵ posé dans le modèle posé par A. Bonnac dans son mémoire (BONNAC, 2020) mais dépend ici de l'individu i. Ce facteur permet ainsi de prendre en compte les menaces provenant à l'extérieur du réseau.

De manière similaire à celle introduite dans HILLAIRET et LOPEZ, 2021, les auteurs définissent deux quantités qui leurs permettent de modéliser des pertes liées à un évènement cyber. La première correspond aux pertes causées par l'infection, ce qui peut par exemple provenir d'un vol de données ou de frais juridiques. Ce type de perte est représenté par la variable $L_\cdot$ sur la figure (2.20). La deuxième correspond aux pertes liées à la remise en état de service de l'assuré, c'est par exemple le cas de pertes d'exploitation ou des coûts engagés par l'assureur pour le rétablissement de l'assuré. Ce type de perte est noté $R_\cdot$ sur la figure (2.20).

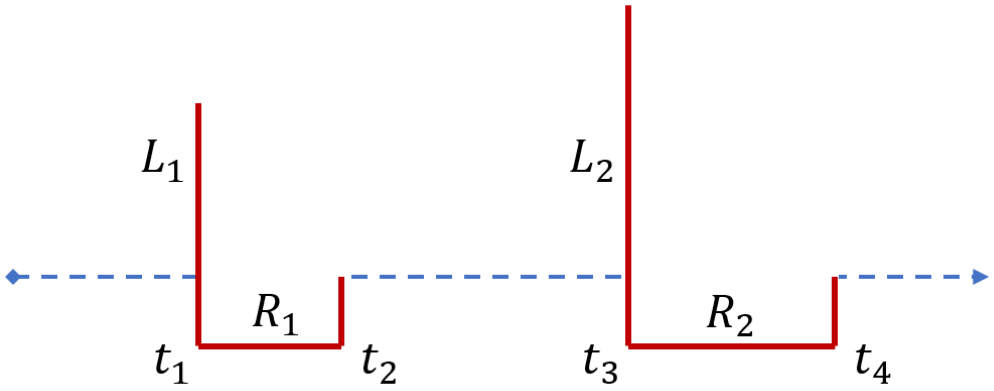


FIGURE 2.20 : Modélisation des pertes pour un évènement lié au cyber.

De la même façon que dans FAHRENWALDT et al., 2018, l'article s'intéresse aux probabilités d'infection des nœuds. Dans un premier temps une méthode pour obtenir les probabilités d'infection d'un nœud est détaillée pour le modèle de Markov continu. Puis dans un second temps, une deuxième méthode permet d'obtenir ces mêmes résultats mais en considérant d'autres lois que l'exponentielle imposée par le modèle de Markov continue.

Ainsi les auteurs calculent les pertes probables générées sur le réseau et suivant certains principes de tarification déduisent une prime pour le contrat.

2.2.3.3 Les arêtes critiques lors d'une attaque

Dans COHIGNAC et KAZI-TANI, 2020, les auteurs considèrent un modèle SIS sur un réseau non orienté et s'intéressent aux arêtes les plus critiques afin de maintenir une bonne connectivité pendant une attaque. Par connectivité, nous sous-entendons la connectivité algébrique, qui correspond à la deuxième plus grande valeur propre de la matrice Laplacienne du graphe (voir annexe A.2). Ainsi le problème se transforme en un problème d'optimisation dont les auteurs apportent la solution.

Notons tout de même que la problématique se concentre sur les arêtes et non sur les nœuds. Ainsi, nous pourrions adapter ces résultats aux du nœuds du graphe pour déterminer, selon un critère de coût, quels nœuds sont contribuent le plus à la transmission du virus au sein du graphe.

2.3 Étude pratique du SIR

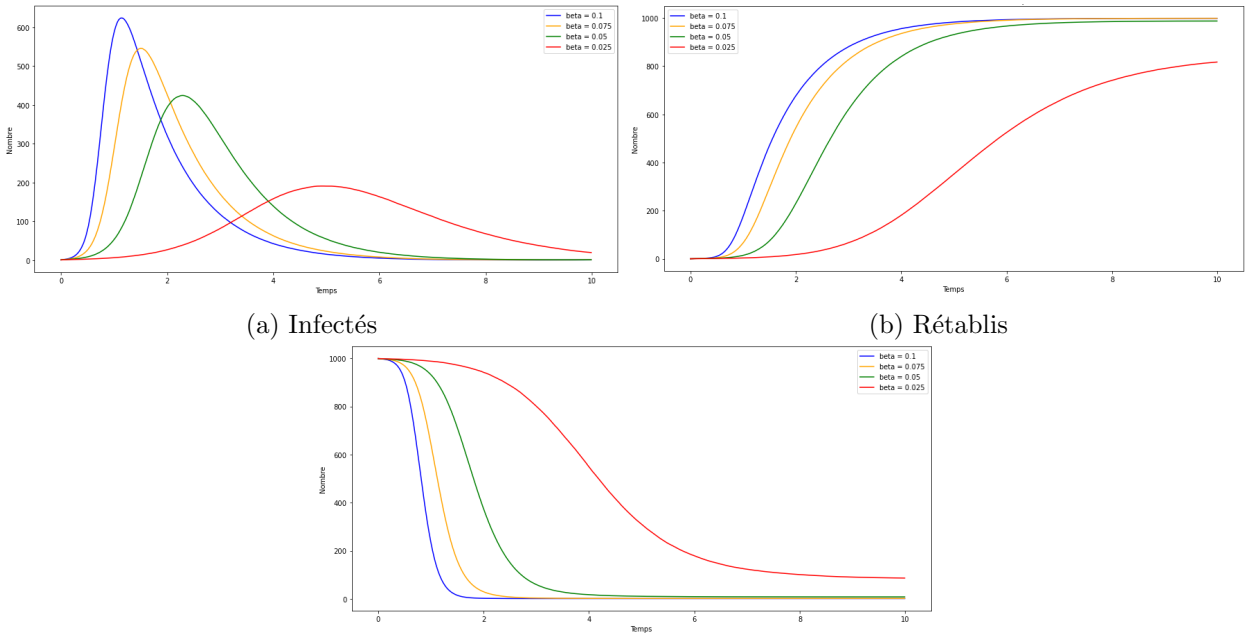
Dans cette section nous proposons une étude pratique du modèle SIR. Nous commencerons par montrer comment évolue la diffusion du virus en fonction des paramètres de contagions β et de rétablissement γ, nous constaterons ensuite comment la topologie du réseau peut également influencer la diffusion, puis finalement, nous montrerons la dépendance qui existe entre la condition initiale et la diffusion du virus.

2.3.1 Étude de sensibilité des paramètres

Nous avons présenté en début de chapitre les modèles SIS et SIR pour décrire la diffusion d'un virus au sein d'une population. Dans cette section nous allons réaliser une brève étude de sensibilité du modèle SIR. Nous commencerons par observer comment évoluent le nombre d'individus des différents compartiments en fonction des paramètres de contagion.

2.3.1.1 Influence du taux de contagion du virus

Commençons donc par observer l'impact du taux de contagion du virus sur 1000 individus reliés selon un graphe de degré 100, c'est à dire que chaque individu est relié à 100 autres nœuds. Pour cela nous allons fixer le taux de rétablissement à $\gamma = 1$ et allons faire varier le paramètre de contagion $\beta \in \{0.025, 0.05, 0.75, 0.1\}$ pour le modèle SIR.



(a) Infectés

(b) Rétablis

(c) Susceptibles

FIGURE 2.21 : Évolution du nombre moyen d'individus par compartiments SIR pour plusieurs taux de contagions.

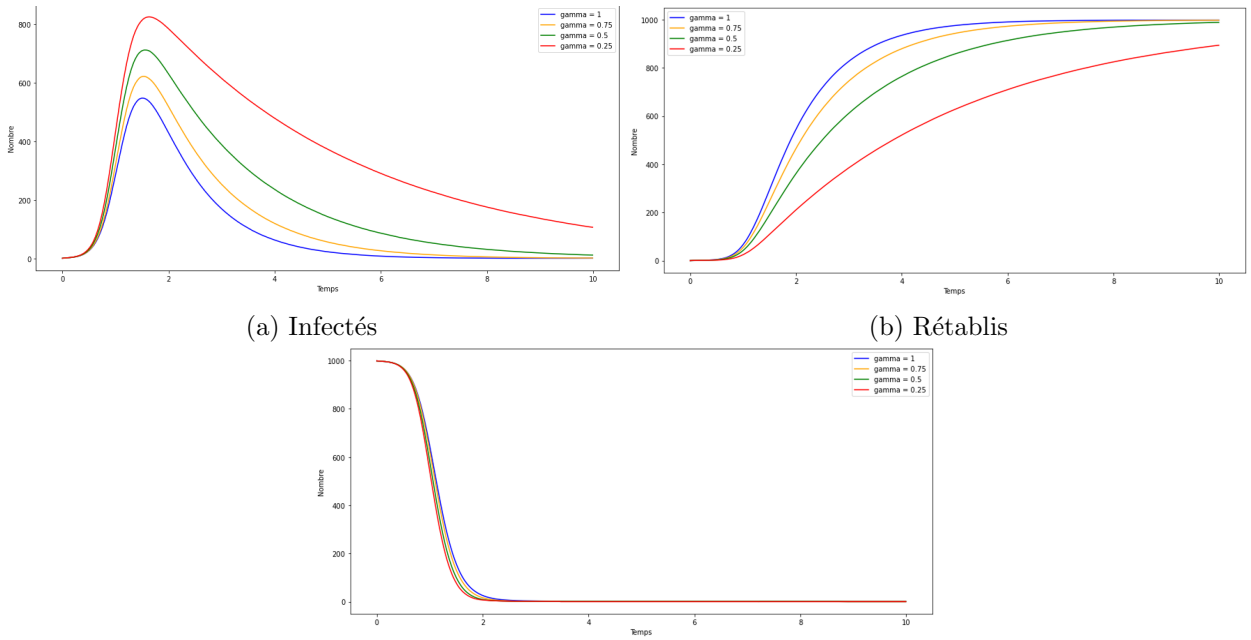
Nous pouvons voir sur la figure (2.21a) que le pic du nombre d'infectés décroît avec la décroissance du taux d'infection β. Cependant nous pouvons remarquer que la diminution de la valeur du pic n'est pas proportionnelle à la diminution du taux d'infection. En effet, l'écart du taux d'infection entre les différentes courbes est constant à 0.025, mais alors que l'écart entre les pics d'infections de la courbe bleu et la courbe jaune se situe autour des 100 individus, celui entre la courbe verte et rouge est plutôt autour des 200 individus.

Nous remarquons également que le temps auquel le pic se produit varie avec la même dynamique que la valeur qu'il prend par rapport au temps. En effet, l'écart entre le temps de réalisation du pic bleu et jaune se situe autour de 0.5 alors que celui entre le pic de la courbe verte et rouge est plutôt autour de 2.5.

L'abaissement du taux d'infection en conservant un taux de rétablissement constant, à un double effet dans la diffusion du virus au sein du réseaux : il influence la valeur du pic du nombre d'infectés mais également le temps auquel il survient. Ainsi un virus avec un taux d'infection plus important va se propager plus rapidement et infecter plus d'individus par unité de temps. Il est tout de même important de noter que les courbes représentent des photos de l'état des individus à chaque instant, elles ne représentent à aucun moment leur nombre cumulé.

2.3.1.2 Influence du taux de rétablissement du virus

A présent nous allons observer l'impact du taux de rétablissement sur la diffusion du virus au sein du même graphe que précédemment. Pour cela nous allons fixer le taux d'infection à $\beta = 0.075$ et allons faire varier le paramètre de rétablissement $\beta \in \{1, 0.75, 0.5, 0.25\}$ pour le modèle SIR.



(a) Infectés

(b) Rétablis

(c) Susceptibles

FIGURE 2.22 : Évolution du nombre moyen d'individus par compartiments SIR pour plusieurs taux de rétablissements.

Nous remarquons ainsi sur la figure (2.22a) que le taux de rétablissement affecte la valeur du pic d'infection mais n'a pas d'impact (à la différence du taux de contagion) sur l'instant auquel ce pic survient. De plus, la décroissance du nombre d'infectés, une fois le pic d'infection passé, va être plus forte avec un taux de rétablissement plus élevé. En effet, plus le taux de rétablissement est faible, plus les individus vont demeurer dans l'état infectieux. Ainsi, si nous souhaitons générer un évènement de contagion très bref, nous pourrons augmenter le taux de rétablissement afin d'obtenir cette décroissance rapide, une fois le pic d'infection passé.

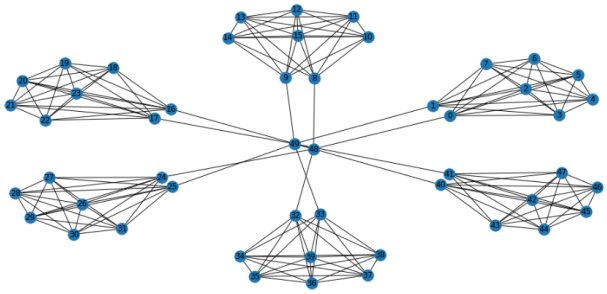
Dans l'étude des paramètres, nous avons choisi un graphe de 1000 nœuds chacun de degré 100 pour propager le virus et nous concentrer sur l'influence des paramètres dans la diffusion du virus. Nous cherchons à étudier l'impact que peut avoir le réseau sur la diffusion du virus dans la population.

2.3.2 Influence du graphe

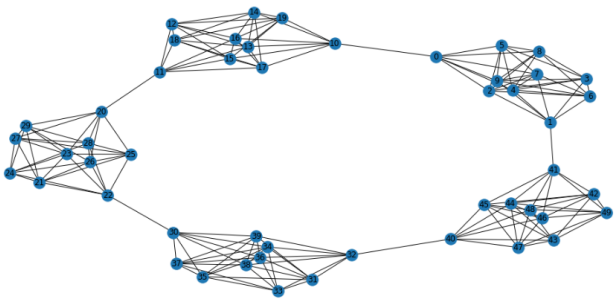
Dans cette section, nous cherchons à montrer l'impact que peut avoir la structure de réseau sur la diffusion du virus. En gardant en tête que le taux avec lequel un nœud susceptible s'infecte dépend du nombre d'infectés qui se trouvent reliés à lui, nous pouvons anticiper le fait qu'un réseau de degré moyen élevé aura tendance à favoriser la diffusion du virus plutôt qu'un réseau de degré moyen plus faible. Nous allons donc fixer les taux de rétablissement et de contagion pour visualiser comment évolue la propagation du virus sur des topologies particulières.

2.3.2.1 Réseaux en étoiles, clusters et de degré 5

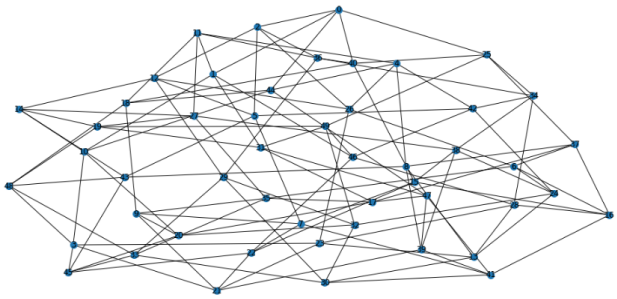
Dans Fahrenwaldt et al., 2018, les auteurs ont exploré plusieurs réseaux pour observer comment la topologie de ces derniers influence les probabilités d'infection des nœuds dans le cas d'un modèle SIS. Nous reprenons ici deux des graphes (clusters et étoile) utilisés dans cet article pour observer comment la structure des réseaux peut influencer la diffusion du virus dans le cadre du modèle SIR.



(a) Étoile



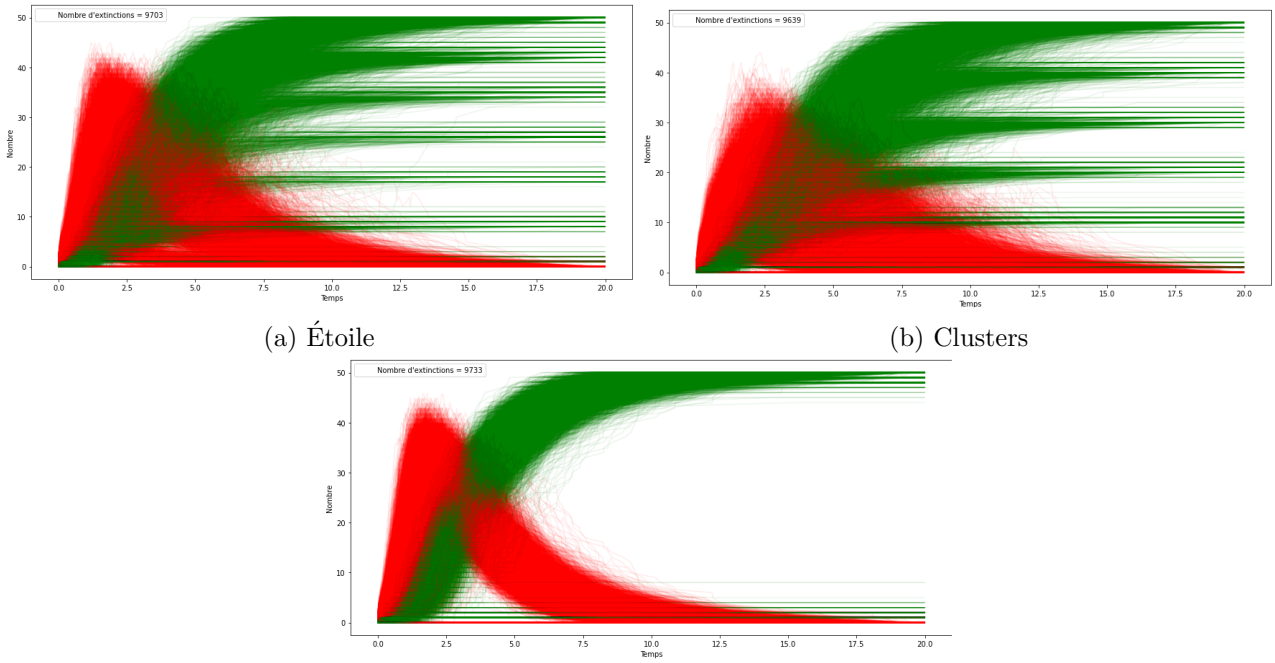
(b) Clusters



(c) Degré 5

Figure 2.23 : Différentes topologies de réseaux.

Nous propageons ainsi sur les graphes présents dans la figure (2.23) un modèle SIR jusqu'au temps $t = 20$, de taux de contagion $\beta = 0.9$ et taux de rétablissement $\gamma = 0.4$. Tous les graphes possèdent 50 nœuds et nous effectuons 10000 simulations pour chacune des topologies.



(a) Étoile

(b) Clusters

(c) Degré 5

FIGURE 2.24 : Simulations (10 000) pour le nombre d'infectés (en rouge) et de rétablis (en vert) dans le modèle SIR pour les réseaux de la figure(2.23). Plus les couleurs ressortent, plus le nombre de simulations à suivre cette trajectoire sont importantes.

Sur les différentes figures ci-dessus nous indiquons le nombre de trajectoires éteintes (plus aucun infecté) avant d'atteindre le temps $t_{final} = 20$. Nous remarquons que ce nombre est très proche d'une structure de réseau à l'autre, autour des 9700 trajectoires. Chaque structure influence la façon dont le virus se propage dans la population, ainsi nous pouvons observer comment le réseau de degré 5, voir figure (2.24c), diffuse le virus de façon plus homogène que les deux autres. Ceci se remarque en fixant un instant particulier, par exemple $t = 5$, puis en observant comment se sont diffusés les trajectoires sur l'axe vertical (imaginaire) définit en ce point. Ainsi, les trajectoires du nombre d'infectés pour le graphe de degré 5 sont mois dispersées que celles des deux autres graphes. Nous pouvons également tracer les distributions à des instants précis pour rendre cette analyse plus rigoureuse, nous le ferons dans la section suivante.

Sur les graphes en forme d'étoile et de clusters, voir figure (2.24a) et (2.24b), nous remarquons que certaines trajectoires convergent vers des intervalles très spécifiques. Ceci se remarque à l'aide des courbes vertes, qui représentent le nombre d'individus rétablis. Nous observons ainsi comment les points d'accès entre amas d'individus modifient l'évolution du virus au sein de la population. Ces points d'accès peuvent représenter certaines infrastructures ou logiciels informatiques centralisant l'accès à d'autres réseaux. C'est entre autres le cas des pare-feux qui sous forme de logiciel ou de matériels informatiques permettent de sécuriser les réseaux locaux.

Sur la figure (2.25) nous pouvons donc observer comment le réseau de degré 5 est celui qui diffuse le virus le plus rapidement, suivi du réseau en étoile et du réseau en clusters.

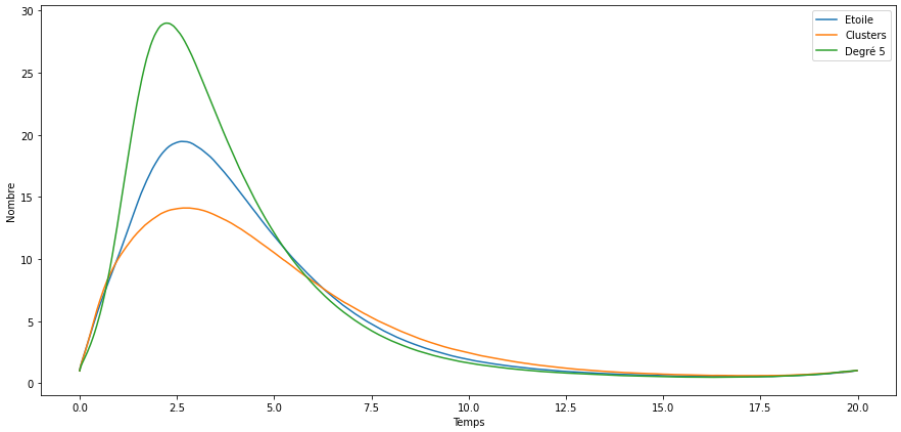


FIGURE 2.25 : Nombre moyen d'infectés par structure de réseau au cours du temps.

Dans les simulation précédentes, l'épidémie était initialisée aléatoirement pour chaque nouvelle trajectoire. Mais à présent nous cherchons à savoir quel est l'impact de la condition initiale sur la diffusion du virus dans la population.

2.3.3 Impact de la condition initiale

Dans cette section nous cherchons à montrer que non seulement la structure du réseau joue un rôle important dans la diffusion du virus mais que de la même façon, le point de départ du virus peut influencer sa diffusion.

2.3.3.1 Différentes conditions initiales sur un réseau internet

Pour cela nous allons considérer 20 entreprises connectées à un réseau internet dont la topologie suit l'approche décrite dans (ELMOKASHFI et al., 2010). Ces 20 entreprises possèdent entre 2 et 59 systèmes informatiques reliés à un réseau local d'entreprise, modélisé par un sous-réseau homogène.

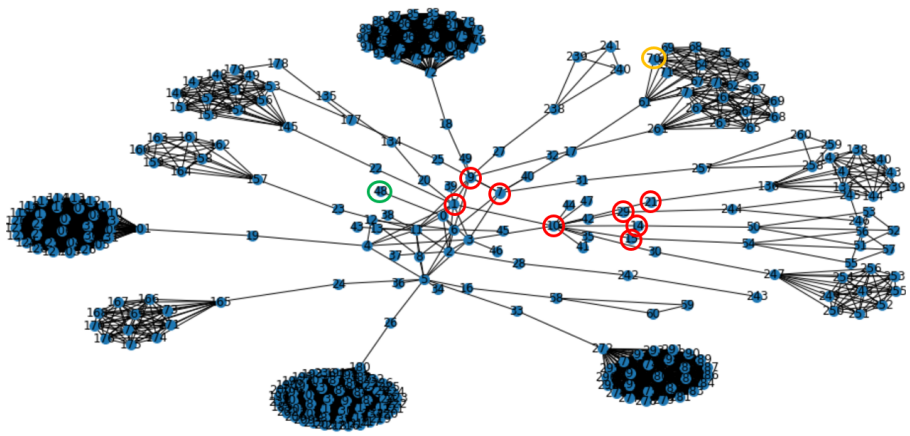
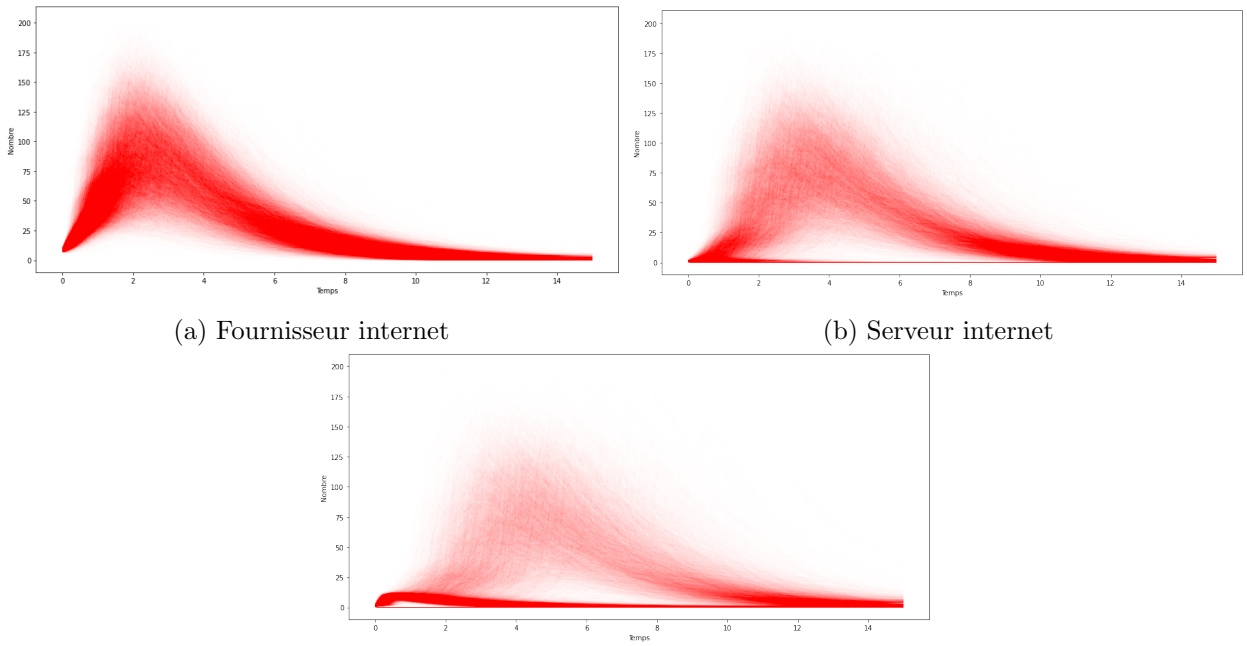


FIGURE 2.26 : Entreprises reliées à internet avec en cercles, les nœuds correspondant aux différentes conditions initiales.

Sur la figure (2.26) nous représentons le graphe utilisé pour modéliser les 20 entreprises reliées entre elles par internet. Les nœuds encerclés par des anneaux de couleurs correspondent aux trois conditions initiales que nous allons utiliser pour étudier la propagation du virus. En rouge, cette condition initiale peut correspondre à l'infection (partielle ou non) d'un fournisseur internet ou d'un gestionnaire de données auquel sont rattachées plusieurs entreprises. En jaune, le virus prend source au sein d'une entreprise de taille moyenne ce qui peut par exemple être causé par l'intrusion d'un individu malintentionné . Finalement, en vert, cette condition initialise le virus directement sur internet à un niveau élevé du réseau, ce qui peut par exemple correspondre à un serveur un peu isolé.

2.3.3.2 Nombre d'infectés au cours du temps

Nous avons ainsi propagé un virus de paramètre d'infection $\beta = 1$ pour lequel nous avons considéré un taux rétablissement $\gamma = 0.4$. Sur la figure (2.27), nous observons l'évolution du virus pour 10000 trajectoires en fonction de l'initialisation. Rappelons que sur les différentes figures, plus les traits de couleurs semblent vifs, plus le nombre de simulations suivant cette trajectoire sont importantes.



(a) Fournisseur internet

(b) Serveur internet

(c) Compagnie moyenne

FIGURE 2.27 : Trajectoires simulées pour le nombre d'infectés au cours du temps pour les différentes initialisations.

Nous remarquons ainsi que l'initialisation par fournisseur internet, figure (2.27a), favorise la diffusion du virus de façon plus certaine que les deux autres. A la suite, l'initialisation par serveur internet a plus tendance générer de grandes infections que si le virus avait été initialisé dans une compagnie moyenne. Notons cependant sur la figure (2.27c), que l'initialisation du virus dans une compagnie moyenne génère un grand nombre d'infecté dans les premiers instants du fait de la structure (homogène) du réseau local utilisé (par hypothèse) dans l'entreprise.

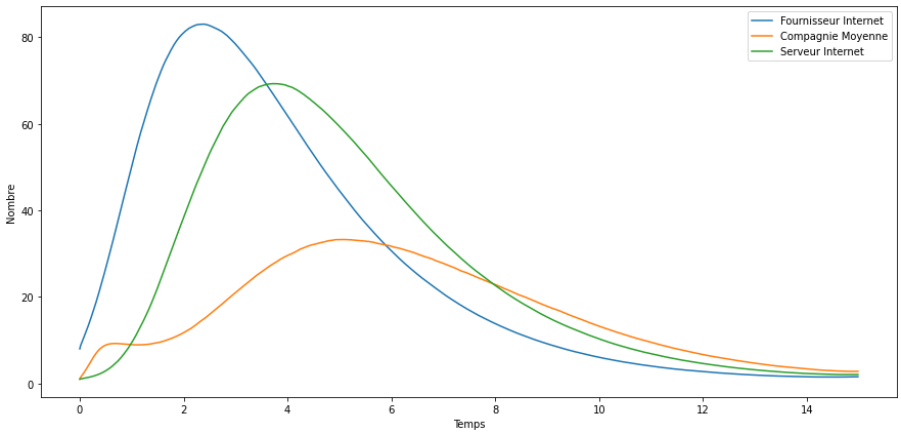
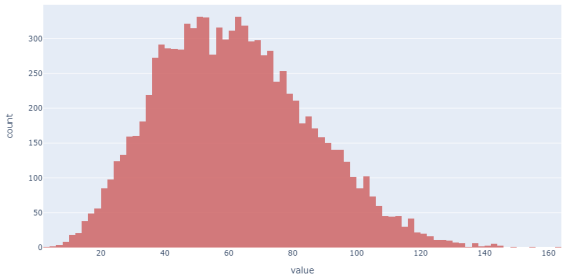
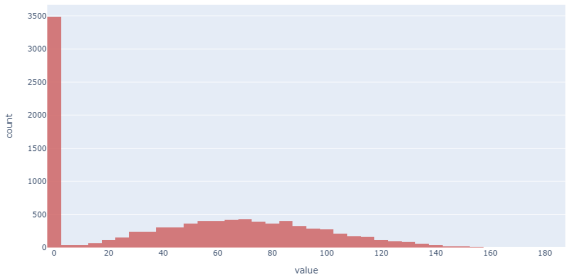


FIGURE 2.28 : Courbes moyennes du nombre d'infectés au cours du temps pour chacune des initialisations.

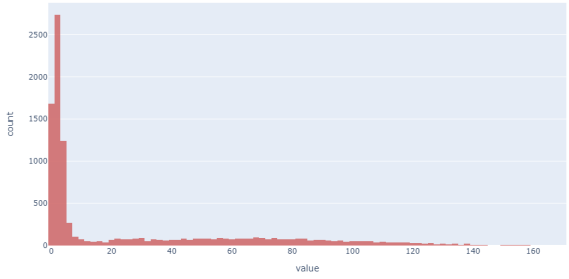
Comme mentionné précédemment lors de l'analyse de l'impact de la structure de réseau sur la diffusion du virus. Nous pouvons tracer les distributions du nombre d'infectés à des instants particuliers pour comparer plus rigoureusement, la diffusion du virus en fonction de l'initialisation (ou du graphe pour l'étude précédente) de ce dernier. Il est important de noter que sur la figure (2.29), l'échelle en abscisse n'est pas la même pour l'histogramme de la compagnie moyenne.



(a) Fournisseur internet



(b) Serveur internet



(c) Compagnie moyenne

FIGURE 2.29 : Distribution du nombre d'infectés au temps $t = 4$ en fonction de l'initialisation du virus.

De plus, nous avons vu plus en amont que dans certaines simulations, le virus est éliminé avant le dernier instant d'observation (ici $t_{final} = 15$). Ainsi, l'initialisation par fournisseur internet élimine le virus dans $44,2\%$ des cas, par Serveur Internet dans $53,83\%$ des cas et finalement, celle par une Compagnie moyenne dans $63,65\%$ des cas.

2.3.4 Algorithmes de simulation

Dans cette section nous allons présenter deux algorithmes qui permettent de simuler de façon exacte la propagation d'un virus au sein d'un graphe. Plus de détails sont fournis dans Kiss et al., 2017, et la librairie python utilisée est détaillée dans Miller et Ting, 2020. Les pseudo-codes de ces deux algorithmes sont fournis en annexes.

2.3.4.1 Algorithme de Gillespie :

L'algorithme de Gillespie est un algorithme très répandu pour la modélisation des processus de Markov où les objets changent de statut. Cet algorithme simule de façon macro l'état du réseau à plusieurs instants.

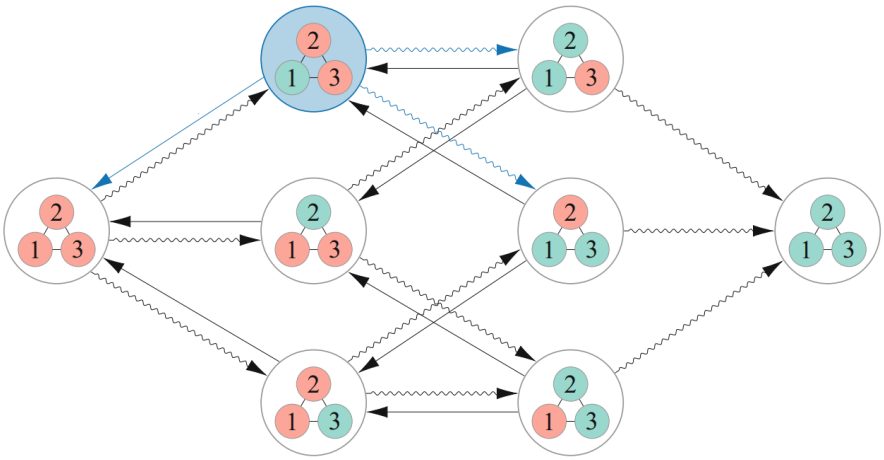


Figure 2.30 : Transitions possibles pour un SIS dans un réseau à trois nœuds (Kiss et al., 2017).

Nous avons vu précédemment que les temps de changement d'état d'un nœud sont distribués selon une loi exponentielle de paramètre :

- Nombre de voisins infectés $\times\ \gamma$ si le nœud en question est dans l'état susceptible (en vert sur la figure (2.30)).
- γ si le nœud en question est dans l'état infecté (en rouge sur la figure (2.30)).

Ainsi notre réseau changera d'état lorsque le premier nœud parmi les trois effectuera une transition vers un autre état. Si l'on se place dans la situation en bleu et à l'instant initial, pour connaître quelle sera le prochain état du réseau, nous pouvons tirer trois variables aléatoires exponentielles ayant un paramètre associé à l'état de chacun des nœuds : $E_1 \sim \mathcal{E}(2\tau)$ $E_2, E_3 \sim \mathcal{E}(\delta)$. Pour connaître le

nouvel état du réseau il suffit donc de modifier l'état du nœud dont le temps de changement d'état arrive en premier, par exemple si $\min(E_1, E_2, E_3) = E_1$ le réseau transite vers l'état d'en bas à gauche. L'algorithme de Gillespie propose d'utiliser le fait que le $\min(E_1, .., E_n)$ de plusieurs variables exponentielles de paramètre $\lambda_1, ..., \lambda_n$ est également distribué selon une loi exponentielle de paramètre égal à $\lambda_1 + ... + \lambda_n$, autrement dit :

$$\text{Soit} \quad E_1 \sim \mathcal{E}(\lambda_1), ..., E_n \sim \mathcal{E}(\lambda_n) \quad \text{alors :} \quad \min(E_1, .., E_n) \sim \mathcal{E}(\lambda_1 + ... + \lambda_n).$$

En effet, si nous notons $Y = \min(E_1, .., E_n)$, où $E_1 \sim \mathcal{E}(\lambda_1), ..., E_n \sim \mathcal{E}(\lambda_n)$, nous cherchons alors à connaître la loi de Y. En gardant en tête que la fonction de répartition $\forall x > 0, \quad F_Y(x) = \mathbb{P}(Y \leq x)$ caractérise la loi de Y, rappelons que la fonction de survie d'une variable aléatoire exponentielle s'écrit, $\forall i \in \{1, ..., n\}, \mathbb{P}(E_i > x) = \int_x^{+\infty} \lambda_i e^{-\lambda_i t} dt = e^{-\lambda_i x}$. Ainsi, les variables E_i étant indépendantes, $\mathbb{P}(Y > x) = \mathbb{P}(\forall i \in \{1, ..., n\}, E_i > x) = \prod_{i=1}^n \mathbb{P}(E_i > x) = \prod_{i=1}^n e^{-\lambda_i x} = e^{-x \sum_{i=1}^n \lambda_i}$. D'où, $F_Y(x) = 1 - e^{-x \sum_{i=1}^n \lambda_i}$. Nous reconnaissons la fonction de répartition d'une loi exponentielle de paramètre $\sum_{i=1}^n \lambda_i$. Finalement, nous avons bien $Y = \min(E_1, .., E_n) \sim \mathcal{E}(\lambda_1 + ... + \lambda_n)$.

Une fois que $\min(E_1, .., E_n)$ est calculé, il faut déterminer si la variable génère une infection ou un rétablissement. Pour cela nous tirons une variable uniforme sur $(0, \lambda_1 + ... + \lambda_n)$ puis si la réalisation se trouve en dessous d'un certain seuil, alors on considère qu'il s'agit d'une infection. Ce seuil correspond à la somme de tous les λ_i qui se rattachent a des taux d'infections.

Pour déterminer quel nœud parmi va changer d'état (s'infecter ou se rétablir) il faut distinguer les deux cas :

- si le nœud va se rétablir, alors on choisit aléatoirement un nœud parmi les nœuds infectés.
- si le nœud va s'infecter, nous choisissons aléatoirement un nœud parmi les nœuds susceptibles en pondérant les probabilités par sa contribution au taux d'infection total.

Le grand inconvénient de cet algorithme c'est que la complexité algorithmique pour déterminer quel nœud parmi les susceptibles s'infecte, allonge le temps de calcul.

2.3.4.2 Algorithme *Event-Driven* :

L'algorithme suivant permet de simuler la propagation du virus au sein du réseau. Il se base sur le principe "d'évènements" et adopte une vision plus micro du processus SIR. Un évènement se caractérise par un nœud, un temps de survenance, une action (transmission ou rétablissement) et un nœud source (pour les cas de transmission). Les évènements sont rangés dans une file de priorité, qui ordonne les évènements par date de survenance. Si un évènement possède un temps de survenance supérieur au temps d'arrêt T_{fin} défini par l'utilisateur, alors il n'est pas ajouté à file $\mathcal{Q}$.

Chaque nœud est associé à un statut (infecté ou susceptible) et à un temps de rétablissement.

Lors de l'initialisation de l'algorithme, tous les nœuds initialement infectés créent des évènements d'infection. Ainsi si le nœud u est initialement infecté, nous créons un évènement Event $\leftarrow$ {nœud : u, temps : 0, action : transmit, source : v} où le temps de rétablissement de v est 0. Puis nous ajoutons cet évènement à la file de priorité $\mathcal{Q}$.

Pour la suite de l'algorithme, nous prenons "E_1", le premier élément dans la pile, puis si :

- il s'agit d'un rétablissement, alors nous changeons le statut du nœud en susceptible.

- il s'agit d'une infection alors :
  - Nous changeons le statut du nœud en infecté.
  - Nous tirons une variable exponentielle de paramètre γ dont la réalisation devient son temps de rétablissement. Nous créons donc un nouvel évènement associé à son rétablissement newEvent $\leftarrow$ {nœud : E_1.nœud, temps : la réalisation de la variable précédente , action : rétablissement} que nous ajoutons à la file $\mathcal{Q}$.
  - Nous cherchons à transmettre le virus aux voisins de $E_1.nœud$ qui sont toujours susceptibles. Pour chacun de ces nœuds voisins, nous allons donc tirer une nouvelle variable exponentielle de paramètre β afin de définir un nouveau temps d'infection du nœud susceptible T'_{inf}. Si cette nouvelle date, T'_{inf}, est inférieure à la date de rétablissement du nœud infectieux mais également au temps d'infection que le nœud avait déjà enregistré (cela arrive quand le nœud susceptible à déjà reçu un contact infectieux provenant d'un autre nœud), alors T'_{inf} devient la nouvelle date d'infection du nœud susceptible générant ainsi un nouvel évènement d'infection.

Nous procédons ainsi de suite, jusqu'à ce que la file de priorité $\mathcal{Q}$ soit vide. Il est important de noter que ce dernier algorithme permet de simuler des modèles dont les temps de changement ne seraient plus distribués selon des lois exponentielles.

Dans cette section nous avons illustré la façon dont certains paramètres peuvent influencer la propagation du virus. Ainsi, un fort taux de contagion accélère sa diffusion tandis qu'un fort taux de rétablissement la ralentit. La structure du réseau utilisée influe également sur la diffusion, de façon générale, plus les individus du réseau sont connectés entre eux, plus le virus se propagera rapidement. Finalement, sur certains réseaux, la condition initiale peut impacter les différentes façons la propagation du virus.

Dans le prochain chapitre nous nous focaliserons sur le cyber silencieux et chercherons à exploiter les éléments de modélisation que nous avons apporté dans ce chapitre pour étudier des évènements d'accumulation cyber.

Chapitre 3

Application au silent cyber

Dans ce chapitre nous nous intéressons au risque d'accumulation (voir, section 1.2.2.2) porté par le cyber silencieux. Comme présenté dans le Chapitre 1, le cyber silencieux intervient sur les garanties n'excluant ou n'affirmant pas le risque cyber. Ainsi, certaines garanties peuvent se retrouver sous évaluées vis à vis du risque qu'elles portent, ce qui peut avoir de lourdes conséquences pour les assureurs. Nous allons donc commencer par présenter une méthode pour estimer l'exposition au silent cyber que nous adapterons pour pouvoir appliquer nos modèles dynamiques de pandémie. Nous réaliserons ensuite une étude pratique sur un portefeuille fictif pour estimer le nombre d'infectés et les pertes probables d'un tel évènement sur un portefeuille d'assurance suivant le temps d'intervention et la structure du portefeuille.

3.1 Évaluation de l'exposition cyber silencieuse

A la suite d'une série de réunions autour du cyber ayant eu lieu entre octobre 2015 et juin 2016, l'autorité de régulation prudentielle (PRA) du Royaume Unis publie le 14 novembre 2016 un document de consultation sur le risque de souscription lié à l'assurance cyber (CP39/16, voir PRA, 2016). Ce document introduit notamment la notion de risque cyber silencieux mais propose également une première déclaration prudentielle (*supervisory statement*) qui établit les premières attentes en terme de gestion du risque de souscription cyber. Le 5 juillet 2017, la PRA publie PS15/17 (PRA, 2017), une déclaration de politique générale (*policy statement*) qui répond aux éventuelles remarques et questions qui avaient pu être faites sur CP39/16 mais qui surtout, introduit une dernière déclaration prudentielle qui depuis fait office de référence. Suite à ces initiatives, la PRA publie le 30 janvier 2018 les résultats d'une enquête mettant en lumière certains axes d'améliorations sur la gestion de l'exposition au risque cyber, (PRA et Bank of England, 2019). C'est dans ce contexte prudentiel, que l'Institute and Faculty of Actuaries (IFoA) publie début Janvier 2020, un cadre d'évaluation pour le cyber silencieux, voir Cartagena et al., 2020. Dans la section suivante nous présenterons plus en détail ce cadre d'évaluation.

3.1.1 Cadre d'évaluation du cyber silencieux (IFoA)

En 2020, l'Institut and Faculty of Actuaries (IFoA) publie dans le British actuarial journal un *discussion paper* pour construire un cadre d'évaluation du silent cyber que nous présentons dans cette section (Cartagena et al., 2020). En complément de la présentation de leur cadre d'évaluation, les

auteurs dirigent le lecteur vers un tableau Excel disponible sur internet pour l'illustrer, les prochaines figures en découlent. Ainsi, ce cadre se décompose en trois grandes étapes :

(i) La première consiste a définir la mesure d'exposition par garantie au sein du portefeuille que l'on cherche à traiter.

(ii) La deuxième vise à quantifier les pertes potentilles en fonction de plusieurs scénarios plausibles.

(iii) La dernière étape décrit quelques points pour inscrire ce cadre d'évaluation dans le *reporting* et en faire un processus à part entière.

Dans la suite, nous nous concentrerons sur les deux premières parties citées précédemment. Le lecteur trouvera en annexe (A.8), toutes les étapes détaillées qui composent le cadre d'évaluation.

L'un des objectifs du cadre proposé par l'IFoA, est de servir de référence, de guide, pour que chaque entité soit en mesure d'établir sa propre évaluation du risque cyber silencieux. Ainsi, nous supposons disposer d'un portefeuille constitué de plusieurs polices de nature diverses (MRH, Multirisques, etc...) pour lesquels nous connaissons :

(i) Sa *Line of business* (LoB) ou branche selon l'annexe 1 dans la directive du Parlement Européen et Conseil de l'Union Européenne, 2009 (voir Annexe A.6).

(ii) Son exposition théorique, définie par l'assureur. Cette exposition peut par exemple correspondre à la limite d'indemnisation contractuelle.

(iii) S'il possède (ou non) une *cyber sub limit,* c'est a dire une limite d'indemnisation spécifique pour le risque cyber.

Pour mesurer l'exposition des différents contrats au cyber silencieux, nous commençons par créer une matrice des clauses. Cette matrice fait le croisement entre la fréquence d'utilisation d'une clause et les différentes *LoBs*. Les clauses et les *LoBs* considérées dans l'évaluation du risque doivent être préalablement définies et en adéquation avec les contrats étudiés.

Au Royaume-Unis des organismes comme la *London Market Association* (LMA), l'*International Underwriring Association* (IUA), l'*International Association of Engineering Insurers* (IMIA) et d'autres, construisent des clauses d'assurances qui sont par la suite utilisées dans les contrats. Citons par exemple la clause CL380, clause d'exclusion des risques cybernétiques, très utilisée dans le domaine maritime (voir l'annexe A.7 pour des exemples de clause).

Comme montré sur la figure (3.1), une fois cette matrice construite, nous avons donc une correspondance directe entre les branches d'activités et la fréquence d'utilisation des clauses considérées.

#	Wordings Intention LMA Classes	Exclusion LMA5272/3/4/5 Cyber Incident Exclusion	Affirmative LMA3141 Electronic and Computer Crime Policy	Exclusion NMA2914/5 Electronic Data Endorsement	Affirmative LSW555 Aviation Hull "War and allied perils"	Affirmative AVN52G Extended Coverage Endorsement	Exclusion AVN48B War/Hijacking and other perils exclusion	Exclusion ANV124 Data Event Clause
1	Aviation Hull						Very High	Unknown
2	Aviation Liability					Very High	Very High	Unknown
3	Aviation War				Very High	Very High	Very High	Unknown
4	Casulaty RI	Very Low						

Figure 3.1 : Matrice des Clauses/*LoBs*.

Sur la figure (3.1), nous disposons pour chaque line of business, de la fréquence d'utilisation des clauses.

Nous associons également à chaque label de fréquence (*low, very low, high, very high,...*) un pourcentage. Connaissant ainsi la branche à laquelle appartient un contrat, nous sommes donc en mesure d'associer à chaque police une fréquence d'utilisation par clause. Ainsi, pour l'Aviation de Guerre, la clause AVN48B est classée *Very High* ce qui correspond à une fréquence de 88%.

						2	2	2	2	2	2	1	1	2	2
#	ID	UW Year	Inception Date	Expiry Date	LoB	NMA2918	NMA2914/5	NMA2914/5 A	NMA2912/8	CL380	JS2015/8	LSW555	AVN52G	AVN48B	LMA5359
1	FAC2243372	2019	21/01/2019	21/01/2020	Property UK Commercial	0%	88%	88%	0%	3%	0%	0%	0%	0%	0%
2	FAC3627392	2018	27/07/2018	27/07/2019	Livestock & Bloodstock	0%	0%	0%	0%	88%	0%	0%	0%	0%	0%
3	FAC6870284	2019	13/06/2019	12/06/2020	Aviation War	0%	0%	0%	0%	0%	0%	88%	88%	88%	0%
4	FAC9580164	2019	13/12/2019	12/12/2020	Livestock & Bloodstock	0%	0%	0%	0%	88%	0%	0%	0%	0%	0%
5	FAC1638000	2019	04/07/2019	03/07/2020	Property D&F	0%	88%	88%	0%	0%	0%	0%	0%	0%	0%
6	FAC873051	2018	23/06/2018	23/06/2019	Aviation Hull	0%	0%	0%	0%	0%	0%	0%	0%	88%	0%
7	FAC5957838	2018	09/07/2018	09/07/2019	Marine War	0%	0%	0%	0%	63%	0%	0%	0%	0%	0%

FIGURE 3.2 : Fréquence d'utilisation des clauses par contrats.

Parmi les clauses cyber, certaines couvrent explicitement les évènements cyber (en vert avec le numéro 1 sur la première ligne de la figure 3.2) tandis que d'autres l'excluent (en rouge avec le numéro 2 sur la première ligne de la figure 3.2). A présent, nous allons définir pour chacun des contrats la vraisemblance d'utilisation de clause comme étant la moyenne des pourcentages d'utilisation des clauses, nous donnerons un exemple dans la suite. Nous définissons également la part des clauses affirmant et excluant le risque cyber par contrat.

							Likelihood Policy is A/Ex if present		2	2	2	2	2	2	1	1	2	2
#	ID	UW Year	Inception Date	Expiry Date	LoB	Likelihood Use of Clause	Affirmative	Excluded	NMA2918	NMA2914/5	NMA2914/5 A	NMA2912/8	CL380	JS2015/8	LSW555	AVN52G	AVN48B	LMA5359
1	FAC1320789	2018	06/03/2018	06/03/2019	Property UK Commercial	60%	0%	100%	0%	88%	88%	0%	3%	0%	0%	0%	0%	0%
2	FAC2762216	2019	03/04/2019	02/04/2020	Livestock & Bloodstock	90%	0%	100%	0%	0%	0%	0%	88%	0%	0%	0%	0%	0%
3	FAC4068159	2019	08/04/2019	07/04/2020	Aviation War	90%	70%	30%	0%	0%	0%	0%	0%	0%	88%	88%	88%	0%
4	FAC6708027	2018	13/06/2018	13/06/2019	Livestock & Bloodstock	90%	0%	100%	0%	0%	0%	0%	88%	0%	0%	0%	0%	0%
5	FAC829245	2018	08/11/2018	08/11/2019	Property D&F	90%	0%	100%	0%	88%	88%	0%	0%	0%	0%	0%	0%	0%
6	FAC9890050	2019	05/01/2019	05/01/2020	Aviation Hull	90%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	88%	0%
7	FAC6060499	2018	02/11/2018	02/11/2019	Marine War	70%	0%	100%	0%	0%	0%	0%	63%	0%	0%	0%	0%	0%

FIGURE 3.3 : Vraisemblance et part d'utilisation des clauses par contrat.

Une fois la vraisemblance d'utilisation des clauses établie (*Likelihood Use of Clause*), nous définissons la vraisemblance au cyber silencieux (*Likelihood Silent*) comme étant la part non captée par la vraisemblance d'utilisation de clause.

$$\text{vraisemblance au cyber silencieux} = 100\% - \text{vraisemblance d'utilisation des clauses}$$

						Output Percentages (adjusted for A/Ex)				Likelihood Policy is A/Ex if present	
#	ID	UW Year	Inception Date	Expiry Date	LoB	Affirmative	Excluded	Likelihood Silent	Likelihood Use of Clause	Affirmative	Excluded
1	FAC7770471	2019	20/02/2019	20/02/2020	Property UK Commercial	0%	60%	40%	60%	0%	100%
2	FAC2618994	2019	10/01/2019	10/01/2020	Livestock & Bloodstock	0%	90%	10%	90%	0%	100%
3	FAC6603963	2019	02/05/2019	01/05/2020	Aviation War	63%	27%	10%	90%	70%	30%
4	FAC3064621	2019	01/08/2019	31/07/2020	Livestock & Bloodstock	0%	90%	10%	90%	0%	100%
5	FAC724903	2018	28/02/2018	28/02/2019	Property D&F	0%	90%	10%	90%	0%	100%
6	FAC4389555	2018	07/01/2018	07/01/2019	Aviation Hull	0%	90%	10%	90%	0%	100%
7	FAC3156310	2018	05/09/2018	05/09/2019	Marine War	0%	70%	30%	70%	0%	100%

FIGURE 3.4 : Vraisemblance au cyber silencieux, aux clauses affirmatives et aux clauses d'exclusion.

L'utilisation de clause étant valable pour une affirmation ou une exclusion du risque cyber. Nous répartissons donc la vraisemblance d'utilisation de clause entre ces deux catégories en la multipliant

par la part des clauses. Par exemple, pour la ligne 3 sur la figure (3.4 et 3.3), nous pouvons observer que deux des trois clauses qui sont utilisées dans la branche d'aviation de guerre sont des clauses d'affirmation du cyber (LSW555 et AVN52G pour l'affirmation et AVN48B pour l'exclusion), d'où les pourcentages 70% et 30%. Si nous faisons la moyenne des fréquences d'utilisation de ces trois clauses, nous retrouvons bien (après arrondi) 90% pour la vraisemblance d'utilisation de clause et donc 10% pour la vraisemblance cyber silencieux. Pour retrouver les 63% et 27% il suffit donc de multiplier 90% par 70% et 30% respectivement. Nous ignorons pourquoi les auteur décident de conserver un arrondi aussi important.

Nous pouvons déduire la part de l'exposition potentiellement liée au cyber affirmatif, silencieux ou même, la part qui ne sera pas concernée par le cyber. Pour ce faire nous multiplions les pourcentages précédents par l'exposition théorique du contrat. Par exemple, si l'exposition du contrat en ligne 3 est de 1M€, alors l'exposition au cyber silencieux serait de 100 000€(10% de 1M€).

Une fois l'exposition théorique déterminée pour chacune des polices, il est possible d'agréger les montants exposés au silent par *LoB*.

Comme nous l'avons précisé dans l'introduction de cette section, les auteurs proposent de construire des scénarios pour déterminer quelles pourraient être les pertes engendrés. Pour cela, ils établissent un tableau permettant d'associer, par le biais d'une fréquence, les différentes *LoB* et plusieurs garanties considérées. Indépendamment, la construction de scénarios passe par l'énumération des garanties à considérer mais également par la définition d'autres caractéristiques qui lui seraient exhaustives. Finalement, pour évaluer quel serait le montant mobilisé, ils croisent les informations sur l'exposition et sur le scénario en question.

Notons que le cadre d'évaluation proposé par l'IFoA est statique dans le sens où la composante temporelle n'est pas considérée dans les scénarios d'évaluation des pertes.

3.1.2 État de la situation en France

En France, l'ACPR a publié en novembre 2019 un communiqué de presse sur la distribution des garanties cyber par les assureurs (ACPR, 2019). Ils y exposent notamment que, "les organismes ne mesurent pas encore suffisamment leur exposition, notamment à travers les garanties implicites contenues dans les contrats en cours". Ils identifient quatre axes d'améliorations :

(i) Evaluer de façon exhaustive l'exposition du portefeuille au risque cyber, notamment en termes de garanties implicites ; si c'est pertinent, intégrer l'évaluation au rapport ORSA.

(ii) Clarifier les définitions et la terminologie relatives aux risques, pour permettre une offre exempte d'ambiguïté vis-à-vis des preneurs d'assurance.

(iii) Construire progressivement les bases statistiques qui permettront de mieux délimiter les garanties et de les tarifer de façon pertinente.

(iv) Sensibiliser et former les acteurs au risque cyber, tant du côté des assurés que des forces commerciales (articulation promotion / prévention).

A notre connaissance, aucun cadre d'évaluation n'a encore été proposé pour le marché français.

Contrairement à ce qui se fait au Royaume-Unis, en France il n'y a pas d'organisme chargé de publier des clauses spécifiques, et la transcription des clauses étrangères (comme celles publiées par

la LMA) n'est pas triviale. Au delà du caractère très apparent qui leur est exigé sur les polices d'assurance (article L. 112-4 du Code des assurances), l'article L.113-1 du Code des assurances prévoit que les exclusions de garantie ne sont opposables à l'assuré qu'à condition d'être formelles et limitées. Le caractère limité ne doit pas conduire à annihiler ou à trop réduire significativement la garantie que l'assureur est censé avoir accordé à l'assuré, l'idée étant que l'assuré ne puisse être trompé (Pennec & Michau, s. d.). Quant au caractère formel, il vise à rendre les clauses claires, précises et sans ambiguïté aucune. Nous pouvons retenir qu'une clause d'exclusion de garantie ne peut être formelle et limitée dès lors qu'elle doit être interprétée (Cour de Cassation - Chambre civile 1, 2001). La rédaction d'une nouvelle clause demande donc une expertise transverse aux métiers de l'assurance, afin de garantir sa légitimité sous la jurisprudence en vigueur, tout en conservant son rôle premier qui est celui de limiter, voir exclure, le risque.

Le cyber silencieux reste un sujet d'actualité en France, en effet, lors de notre participation au Printemps de l'Assurance 2022, organisé par l'Université Paris Dauphine, nous avons pu avoir la confirmation de différents acteurs assurantiels sur la poursuite d'efforts dans le traitement du cyber silencieux. Ce sujet est adressé pour les affaires nouvelles chez certains acteurs mais il reste tout de même un stock sur lequel il faut encore travailler.

Le cadre d'évaluation proposé par l'IFoA (Cartagena et al., 2020) permet d'estimer les pertes liées au cyber silencieux dans un cadre très général. Il permet essentiellement d'uniformiser au niveau du Royaume-Uni, les étapes à inclure pour une évaluation complète des risques liés au cyber silencieux. Cependant, le marché français ne dispose pas d'acteurs comme la LMA, l'IUA et l'IMIA qui proposent des clauses pour les polices d'assurance. Ceci pénalise fortement l'application telle quelle dû cadre d'évaluation proposé par l'IFoA au marché français. Le 7 septembre 2022, la Direction Générale du Trésor a publié un rapport (Direction générale du Trésor, 2022) dans lequel les auteurs soulignent l'importance de la rédaction des clauses de couverture et d'exclusion, voir section (3.3.2.1).

3.1.3 Proposition d'un cadre d'évaluation

Ne pouvant donc pas transposer tel quel le cadre d'évaluation proposé par l'IFoA, nous avons décidé de nous adapter à ce qui pourrait être la situation d'une compagnie d'assurance en France. Pour des soucis de clarté, nous distinguerons dans la suite le type de police du libellé du contrat d'assurance : le contrat sera une matérialisation de la police d'assurance. Cette distinction provient du fait que, par exemple, une police MRH standard de 2015 n'est pas construite de la même façon qu'une police MRH standard de 2022 du fait des changements législatifs, des nouveaux risques, etc. Dans cet exemple, la date sert de critère discriminant mais (afin de rester le plus général possible) d'autres critères plus complexes pourraient être pris en compte pour faire cette distinction. De plus, nous supposons pouvoir disposer des expositions théoriques pour chacune des garanties.

Nous proposons d'évaluer le risque lié au cyber silencieux en 4 étapes. Afin d'illustrer les quatre étapes de la figure suivante, nous allons considérer une base de donnée fictive, inspirée de données réelles, contenant plusieurs polices d'assurance pour lesquelles nous disposons des garanties qu'elles incluent et des expositions théoriques pour chacune des garanties. Ce type de données peut être le résultat d'une agrégation de plusieurs portefeuilles chez un assureur.

FIGURE 3.5 : Étapes d'évaluation du risque cyber silencieux.

Sur la figure (3.6), nous avons dans la colonne de gauche, trois exemples de produits, l'Auto 2 roues, l'Auto 4 roues et la MRH standard, mais d'autres produits constituent cette base. Les autres colonnes représentes les différentes garanties qui composent les produits d'assurance, leurs signification n'est pas essentielle à la compréhension du cadre d'évaluation. Cependant, nous apporterons au fur et à mesure de la présentation du cadre d'évaluation, plus d'éléments explicatifs sur cette base.

3.1.3.1 Étape 1 : l'exposition

En s'adaptant aux données disponibles, la première étape consiste à définir une mesure d'exposition adéquate. Dans notre exemple, l'exposition correspondra au montant réellement indemnisable par l'assureur. Par exemple, pour une garantie contenant une limite d'indemnisation, l'exposition correspond à cette limite.

Dans le cadre proposé par l'IFoA, les auteurs transfèrent l'exposition des polices au *LoBs* puis aux garanties. Comme nous le précisions plus haut, nous supposons disposer des données transcrites dans la figure (3.6), ce qui est représentatif de données réelles d'assurance. Cependant, afin que notre cadre soit le plus général possible, nous supposons que les structures données accessibles chez les assureurs peuvent être différentes.

Ainsi, nous définissons à quel niveau est ce que l'on va agréger les données. Cette agrégation peut par exemple se faire au libellé des contrats, notamment si chaque type de police dispose de plusieurs contrats différents : pour une police auto 2 roues nous disposons dans le portefeuille de contrats A2_1, A2_2, A2_3 que nous agrégeons en un groupe de contrat A2. Dans le cadre de notre exemple, nous ne regroupons pas les données puisque le niveau de granularité est déjà assez bas, c'est à dire que nous avons déjà les informations par police (Auto 2 roues, MRH, etc...), voir figure (3.6). L'idée derrière ce regroupement est de pouvoir travailler à un niveau qui nous permette d'obtenir une vision globale sur l'exposition au cyber silencieux par police.

Exposition par produits et garanties	2260 : RC - 2 ROUES CORPORELS	2265 : RC - 2 ROUES MATERIELS	2299 : GLOBAL RC AUTOS	2310 : DOMMAGES AUTO - 4 ROUES	2311 : PROT.CORP ORELLE AUTO	2320 : DOMMAGES - CARAVANE	2360 : DOMMAGES AUTO - 2 ROUES	2361 : PROT.CORP ORELLE 2 roues	2381 : DOMMAGES ACCIDENTEL S UTILITAIRES AGRICO	2399 : GLOBAL DOMMAGE AUTO	2399 : GLOBAL DOMMAGES AUTOS	2400 : RC MRH
A2	181 290,00	67 390,00	63 250,00	-	-	-	153 350,00	60 980,00	-	-	93 520,00	-
A4	-	-	84 740,00	253 530,00	242 080,00	-	-	-	-	-	86 530,00	-
SMRH												200 000,00

FIGURE 3.6 : Polices et expositions théoriques par garanties en €.

Ainsi pour chaque police, nous obtenons une exposition pour chacune des garanties qu'elle contient.

3.1.3.2 Etape 2 : Matrice des clauses

Lors de cette étape, l'objectif est de pouvoir associer à chaque garantie une fréquence d'utilisation de clause d'exclusion ou d'affirmation du risque cyber afin d'en déduire un taux silencieux. Nous allons donc regarder pour chaque garantie la fréquence avec laquelle une clause cyber est utilisée ou non, et ce, pour chaque élément du niveau d'agrégation défini précédemment, dans notre cas servant d'exemple : pour chaque type de police.

Ce travail d'évaluation de la fréquence d'utilisation de clause peut éventuellement être approximé. Dans notre exemple nous disposons du type de police (Auto 2 roues, Auto 4 roues, MRH Standard, etc...) qui définissent des classes. Ces dernières sont constituées de plusieurs contrats eux-mêmes construits différemment les uns des autres (SMRH_1, SMRH_2, SMRH_3, etc). Supposons que nous disposons en interne des proportions suivantes :

TABLE 3.1 : Exemple des proportions constituant la police MRH standard.

Contrat	Proportion
SMRH_1	30 %
SMRH_2	20 %
SMRH_3	50 %

De plus, supposons que les fréquences d'utilisation de clause pour la garantie "Incendie MRH Standard" sont celles décrites dans la table qui suit :

TABLE 3.2 : Exemple de fréquence d'utilisation de clauses par type de contrat pour la garantie Incendie MRH standard.

Contrat	Fréquence de clauses cyber affirmatives	Fréquences de clauses cyber exclusives
SMRH_1	20 %	20 %
SMRH_2	5 %	70 %
SMRH_3	20 %	80 %

Notons que nous considérons des fréquences, car certaines garanties possèdent des options complémentaires mais la fréquence peut tout simplement être remplacée par la présence ou non de clause.

Ainsi, en utilisant les informations de la Table (3.3) et la Table (3.2), nous pouvons obtenir les fréquences d'utilisation de clauses affirmatives ou exclusives par type de police.

TABLE 3.3 : Exemple de calcul pour la garantie “Incendie MRH standard”.

Type de police	Fréquence de clauses affirmatives	Fréquences de clauses d'exclusion
SMRH	$20\%\times30\% + 5\% \times 20\% + 20\% \times 50\%$	$20\%\times30\% + 70\% \times 20\% + 80\% \times 50\%$
i.e.	$= 17\ \%$	$= 60\ \%$

Nous avons donc une correspondance entre le type de police et la fréquence d'utilisation, ou non, de clause par garantie.

Nous définissons finalement le **taux silencieux** comme la fréquence de non utilisation de clause, d'où pour la police SMRH nous avons :

$$\text{Taux\_Silent\_SMRH} = 100\% - 17\% - 60\% = 23\%.$$

Produits	Total Garanties	2310 : DOMMAGES AUTO - 4 ROUES	2311 : PROT.CORP ORELLE AUTO	2319 : TAXE ATTENTAT 4 ROUES	2320 : DOMMAGES - CARAVANE	2360 : DOMMAGES AUTO - 2 ROUES	2361 : PROT.CORP ORELLE 2 roues	2380 : DOMMAGES AUTO - 4 ROUES	2381 : DOMMAGES ACCIDENTELS UTILITAIRES AGRICO	2389 : TAXE ATTENTAT 4 ROUES	2399 : GLOBAL DOMMAGE AUTO
A2	4					1	1				
Affirmation		0%	0%	0%	0%	20%	20%	0%	0%	0%	0%
Exclusion		0%	0%	0%	0%	20%	60%	0%	0%	0%	0%
Silent		100%	100%	100%	100%	60%	20%	100%	100%	100%	100%
A4	5	1	1								
Affirmation		10%	10%	0%	0%	0%	0%	0%	0%	0%	0%
Exclusion		50%	60%	0%	0%	0%	0%	0%	0%	0%	0%
Silent		40%	30%	100%	100%	100%	100%	100%	100%	100%	100%

FIGURE 3.7 : Exemple d'application par type de police : Auto 2 roues et Auto 4 roues.

Sur la figure (3.7), les cases en vert foncé contenant le chiffre 1, représentent les garanties qui sont inclues dans la police d'assurance. Si la case est vide (couleur blanche), alors la garantie n'est pas incluse dans la police.

Ce taux va nous permettre de quantifier l'importance du cyber silencieux dans la police et pour chacune de ses garanties.

3.1.3.3 Etape 3 : Application de la matrice aux expositions

L'objectif de cette étape est de déterminer l'exposition théorique au silent cyber pour chaque garantie. Pour cela nous allons distribuer l'exposition par garantie calculée à l'étape 1 aux différentes clauses, affirmatives et exclusives, puis nous en déduisons la part exposée au cyber silencieux. Par exemple, pour une exposition de 1M€sur la garantie “Incendie MRH standard”, nous avons 170 000€(i.e. $1M \times 17\%$€) d'exposition théorique au cyber affirmatif, 600 000€(i.e. $1M \times 60\%$€) non imputables au risque cyber et donc potentiellement 230 000€(i.e. $1M - 170K - 600K$€) resteraient exposés à un risque cyber non-affirmatif (cyber silencieux). Cependant il faut tenir compte de l'agrégation qui a pu être faite à l'étape 1 et 2 par type de police, on suppose donc que l'exposition dans les données est répartie de façon homogène sur l'ensemble des contrats constituant le type de police. Si ce n'est pas la cas, alors l'agrégation devra se faire une fois que les expositions théoriques affirmatives, exclusives et silencieuses sont déterminées.

3.1.3.4 Etape 4 : Application de scénarios

L'objectif de cette dernière étape est d'estimer la perte probable de l'assureur suivant des scénarios. Nous construisons des scénarios afin d'évaluer les impacts qu'ils pourraient avoir sur les garanties.

Pour ce faire, nous reprenons la méthode des scénarios statiques et déterministes comme appliquée dans le cadre d'évaluation de l'IFoA. Mais nous pouvons tout de même pousser la modélisation (statique de l'IFoA) en appliquant une distribution de probabilité sur l'exposition par garantie et par type de police sous condition de disposer des éléments nécessaires pour justifier le choix des distributions appliquées. Ainsi nous pouvons obtenir une perte probabilisée.

Un exemple de scénario serait une attaque cyber qui entraînerait un incendie dans une entreprise. Ainsi, plusieurs garanties pourraient être activée par ce sinistre comme la garantie incendie professionnelle, perte d'exploitation, RC professionnelle, assistance professionnelle ou encore la protection juridique. Nous apporterons plus de détails dans la section des compléments.

Dans la prochaine section du chapitre, nous construirons un scénario centré sur la perte d'exploitation pour une police multirisques professionnels. Nous nous concentrerons donc sur des scénarios basés sur l'utilisation de Malwares qui pourraient entraîner des phénomènes d'accumulation. Mais avant, nous allons apporter quelques remarques et compléments à la modélisation précédente.

3.1.4 Remarques et compléments

3.1.4.1 Des scénarios statiques

Dans l'article publié par l'IFoA (Cartagena et al., 2020), la modélisation des scénarios se base sur des hypothèses déterministes en utilisant les montants estimés sur les expositions silencieuses. Dans son mémoire (de Brive, 2022), G. Beaud de Brive construit un scénario "Black-out" inspiré de celui construit par les Lloyd's of London et l'université de Cambridge (Ruffle et al., 2015) afin de quantifier l'exposition des couvertures d'assurance pour le risque d'accumulation cyber. Dans la mise en place de ce scénario, G. Beaud de Brive établit une courbe représentative du temps passé sans électricité et propose certaines couvertures qui seraient directement impactées. Les montants exposés dans sa modélisation pourraient donc être enrichis par ceux relevant du silent cyber estimés à l'aide du cadre d'évaluation présenté dans la section précédente (section 3.1.3). De la même façon mais dans un autre mémoire (Laurent, 2022), R. MATERA LAURENT construit des scénarios avec un outil interne pour directement évaluer l'impact qu'ils auraient sur des couvertures exposées au cyber silencieux dans certaines captives d'assurance.

Au delà des scénarios présentés dans les mémoires précédents, les Lloyd's of London publient chaque année leur "Realistic Disaster Scenarios (RDS) specifications" (Lloyd's, 2022) avec un volet consacré au risque cyber. Pour chaque scénario décrit, le document inclut diverses informations comme la description de l'évènement, les pertes, les secteurs les plus impactés, etc... Dans le volet cyber, nous retrouvons le scénario Business Blackout II, qui est dans la continuité du scénario utilisé par G. Beaud de Brive dans son mémoire, mais également deux autres scénarios comme le Cloud Cascade et le Ransomware Contagion.

En France, le Forum des Compétences "a pour objectif d'organiser une compétence globale autour de la sécurité des systèmes d'information des acteurs de la Banque, de la Finance et des Assurances et d'ériger le sujet au rang de Culture d'entreprise". Dans ce contexte, ils ont publié deux livrables accompagnés de tableaux de synthèses autour des scénarios cyber. Le premier concerne les scénarios fon-

damentaux d'attaques cyber et la cartographie des risques (Forum des compétences et EGERIE, 2022), et le deuxième traite les scénarios de risques importants sur la corruption de données (Forum des compétences, 2022). Il est important de noter que tous les scénarios listés dans les différents livrables ont diverses origines, certains sont par exemple, provoqués par des mauvaises manipulations tandis que d'autres ont des origines malveillantes. Ils ne sont donc pas tous pertinents pour chercher à déterminer les pertes qu'il pourraient engendrer aux assureurs à travers des couvertures silencieuses.

3.1.4.2 Le taux silencieux : la clef de voûte

L'obtention des fréquences d'utilisation de clause dépend de la façon dont la société construit ses contrats et des données dont elle dispose. S'il existait une base de donnée contenant, pour chaque contrat, les informations sur l'utilisation des clauses, elle pourrait alors être exploitée pour déterminer de façon exacte quels contrats sont exposés au silent cyber. Si ce n'est pas le cas, mais que nous disposons cependant d'une version numérique des contrats, nous pourrions alors utiliser des algorithmes de machine learning et quelques notions de NLP, pour identifier la présence ou non d'une clause par garantie et approximer plus fidèlement le taux silencieux. Malgré cela, l'avis d'expert devra être utilisé soit pour donner un retour sur la pertinence des résultats obtenus, soit pour donner directement les fréquences d'utilisation.

Comme nous le verrons lors de la mise en place du modèle épidémiologique, nous utiliserons le taux silencieux pour simuler l'activation de garanties.

Dans cette section nous venons de présenter les étapes essentielles pour l'évaluation des risques du cyber silencieux. Le cadre est très général et doit donc être adapté aux données disponibles chez les assureurs. De plus, de ces données découle également la façon dont taux silencieux peut être estimé.

3.1.5 Construction d'un portefeuille fictif

L'objectif principal de notre mémoire étant de voir comment est ce que le risque d'accumulation cyber peut impacter un portefeuille non cyber, nous cherchons à appliquer le modèle de Markov Continu (voir section 2.1.4.1) à un portefeuille d'assurance. Ne disposant pas de ce genre de données en interne, nous avons du construire notre propre portefeuille pour appliquer le modèle. Ainsi dans cette partie nous présentons les étapes que nous avons suivies pour construire un portefeuille fictif suffisamment réaliste. Nous appliquons le cadre d'évaluation de l'exposition silent, défini dans la section (3.1.3), directement lors de la construction de la base de données afin que dans la section (3.2) nous puissions nous concentrer sur la modélisation du risque d'accumulation. Ainsi nous pourrons construire des scénarios que nous appliquerons dans la section (3.2.2) afin d'analyser comment se diffuse le virus et les pertes qu'il génère.

3.1.5.1 Motivations et éléments de construction

Pour pouvoir construire notre modèle nous devons disposer d'un portefeuille qui soit représentatif d'une certaine réalité des données disponibles dans les sociétés d'assurances. Ne disposant pas de ce genre de données, nous avons fait le choix de construire notre base de données. Nous nous sommes inspirés de bases de données réelles anonymisées que nous avons pu consulter ainsi que de plusieurs conditions générales qui sont disponibles en ligne.

Avant de pouvoir construire notre portefeuille nous avons créé une matrice synthétisant pour chaque produit, les garanties qui sont, ou peuvent être, incluses dans les contrats lors de la souscription. Sur la figure (3.8), les cellules vertes sont des garanties obligatoires, incluses directement dans le produit d'assurance, tandis que les cellules en jaune concernent les garanties optionnelles lors de la souscription du contrat. Cette matrice a été construite en consultant un grand nombre de conditions générales réellement commercialisées et disponibles en ligne.

Matrice des garanties possibles par produit		Locaux et leur contenu										
Produits	**Libellé**	Tempête	Grêle	Avalanche, poids de la neige, gel, inondation	Evènements climatiques	Bris des vitres, vitrines et enseignes	Choc de véhicules	Dommages aménagements extérieurs	Vol hors domicile	Vol tentative de vol et vendalisme	Perte du contenu des congélateurs et caves à vin	Bris de matériel	Bris de matériel et infections informatiques
MP_2022	ASSURANCE MULTIRISQUE PROFESSIONNELS				1	1	1	2		2		2	
MBTP_2022	ASSURANCE MULTIRISQUE DES PROFESSIONNELS DU BÂTIMENTET DES TRAVAUX PUBLICS				1	1	1	2		2		2	
MRH:Init_2012	ASSURANCE MULTIRISQUE TEMPO HABITATION : Initiale	1	1			1	1			1			
MRH:Clas_2012	ASSURANCE MULTIRISQUE TEMPO HABITATION : Classique	1	1			1	1			1			
MRH:Full_2012	ASSURANCE MULTIRISQUE TEMPO HABITATION : Intégrale	1	1	1		1	1		1	1	1		

FIGURE 3.8 : Extrait de la matrice des garanties par produits.

A partir de cette matrice, nous sommes en mesure de construire 17 types de contrats d'assurance qui incluent diverses garanties parmi les 69 que nous avons répertorié.

La description des produits dans la figure (3.8) s'est faite de la façon suivante : X_0 : x_1 : (...) : x_i_A. Le premier terme X_0 est toujours écrit en majuscules et fait référence au nom du produit qui commercialise la police. Certaines polices (au niveau X_0) proposent des extensions ou de nouvelles garanties qui se hiérarchisent en plusieurs niveaux. Ainsi, X_1, X_2, jusqu'à X_i font référence au niveau de détail du contrat. Finalement, A représente l'année de rédaction de la police.

Prenons l'exemple de *MRH :Full_2012*. Pour cette police le caractère *MRH* fait référence à l'assurance mulitirisque tempo habitation (nom réel du contrat inscrit sur les conditions générales) et octroie un certain nombres garanties à l'assuré. Le caractère *Full* va venir compléter les garanties de l'assuré, mais peut également permettre à l'assurer d'acquerir de nouvelles garanties qui ne lui auraient pas été accessibles s'il n'avait pas acquis le caractère *Full* (garanties en jaune, pas visibles dans l'extrait de la figure (3.8)). Finalement, l'année de création de ce contrat est de 2012 mais ne correspond pas à l'année de souscription. En effet nous, considérons que certains contrats n'ont pas été modifié du fait de la tacite reconduction, ce qui les emmène à être exposé au risque cyber silencieux.

La colonne Libellé sur la figure (3.8) donne les descriptions auxquelles font références les différents sigles X_i.

3.1.5.2 Présentation

En utilisant la matrice précédente, nous sommes en mesure de constituer des portefeuilles par produit, dans lesquels nous indiquons pour chaque garantie, quels sont les montants exposés. Par montants exposés, nous entendons la quantité maximale indemnisable, par exemple si une garantie contient une limite d'indemnisation, alors l'exposition correspondra à cette limite. Au cours de l'étude nous avons construit plus d'une dizaine de portefeuilles qui ne possèdent pas tous les mêmes caractéristiques afin de pouvoir étudier comment se diffusait le virus. Durant la partie application, nous en utiliserons

certains pour identifier quelques paramètres important sur la diffusions des virus. Sur la figure (3.9), nous avons quelques lignes pour le portefeuille Multirisques Professionnels.

Multirisques Professionnels											
No_Contrat	Grêle	Avalanche, poids de la neige, gel, inondation	Evènements climatiques	Bris des vitres, vitrines et enseignes	Choc de véhicules	Dommages amménagements extérieurs	Vol hors domicile	Vol tentative de vol et vendalisme	Perte du contenu des congélateurs et caves à vin	Bris de matériel	Bris de matériel et infections informatiques
MP_2019: 773958318	-	-	449 378,00	942 301,00	432 230,00	756 724,00	-	-	-	-	487 376,00
MP_2019: 4477225817	-	-	392 535,00	336 113,00	989 196,00	48 969,00	-	-	-	-	-
MP_2019: 5012578793	-	-	684 154,00	736 806,00	32 668,00	75 400,00	-	-	-	-	843 571,00
MP_2019: 3920573794	-	-	738 825,00	737 263,00	469 771,00	512 267,00	-	-	-	-	380 360,00
MP_2019: 7029031177	-	-	682 983,00	658 998,00	466 337,00	-	-	-	-	-	-
MP_2019: 3394212725	-	-	266 418,00	705 172,00	164 855,00	-	-	12 075,00	-	-	-
MP_2019: 6350782039	-	-	759 542,00	580 677,00	175 646,00	161 031,00	-	-	-	-	-
MP_2019: 5439713871	-	-	613 002,00	588 761,00	586 428,00	-	-	-	-	-	-
MP_2019: 4781945843	-	-	452 661,00	78 729,00	570 354,00	189 456,00	-	-	-	-	31 882,00
MP_2019: 5163890721	-	-	209 344,00	241 388,00	656 540,00	-	-	832 130,00	-	-	-

FIGURE 3.9 : Montants exposés en euros pour les premiers contrats dans le portefeuille Multirisque Professionnels.

Sur la première colonne de la figure (3.9), nous remarquons que le numéro de contrat (No_Contrat) reprend la structure de description du produit de la figure (3.8) mais rajoute un numéro d'identification pour l'assuré.

Dans la section 3.2.2, nous considérons un scénario pertes d'exploitation. Cette garantie communément incluse dans les contrats professionnels vise à indemniser le manque à gagner des entreprises du fait d'une interruption d'activité. A des fins de modélisation, nous considérons que la garantie perte d'exploitation ne contient pas de franchise.

3.1.5.3 Les données de souscription

Dans le Chapitre 2, nous avons construit des modèles épidémiologiques en exploitant les structures de réseaux pour modéliser l'environnement dans lequel un virus peut se propager, (voir section 2.3). De façon générale, les données de souscription jouent un rôle déterminant pour comprendre le comportement du profil des assurés. Nous allons considérer par la suite le secteur de l'assuré dans notre modélisation. Cette donnée est récupérée à la souscription et nous permet de définir une structure de réseau (voir section 2.2.3.1).

Multirisques Professionnels	Information de souscription										
No_Contrat	Secteur	Grêle	Avalanche, poids de la neige, gel, inondation	Evènements climatiques	Bris des vitres, vitrines et enseignes	Choc de véhicules	Dommages amménagements extérieurs	Vol hors domicile	Vol tentative de vol et vendalisme	Perte du contenu des congélateurs et caves à vin	Bris de matériel	Bris de matériel et infections informatiques
MP_2019: 773958318	Energy	-	-	449 378,00	942 301,00	432 230,00	756 724,00	-	-	-	-	487 376,00
MP_2019: 4477225817	Services	-	-	392 535,00	336 113,00	989 196,00	48 969,00	-	-	-	-	-
MP_2019: 5012578793	Mining	-	-	684 154,00	736 806,00	32 668,00	75 400,00	-	-	-	-	843 571,00
MP_2019: 3920573794	Services	-	-	738 825,00	737 263,00	469 771,00	512 267,00	-	-	-	-	380 360,00
MP_2019: 7029031177	Manufacturing	-	-	682 983,00	658 998,00	466 337,00	-	-	-	-	-	-
MP_2019: 3394212725	Energy	-	-	266 418,00	705 172,00	164 855,00	-	-	12 075,00	-	-	-
MP_2019: 6350782039	Services	-	-	759 542,00	580 677,00	175 646,00	161 031,00	-	-	-	-	-
MP_2019: 5439713871	Services	-	-	613 002,00	588 761,00	586 428,00	-	-	-	-	-	-
MP_2019: 4781945843	Manufacturing	-	-	452 661,00	78 729,00	570 354,00	189 456,00	-	-	-	-	31 882,00
MP_2019: 5163890721	Energy	-	-	209 344,00	241 388,00	656 540,00	-	-	832 130,00	-	-	-

FIGURE 3.10 : Montants exposés en euro des premiers contrats dans le portefeuille Multirisques Professionnels avec les données de souscription.

Les secteurs retenus pour décrire les assurés sont définis de manière à pouvoir exploiter la structure de réseau qui peut être définie à partir de la matrice présentée dans la section 2.2.2.2. Ainsi, nous considérons les secteurs *Mining*, *Services*, *Construction*, *Energy* et *Manufacturing*. Nous cherchons à

garder cette cohérence entre données souscription et le réseau considéré car chez les assureurs, nous pensons qu'elles peuvent jouer un rôle déterminant dans la conception des liens entre assurés.

Dans la première configuration de notre portefeuille, nous allons considérer une équi-distribution des assurés par secteur d'activité. De cette façon nous avons donc :

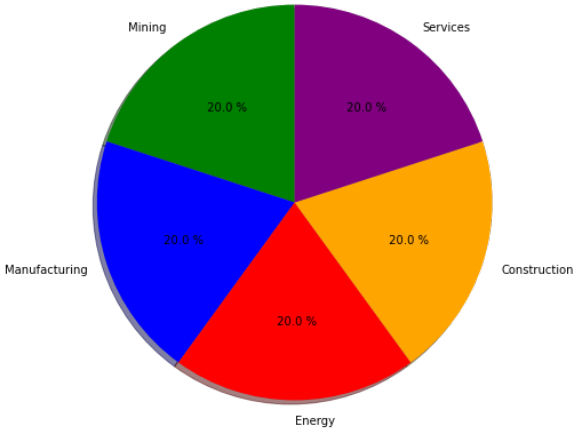


FIGURE 3.11 : Distribution des assurés par secteur d'activité pour le portefeuille Multirisques Professionnels.

Cette hypothèse va en effet nous permettre d'illustrer en quoi les données de souscriptions (le secteur des assurés) peuvent impacter la diffusion du virus (voir section 3.2.3.2).

3.1.5.4 Les taux silencieux

Finalement, comme nous l'avons souligné à plusieurs reprises, les risque cyber silencieux sont provoqués lorsqu'une garantie n'exclue ou n'affirme pas correctement le risque cyber (voir 1.3.2). Ainsi nous établissons pour chaque garantie un taux silencieux qui va représenter le pourcentage de garanties générant des expositions silencieuses. Dans DAFFRON et al., 2019, les auteurs présentent un scénario "Bashe Attack" et considèrent que 33% des garanties pertes d'exploitations seraient en mesure d'engendrer des coûts cyber silencieux. Encore une fois, l'estimation de ce taux mérite une évaluation propre à chaque assureur afin d'être le plus fidèle possible à la réalité de ses risques.

Multirisques Professionnels	Information de souscription											
No_Contrat	Secteur	Grêle	Avalanche, poids de la neige, gel, inondation	Evènements climatiques	Bris des vitres, vitrines et enseignes	Choc de véhicules	Dommages amménagements extérieurs	Vol hors domicile	Vol tentative de vol et vendalisme	Perte du contenu des congélateurs et caves à vin	Bris de matériel	Bris de matériel et infections informatiques
Taux Silent -->		0%	0%	0%	10%	13%	5%	2%	2%	40%	33%	12%
MP_2019: 773958318	Energy	-	-	449 378,00	942 301,00	432 230,00	756 724,00	-	-	-	-	487 376,00
MP_2019: 4477225817	Services	-	-	392 535,00	336 113,00	989 196,00	48 969,00	-	-	-	-	-
MP_2019: 5012578793	Mining	-	-	684 154,00	736 806,00	32 668,00	75 400,00	-	-	-	-	843 571,00
MP_2019: 3920573794	Services	-	-	738 825,00	737 263,00	469 771,00	512 267,00	-	-	-	-	380 360,00
MP_2019: 7029031177	Manufacturing	-	-	682 983,00	658 998,00	466 337,00	-	-	-	-	-	-
MP_2019: 3394212725	Energy	-	-	266 418,00	705 172,00	164 855,00	-	-	12 075,00	-	-	-
MP_2019: 6350782039	Services	-	-	759 542,00	580 677,00	175 646,00	161 031,00	-	-	-	-	-
MP_2019: 5439713871	Services	-	-	613 002,00	588 761,00	586 428,00	-	-	-	-	-	-
MP_2019: 4781945843	Manufacturing	-	-	452 661,00	78 729,00	570 354,00	189 456,00	-	-	-	-	31 882,00
MP_2019: 5163890721	Energy	-	-	209 344,00	241 388,00	656 540,00	-	-	832 130,00	-	-	-

FIGURE 3.12 : Montants exposés pour les premiers contrats dans le portefeuille Multirisques Professionnels avec les données de souscription et les taux silencieux.

Sur la figure (3.12) nous présentons quelques ligne représentatives des portefeuilles que nous avons pu étudier.

Dans cette section nous avons présenté les éléments qui nous ont permis de construire un portefeuille fictif. Pour chaque assuré, nous disposons des montants exposés et des taux silencieux par garantie. En guise de donnée de souscription, nous considérons pour chaque assuré, son secteur d'activité. Ces éléments seront exploités et détaillés dans la prochaine section sur la modélisation du risque d'accumulation.

3.2 Modélisation du risque d'accumulation

Dans cette section nous allons exploiter les différents portefeuilles que nous avons construits pour appliquer le modèle de Markov continu avec réseaux présenté dans le Chapitre 2 (voir 2.1.4.1 et 2.3). Au cours de la diffusion du virus, nous nous intéresserons à deux quantités, le nombre d'infectés au cours du temps et les coûts générés par le virus chez les assurés. Nous cherchons à montrer comment évoluent ces résultats en fonction des caractéristiques du modèle. Cependant, le réseau pour modéliser les interactions entre assurés sera invariant et se fera à l'aide d'une matrice d'adjacence (voir 2.1.3.4). Comme nous l'avons vu lors de la *Construction* des portefeuilles (voir 3.1.5), nous disposons pour chaque assuré des montants exposés par garantie. Nous allons donc considérer que chaque assuré est un individu dont les états sont ceux du modèle SIR.

3.2.1 Modélisation épidémiologique par scénarios

Dans cette partie nous allons présenter les différentes étapes de la modélisation que nous avons effectuée pour estimer les pertes causées par le risque d'accumulation cyber sur des polices non cyber. Cette section est primordiale puisqu'elle fait le lien entre le modèle de Markov continu du Chapitre 2 (voir 2.1.4.1) et la gestion du silent cyber présenté en début de Chapitre 3 (voir 3.1) et introduit dans le Chapitre 1 (voir 1.3.2).

3.2.1.1 Caractérisation d'un scénario

L'un des principaux défis des modèles de réseaux est le calibrage des paramètres. Une méthode rigoureuse de calibrage impliquerait très certainement l'accès à des données techniques de cybersécurité, de réseaux informatiques et d'informations de souscription pour connaître précisément l'environnement du risque et tenter de reproduire des évènements d'accumulation comme Wannacry (voir 1.3.2). C'est la raison pour laquelle nous adoptons une méthode par scénario. Cette méthode nous indique les leviers à maîtriser afin de mieux gérer ce risque.

Nous caractérisons un scénario par un virus informatique qui possède un certain taux de contagion β qui lui est propre, une structure de réseau qui modélise l'environnement dans lequel ce virus va évoluer (force du lien d'interaction entre assurés), et finalement les garanties couvertes par une police d'assurance (non-vie) que le virus va déclencher lorsqu'il infecte un nœud.

3.2.1.2 Modéliser l'activation des garanties d'un assuré

Le modèle considère les assurés comme les individus d'un modèle épidémiologique (Markov continu avec structure de réseau). Ainsi, nous allons considérer qu'un assuré est caractérisé par son secteur d'activité (comme pour la modélisation faite dans 2.2.2.2), par les garanties qui le couvrent (et qui sont présentes dans le scénario d'étude), ainsi que par les montants des expositions associés et le portefeuille auquel il appartient.

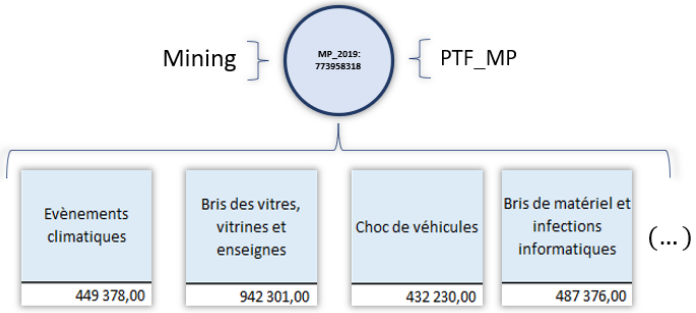


FIGURE 3.13 : Informations contenues dans un nœud représentant l'assuré.

L'application étant sur le cyber silencieux, le modèle doit pouvoir différencier quels individus infectés sont en mesure d'activer une garantie au sein de leur contrat ou non. Le modèle retenu va donc être très semblable au modèle présenté dans la figure (2.15). Cependant ce modèle discrimine les individus selon un taux T_s (voir 2.1.4.2). Or, comme nous l'avons vu dans le Chapitre 1 (1.3.2), les expositions au cyber silencieux prennent source au sein des garanties et non au sein des individus en soit comme le fait la modélisation présenté sur la figure (2.15) . Sur la figure (3.14) nous détaillons le processus d'activation de garanties. Notons que le processus de diffusion SIR est indépendant du processus d'activation des garanties silencieuses.

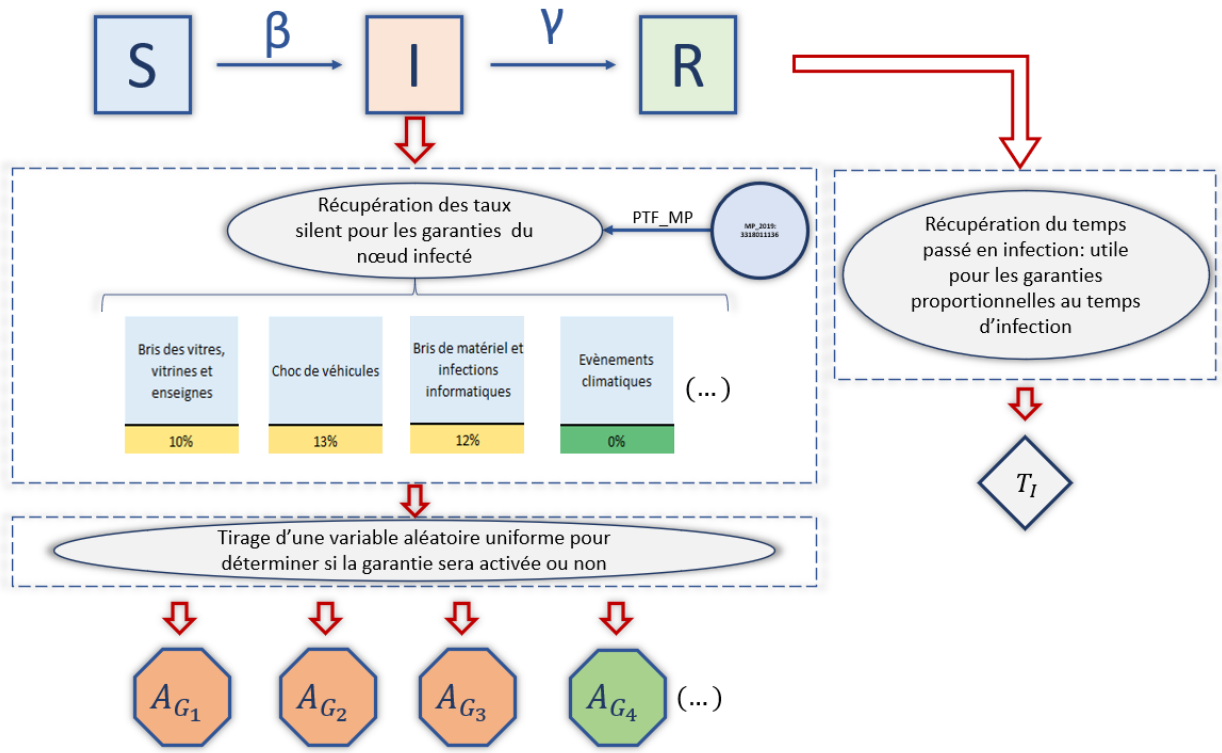


FIGURE 3.14 : Processus d'activation des garanties sur le modèle SIR.

Comme nous pouvons le voir également sur la figure (3.14), nous récupérons le temps de passé en infection par un nœud. Ce temps correspond à la différence entre la date de rétablissement et la

date de début d'infection. Il est utile pour calculer les pertes engendrées par une garantie dont le montant d'indemnisation est proportionnel au temps comme ça peut être le cas pour la garantie perte d'exploitation.

Rappelons que dans Daffron et al., 2019, un taux de 33% avait été retenu pour déterminer les pertes associées au cyber silencieux.

3.2.1.3 Évaluation des pertes par garanties

Une fois que les garanties activées sont déterminées, un deuxième processus entre en jeu pour évaluer les pertes probables. Nous définissons pour chaque garantie, une loi de probabilité pour simuler une demande d'indemnisation de la part de l'assureur, dans notre application nous utiliserons des lois Gamma.

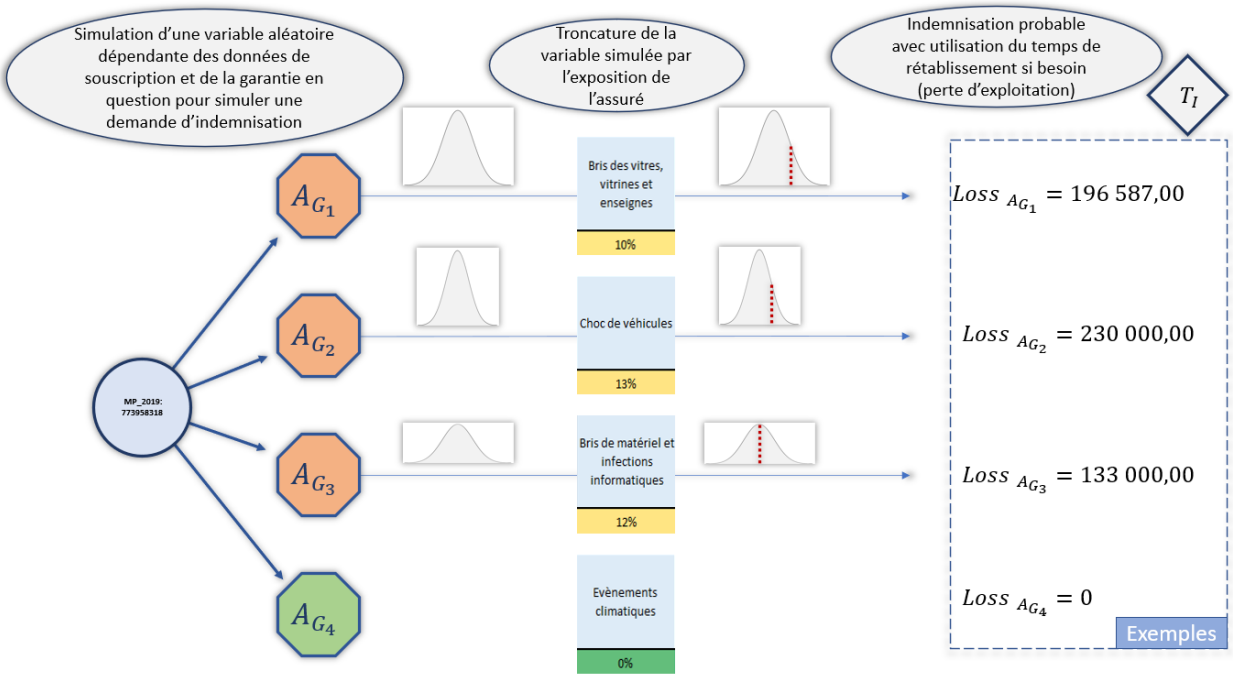


Figure 3.15 : Processus de génération d'une perte pour les garanties d'un nœud infecté.

Au sein d'une compagnie d'assurance, nous pourrions, par exemple, utiliser des données historiques pour calibrer une loi qui corresponde au mieux à la sinistralité observée par garantie.

Soulignons qu'une attention particulière est nécessaire lors de la modélisation des pertes. En effet, la façon dont sont modélisées ces pertes doit être cohérente avec la définition que nous avons pris pour l'exposition (par exemple, présence d'une franchise ou non).

3.2.1.4 La structure de réseau

Nous avons présenté dans le Chapitre 2 (voir 2.1) plusieurs modèles épidémiologiques, dont certains exploitent une structure de réseau afin de prendre en compte l'hétérogénéité des contacts entre les individus. Ainsi, pour mieux représenter la réalité des contacts qui existent entre assurés, nous allons

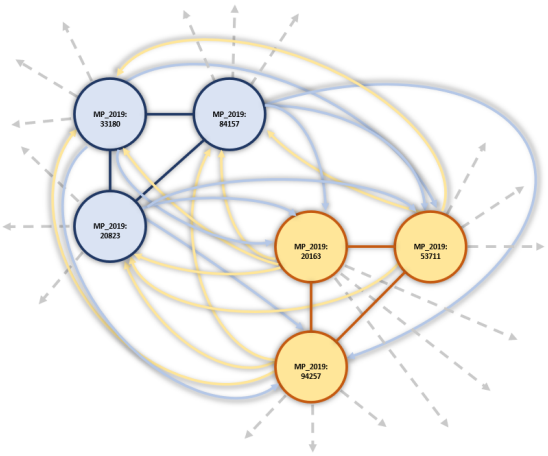
considérer une structure de réseau dépendante des données de souscription. Comme nous disposons uniquement du secteur, nous allons donc nous baser sur ce critère là pour construire le réseau.

Dans le Chapitre 2 (voir 2.2.2.2), nous avons présenté une structure de réseau proposée dans Hillairet et al., 2021, qui nous le rappelons, dépend du secteur des assurés. Nous allons donc reprendre la même structure mais nous effectuons une normalisation selon un secteur de référence. Cette normalisation permet de propager le virus à un nœud susceptible de ce secteur avec un taux d'infection $\beta \times$ le nombre de voisins infectés appartenant au même secteur.

Sectors	Mining	Manufacturing	Energy	Construction	Services
Mining	1	4,61672	0,7082	2,25079	1,9795
Manufacturing	0,0994	0,83123	0,04259	0,17035	0,55363
Energy	0,21293	0,58359	0,90063	0,23659	0,71293
Construction	0,02997	0,10726	0,01104	0,22239	0,14353
Services	0,00473	0,06624	0,00631	0,02681	0,25394

Figure 3.16 : Matrice pour les poids d'adjacence du réseau (d'après les données de Hillairet et al., 2021).

Notre structure de réseau se compose de plusieurs nœuds (porteurs d'informations, voir figure (3.13)), tous reliés les uns aux autres par des arcs pondérés. Sur la figure (3.17a), les arcs ont un poids différent selon leur couleur (caractérise le secteur), ce qui se traduit par des valeurs différentes de 1 dans la matrice d'adjacence du réseau. Par exemple, si nous considérons que les nœuds jaunes appartiennent au secteur *Mining* et les nœuds bleus au secteur *Manufacturing*, la matrice d'adjacence serait celle de la Firgure (3.17b).



(a) Exemple d'une portion de potentiel réseau dont les arcs ont des poids différents selon la couleur.

$$\begin{bmatrix} 0 & 1 & 1 & 4{,}62 & 4{,}62 & 4{,}62 \\ 1 & 0 & 1 & 4{,}62 & 4{,}62 & 4{,}62 \\ 1 & 1 & 0 & 4{,}62 & 4{,}62 & 4{,}62 \\ 0{,}1 & 0{,}1 & 0{,}1 & 0 & 0{,}83 & 0{,}83 \\ 0{,}1 & 0{,}1 & 0{,}1 & 0{,}83 & 0 & 0{,}83 \\ 0{,}1 & 0{,}1 & 0{,}1 & 0{,}83 & 0{,}83 & 0 \end{bmatrix}$$

(b) Matrice d'adjacence résultant de la structure de réseau (a).

Figure 3.17 : Exemple d'un potentiel réseau dont les arcs ont des poids différents selon la couleur.

Notons qu'au sein d'un même secteur, les liens entre les nœuds sont non orientés. Ce qui est représenté par une seule arête non fléchée entre deux nœuds de même couleur. De plus, la matrice d'adjacence a les coefficients diagonaux nuls puisqu'aucun nœud n'a de flux sur lui même.

Dans cette section nous avons présenté la modélisation que nous faisons pour appliquer le modèle de pandémie sur un réseau représentatif des liens potentiels entre assurés. Une forte hypothèse est faite sur ces liens (données de l'OCDE) que nous justifions dans la section (2.2.2.2). De plus, nous avons illustré comment est ce que nous modélisons les pertes provenant du cyber silencieux. Soulignons lors de cette dernière étape, que la simulation des pertes doit être cohérente avec la définition prise pour l'exposition. Tout au long de cette modélisation, nous avons tenté de rester le plus général possible afin qu'elle puisse s'adapter au mieux à d'autres cas, plus concret, d'une société d'assurance réelle.

3.2.2 Application au scénario perte d'exploitation

A présent nous définissons un scénario simplifié afin d'illustrer la modélisation précédente et d'illustrer quelques résultats. Nous considérons un virus avec un taux de propagation β de 0.01 qui paralyse le système informatique d'une entreprise quelconque, activant ainsi uniquement la garantie perte d'exploitation. Nous commençons par considérer un rétablissement du système informatique moyen après un jour d'infection, ce qui conduit à fixer le paramètre de rétablissement γ, définit dans la section 2.1.4.1, à : $\gamma = 1$.

3.2.2.1 Le portefeuille considéré

Le scénario étant déjà spécifié, nous soumettons à cet environnement un portefeuille fictif constitué de 1000 polices multirisques professionnels. Chaque assuré est caractérisé par son secteur d'activité dont nous donnons les proportions au sein du portefeuille dans la figure (3.11). De plus, comme il s'agit d'un portefeuille fictif, nous considérons que les montants journalier des pertes chez les assurés suivent une loi Gamma notée $G \sim \mathcal{G}(a, b)$ dont les paramètres a et b dépendent des informations à la souscription (c'est à dire le secteur d'activité dans notre cas). Rappelons que pour une variable aléatoire $G \sim \mathcal{G}(a, b)$, le domaine de définissions de la densité est strictement positive, elle prend donc des valeurs strictement positives (essentiel pour modéliser une perte). De plus, nous avons $\mathbb{E}[G] = a \times b$ et $\mathbb{V}ar[G] = a \times b^2$. Ainsi nous sommes en mesure de simuler une perte journalière pour un assuré en perte d'exploitation.

Table 3.4 : Paramètres pour les lois générant les pertes par secteurs.

Secteur	a	b	Espérance	Variance
Mining	200 000,00	0,5	100 000,00	50 000,00
Manufacturing	10 000,00	0,5	5 000,00	2 500,00
Energy	40 000,00	0,5	20 000,00	10 000,00
Construction	20 000,00	0,5	10 000,00	5 000,00
Services	20 000,00	0,5	10 000,00	5 000,00

Nous pouvons observer sur la Table (3.4) que les pertes journalières du secteur *Mining* sont (en espérance) les plus importantes, tandis que celles du secteur *Manufacturing* sont les plus faibles. De plus, l'espérance des pertes pour les secteurs *Manufacturing*, *Energy*, *Construction* et *Services* sont relativement proches les unes des autres vis à vis du coût (toujours en espérance) journalier du secteur *Mining*. Ainsi nous pouvons nous attendre à d'importants frais d'indemnisation pour ce secteur.

Comme nous l'avons montré sur la figure (3.15), une fois le sinistre simulé, nous devons confronter

cette valeur au montant qui est contractuellement indemnisable par l'assureur. Ainsi sur la figure (3.18), nous considérons que les montants exposés correspondent au montant maximal indemnisable par l'assureur, ce qui peut par exemple provenir d'une limite d'indemnisation. Dans notre modélisation, nous considérons que cette valeur dépend uniquement du secteur d'activité de l'assuré.

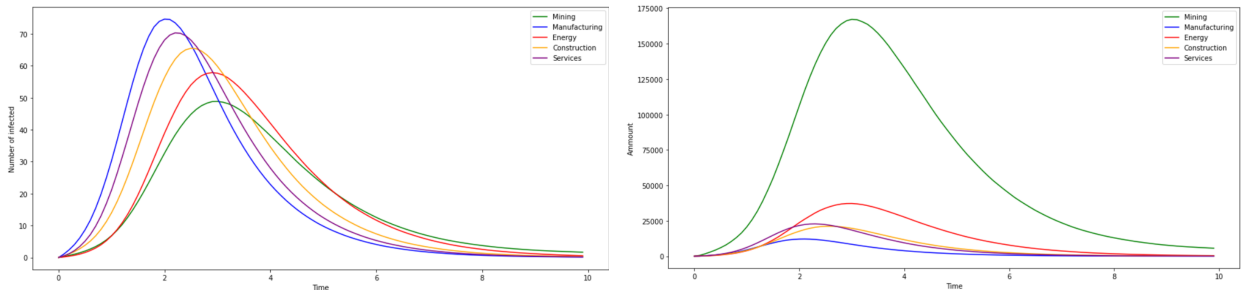
Multirisque Professionnels	Données de souscription	
No_Contrat	**Secteur**	Perte d'exploitation
Silent Pourcentage -->		0,32
MP_2019: 3318011136	Mining	150 000,00
MP_2019: 3318011136	Mining	150 000,00
MP_2019: 4903087815	Manufacturing	7 500,00
MP_2019: 4593390988	Manufacturing	7 500,00
MP_2019: 6372174745	Energy	30 000,00
MP_2019: 3315118185	Energy	30 000,00
MP_2022: 3986174666	Construction	15 000,00
MP_2022: 82552042	Construction	15 000,00
MP_2022: 4493508818	Services	15 000,00
MP_2022: 2013765611	Services	15 000,00

FIGURE 3.18 : Extrait du portefeuille pour le scénario perte d'exploitation.

Notons que sur la figure (3.18), les expositions correspondent à l'indemnisation maximale journalière. De plus, notre portefeuille est constitué de 1000 assurés équi-distribués dans les différents secteurs d'activités, voir figure (3.11).

3.2.2.2 Évolution du nombre d'infectés et des pertes engendrées

Nous effectuons 10 000 simulations en initialisant l'infection aléatoirement parmi les assurés à chaque nouvelle simulation. Ainsi, pour chaque nouvelle simulation, un individus parmi les 1000 présents dans le portefeuille est choisi pour que sont état soit "Infecté" dès le début. Sur la figure (3.19b), la perte instantanée correspond au montant à indemniser par l'assureur à un instant précis.



(a) Nombre d'infectés par secteur au cours du temps. (b) Perte instantanée par secteur au cours du temps.

FIGURE 3.19 : Évolution du nombre d'infectés et de la perte instantanée par secteur au cours du temps.

Sur la figure (3.19a), nous remarquons que le nombre d'infectés évolue différemment selon le secteur que nous considérons. Certains secteurs, comme *Manufacturing*, ont leur pic d'infection très rapidement tandis que d'autres, comme *Mining*, l'ont beaucoup plus tard. Ceci est directement dû à la structure de réseau et aux poids placés sur les arcs. En effet, si nous avions laissé une structure homogène et sans poids, nous aurions eu cinq courbes superposées les unes aux autres. Rappelons que la structure du réseaux est décrite dans la section 3.2.1.4.

Sur la figure (3.19b) nous pouvons observer comment les assurés du secteur *Mining*, moins infectés que les autres secteurs, sont ceux qui pèsent le plus dans la solvabilité de l'assureur. De plus nous remarquons qu'un secteur en particulier semble se détacher des autres, avec un pic de coûts à plus de 160 000 euros. En effet, lorsque 10 assurés du secteur *Mining* (environ à $t = 1.5$) sont infectés, les pertes instantanées sont plus élevées que le maximum des pertes du secteur *Energy* avec 50 infectés.

Nous observons plus généralement que les pics des pertes pour les secteurs *Manufacturing*, *Energy*, *Construction*, et *Services*, sont respectivement autour de 12 500€, 37 000€, 20 000€et 25 000€.

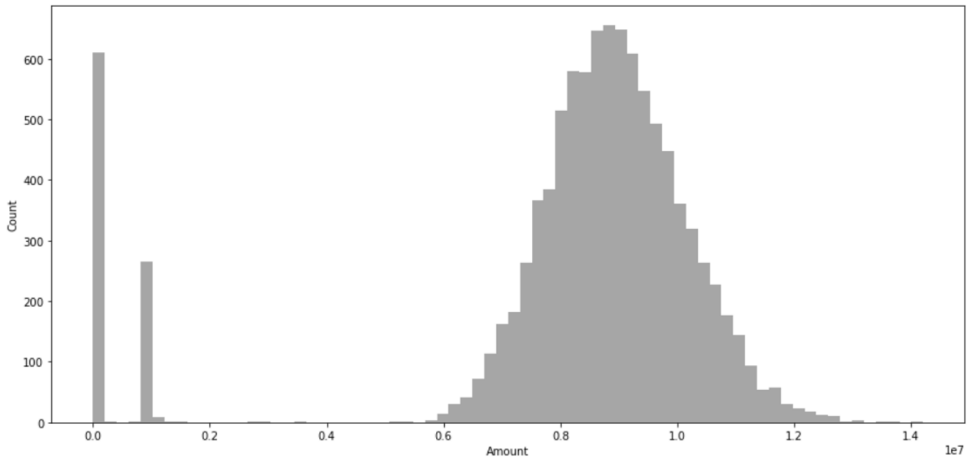


FIGURE 3.20 : Densité des pertes cumulées 10 jours après le début de l'infection.

Ainsi pour ce scénario, la perte moyenne s'élève à **8 204 785€**. Tout de même, une quantité non négligeable de trajectoires a généré des pertes cumulées relativement faibles, voire nulles. Ceci provient essentiellement du fait, que tous les départ d'infection à $t = 0$ n'entraînent pas systématiquement un effet pandémique et peuvent s'éteindre très rapidement. Ce qui fait que lorsque le virus se comporte comme une épidémie, nous aurions plutôt tendance à estimer une perte moyenne autour de 9M€. Remarquons que l'exposition sur l'ensemble du portefeuille est de 43,5M€.

3.2.3 Effets des mesures de réaction

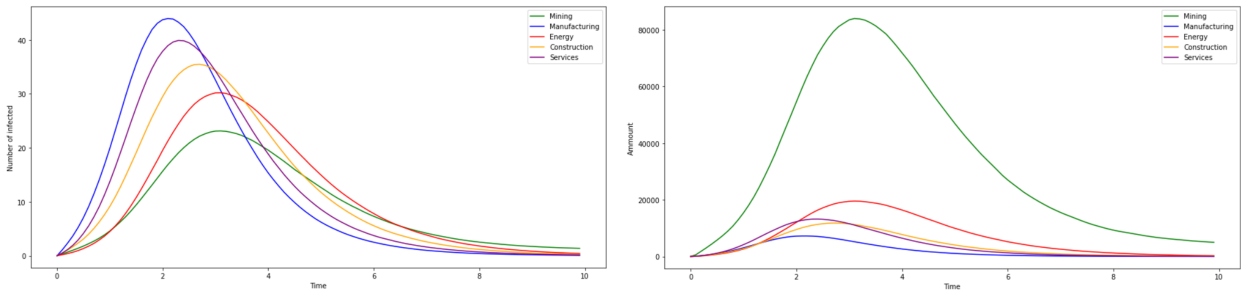
A présent nous étudierons les impacts que peuvent avoir deux mesure de réaction sur la diffusion et les pertes causées par le virus. La première va consister à augmenter le taux de rétablissement des individus infectés, la deuxième se base sur la modification de la composition du portefeuille.

3.2.3.1 Augmentation du taux de rétablissement

La mesure la plus directe à prendre pour atténuer de façon générale le risque d'accumulation est d'augmenter le taux de rétablissement des assurés infectés. Ceci peut passer par une veille de l'état des systèmes informatiques des assurés, facilitant ainsi une intervention de remise en état adapté, rapide et ciblée. En effet, en incitant les assurés à mettre à jour leurs logiciels et systèmes d'exploitation, les assureurs pourraient même dans certains cas, atténuer le risque d'infection sans pour autant l'éliminer.

Rappelons que sur la figure (1.4), nous pouvons remarquer qu'un an après la sortie d'un patch contre la vulnérabilité Eternal Blue, faille entre autres exploitée par Wannacry, beaucoup d'ordinateurs restaient vulnérables partout dans le monde.

De cette façon, nous allons simuler la propagation du virus dans le même environnement que précédemment en augmentant la capacité d'intervention de l'assureur. Ainsi, les assurés infectés rétablissent leurs activités non plus au bout d'un jour mais au bout d'une grosse demi journée ($0,67$ jours), $\gamma = 1,5$. En effet, rappelons que $\gamma = 1/\text{Temps moyen d'infection}$.



(a) Nombre d'infectés par secteur au cours du temps. (b) Perte instantanée par secteur au cours du temps.

FIGURE 3.21 : Évolution du nombre d'infectés et de la perte instantanée par secteur au cours du temps avec intervention de l'assureur.

Sur la figure (3.21a) nous pouvons voir comment les pics d'infection pour chaque secteur est deux fois moins important. Cette baisse sur les pics d'infection permet de réduire le nombre de garanties activées par le scénario et donc de réduire le montant à indemniser. De plus, comme le montant d'indemnisation de la garantie perte d'exploitation est proportionnel au temps, les pertes des individus infectés diminuent directement par la simple augmentation du paramètre de rétablissement γ. Ainsi, en augmentant le paramètre γ, un double effet bénéfique se produit pour l'assureur :

(i) les probabilités qu'un nœud susceptible se retrouve infecté (puisque ses nœuds voisins en étant moins de temps infecté, on donc également moins de temps pour l'infecter) diminuent.

(i) les pertes liées à la perte d'exploitation, qui est proportionnelle au temps d'arrêt d'activité de l'entreprise assurée, diminuent.

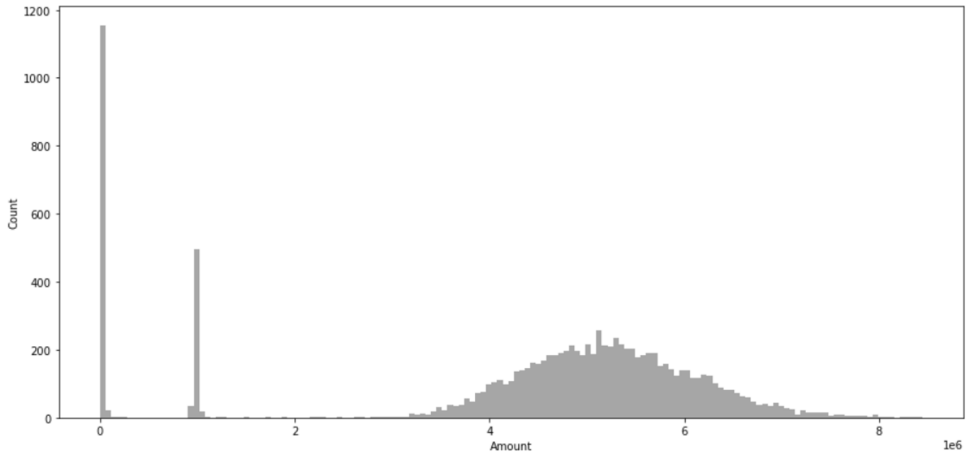


FIGURE 3.22 : Densité des pertes cumulées après 10 jours du début de l'infection avec intervention de l'assureur.

Nous remarquons sur la figure (3.22), que le nombre de trajectoires dont les pertes sont proches de 0 augmentent. Ceci est la conséquence directe du temps d'infection plus faible. De plus, lorsque les garanties de pertes d'exploitation sont activées, les montants d'indemnisation sont plus faibles. En diminuant ainsi le temps nécessaire au rétablissement des assurés infectés, nous diminuons les pertes moyennes de moitié, celles-ci passent de 8 204 785€à **4 400 217€**.

3.2.3.2 Modifier la composition du portefeuille

La question que nous nous posons à présent est quel serait l'impact d'une modification du portefeuille sur les pertes totales ? Nous avons vu sur la figure (3.16) que le secteur qui transmet le plus le virus aux autres est le secteur *Mining*. Mais ce secteur est également le plus coûteux, voir Tables (3.4).

Nous allons donc apporter quelques modification à notre portefeuille en réduisant le nombre d'assurés appartenant à ce secteur. Ainsi, nous conservons un portefeuilles de 1000 assurés dont les secteurs sont répartis ainsi :

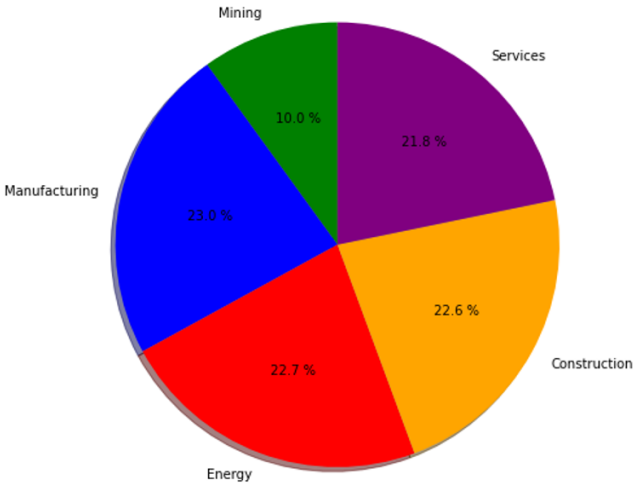
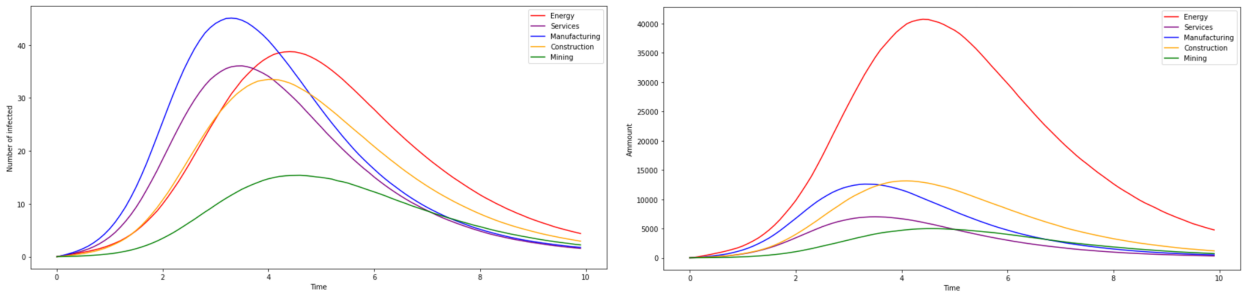


FIGURE 3.23 : Répartition des assurés selon les différents secteurs au sein du nouveau portefeuille.

Sur ce nouveau portefeuille restructuré, nous faisons propager le même virus qu'en début de section (3.2.2), c'est a dire de paramètre d'infection $\beta = 0.01$ et de taux rétablissement $\gamma = 1$.



(a) Nombre d'infectés par secteur au cours du temps. (b) Perte instantanée par secteur au cours du temps.

FIGURE 3.24 : Évolution du nombre d'infectés et de la perte instantanée par secteur au cours du temps sur un portefeuille restructuré.

Ainsi nous pouvons voir sur la figure (3.24a), que les pics d'infections sont réduits de moitié, et ce, tous secteurs confondus. Cependant, remarquons que les pics arrivent dans un ordre différent et à des temps différents que ceux exposées sur la figure (3.19a). En effet, sur les figures (3.19a,3.21a) nous

observons le premier pic d'infection de *Manufacturing* arriver autour du deuxième jour tandis qu'avec cette nouvelle structure de portefeuille nous avons les premiers pics (*Manufacturing* et *Services*) arriver presque simultanément autour du 3ème jour. Avec ces observation nous constatons de nouveau l'impact que peut avoir la structure du réseau sur la propagation du virus.

De plus, comme nous avons considérablement diminué la proportion d'assurés appartenant au secteur *Mining* (le secteur le plus cher à indemniser quotidiennement), nous observons sur la figure (3.24b) que le nouveau secteur ayant la perte instantanée la plus important est le secteur *Energy*.

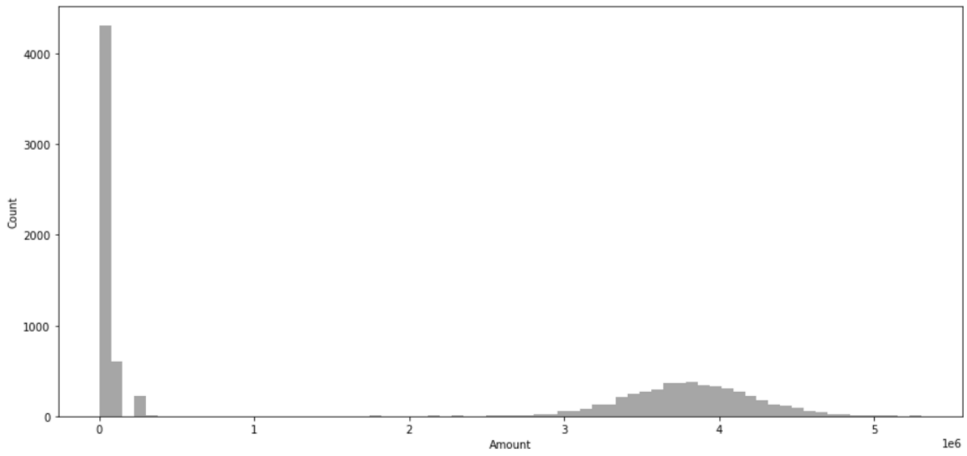


FIGURE 3.25 : Densité des pertes cumulées après 10 jours du début de l'infection sur un portefeuille restructuré.

Ainsi en restructurant le portefeuille, nous diminuons la perte moyenne cumulée qui passe de 8 204 785€à **1 872 643€**. Cependant un double effet est à l'origine de cette diminution aussi considérable : en effet le secteur *Mining* est un important vecteur de transmission pour les autres assurés mais il est également le plus cher à indemniser. En diminuant la part d'assurés dans ce secteur, nous jouons donc sur ces deux leviers pour favoriser une faible perte moyenne pour l'assureur.

Simulation	**Description**	**Perte moyenne**	**Commentaire**
Référence	Equi-répartition des 1000 assurés dans les 5 secteurs, figure (3.11). Taux de contagion $\beta = 0.01$. Taux de rétablissement $\gamma = 1$.	8 204 785€	Importante distribution des pertes concentrées autour des 9M€, figure(3.20).
Taux de rétablissement	Equi-répartition des 1000 assurés dans les 5 secteurs, figure (3.11). Taux de contagion $\beta = 0.01$. Taux de rétablissement $\gamma = 1.5$.	4 400 217€	Faible distribution autour de 5M€, figure(3.22). Nombre important de pertes autour de 0.
Restructuration	Distribution des 1000 assurés dans les 5 secteurs, figure (3.23). Taux de contagion $\beta = 0.01$. Taux de rétablissement $\gamma = 1$.	1 872 643 €	Très faible distribution autour de 3 M€, figure(3.25). Nombre très important de pertes autour de 0. Double effet généré par la restructuration.

3.3 Pour aller plus loin

Dans cette section nous présentons dans un premier temps les différents points qui permettent d'accroître la prise en compte du risque d'accumulation en lien avec la modélisation que nous avons développée. Puis dans un second temps nous présenterons brièvement le rapport de la Direction Générale du Trésor.

3.3.1 Autour du modèle

Le scénario perte d'exploitation que nous avons utilisé dans la section (3.2) modélise les interactions entre assurés par un réseau dépendant des secteurs auxquels ils appartiennent.

3.3.1.1 L'importance du réseau

La matrice utilisée pour construire le réseau est inspirée de celle introduite dans Hillairet et al., 2021 et que nous avons présenté dans la section (2.2.2.2). Cependant, nous avons illustré dans la section (2.3.2) l'importance du graphe dans la diffusion du virus. Ainsi, les données de souscription présentes chez les assureurs pourraient permettre de construire des structures de réseaux bien plus précises et adaptées.

Une première donnée sur les assurés qui pourrait être utilisée est le lien professionnel qui existe entre eux. En effet, certaines entreprises sous-traitent une partie de leur activité ou, collaborent avec d'autres entreprises sur certains projets. Du fait des disparités en matière de cyberprotection entre les entreprises, ces informations permettraient de mieux cerner l'environnement dans lequel se propage un virus informatique.

Des outils mathématiques permettent également de construire des réseaux en se basant sur des données. Dans Ver Steeg et Galstyan, 2012, les auteurs utilisent l'entropie de transfert pour construire à partir de données un réseau orienté. L'entropie de transfert est une statistique non-paramétrique qui permet de mesurer le transfert d'informations entre deux processus aléatoires. Ainsi, si nous disposons des temps de survenance d'incidents cyber pour un grand nombre d'assuré, nous pouvons déduire une structure de réseau.

3.3.1.2 Ne pas se limiter à un seul portefeuille

Dans notre scénario perte d'exploitation, section (3.2.2), nous avons considéré un seul portefeuille de police multirisques professionnels. Cependant, nous avons montré tout au long du chapitre 1, que le risque cyber n'est pas limité en espace : un même virus peut se propager tout autour du globe sans distinguer les individus qu'il infecte. Ainsi, nous pouvons envisager de ne pas se cantonner à un seul portefeuille mais de considérer un plus grand nombre d'assurés.

Considérer un plus grand nombre d'assurés permettrait de modéliser plus fidèlement l'environnement dans lequel se propage un virus informatique. Cependant, accroître le nombre d'individus dans la modélisation impliquerait l'utilisation d'un réseau également plus grand et plus complexe, ce qui entraînerait des temps de calcul bien plus élevés.

Pour pallier ce problème, nous pouvons modifier la conception du réseau. En effet, lors de l'application du scénario perte d'exploitation, section (3.2.2), nous avons construit un réseau où chaque

noeud représente un individu (assuré). Mais sur un grand nombre d'assurés, nous pouvons envisager de les regrouper par catégorie de risques. Nous ne développerons pas plus cette piste de modélisation puisqu'elle est très dépendante des données disponibles chez les assureurs mais reste tout de même envisageable en pratique.

Au-delà de considérer un plus grand nombre d'assurés, nous pouvons introduire dans le réseau, des noeuds externes qui jouent un rôle important dans le risque cyber. Par exemple, nous pouvons considérer un centre de stockage de données duquel dépendent certains assurés. Ainsi, une attaque par rançongiciel sur ce centre aurait un impact économique direct sur les assurés qui en dépendent.

La modélisation par réseau permet certes de diffuser un virus informatique, mais permet plus généralement de modéliser des liens d'influence entre les assurés. Ainsi, d'autres applications que le risque cyber pourraient être faites en utilisant la modélisation par réseau. Par exemple, nous pourrions modéliser comment une entreprise en perte d'exploitation entraîne des difficultés dans la continuité d'activité d'autres entreprises.

3.3.1.3 D'autres développements

Finalement, d'autres modèles compartimentaux peuvent être utilisés. Le $SEIR$ par exemple permet de modéliser l'état d'exposition d'un individu. Dans ce modèle, avant qu'un individu se retrouve infecté il devra être exposé au virus. Ceci permet par exemple de modéliser des mesures qui diminuent l'exposition des individus a un virus. Par exemple, lors d'un confinement, les individus susceptibles ne sont pas exposés aux virus de la même façon que lorsque les interactions sociales sont maintenues.

Cependant, appliquée aux réseaux, cette modélisation du degré d'interaction peut passer par l'utilisation de réseaux dynamiques. Par exemple, en début d'épidémie tous les individus sont reliés à 100 voisins mais à un certain moment lors de la diffusion du virus, la mise en place de certaines mesures réduit ce nombre de liens à 10 par individus.

En complément, si l'étude des interactions entre assurés est suffisamment fiable. Nous pouvons étudier quels assurés représentent un enjeu important lors de la diffusion d'un virus sur le réseau induit par les interactions entre assurés. En effet, comme nous l'avons présenté dans la section (2.2.3), certains auteurs s'intéressent aux probabilités d'infection des nœuds lorsqu'un virus se propage au sein du réseau. Ainsi, si les assureurs disposent d'une représentation fiable des interactions qui existent entre assurés (mais peut également prendre en compte d'autres facteurs externes : data centres, autres entreprises non assurées...) nous pouvons nous intéresser aux nœuds les plus contributifs à la diffusion du virus.

3.3.1.4 Des pistes de calibration

Dans ce mémoire nous n'avons pas développé le problème de calibration des paramètres pour le modèle épidémiologique. Cependant, une méthode de calibration des modèles épidémiologiques appliqués au risque cyber a été proposé dans un mémoire de l'institut des actuaires (RIGAUD, 2022). Dans ce mémoire, l'auteur applique une méthode de calibration bayesienne, développée dans Marin et al., 2012, au risque d'accumulation cyber. Cette calibration vise à trouver la loi des paramètres du modèle permettant de reproduire des évènements d'accumulation cyber comme NotPetya. De prochains travaux pourraient s'intéresser à l'adaptabilité de ces méthodes au cas des modèles de réseaux.

La modélisation par réseau est riche en applications et ne se cantonne pas au domaine du risque cyber. Cependant, réussir à représenter de manière fiable les interactions, ou liens, qui existent entre assurés demande un travail propre à chaque assureur afin de maintenir la cohérence entre la modélisation et les données disponibles.

3.3.2 Le développement de l'assurance du risque cyber

De nos jours, l'assurance cyber continue de se développer et suscite même des réflexions au plus haut sommet de l'Etat. Développer le marché de l'assurance cyber passe notamment par des concertations nationales comme a pu l'être celle menée par la Direction Générale du Trésor en juillet 2021 (Direction générale du Trésor, 2021). Le 7 septembre 2022, un rapport sur le développement de l'assurance cyber, et fondé sur cette concertation, est remis.

3.3.2.1 Le rapport de la Direction Générale du Trésor

Lors de la concertation de 2022, plusieurs acteurs importants du secteur de l'assurance ont constitué le groupe de travail. Nous retrouvons l'Autorité de Controle Prudentiel et de Résolution (ACPR), des réassureurs comme la Caisse centrale de réassurance (CCR) et SCOR, des assureurs tels que Axa France et Generali mais également des institutions académiques comme Sorbonne Université et l'ENSAE.

A l'issu du travail réalisé, le groupe de travail a déduit un plan d'action qui se décline en quatre axes :

(i) clarifier le cadre juridique de l'assurance du risque cyber

(ii) favoriser une meilleure mesure du risque cyber

(iii) améliorer le partage de risque entre assurés, assureurs et réassureurs

(iv) accroître les efforts de sensibilisation des entreprises au risque cyber

Nous ne rentrerons pas dans les détails de chacun de ces axes mais nous apportons tout de même des éléments qui justifient l'intérêt de notre problématique. En effet, le premier axe invite les assureurs à "rendre plus explicites les clauses de couverture et d'exclusion des risques cyber et à mieux évaluer l'exposition de leur portefeuille d'assurance au risque". Ce point cherche à inciter les assureurs à mieux travailler leur rédaction de clause pour atténuer le risque d'exposition silencieuse. De plus l'évaluation des expositions des portefeuilles au risque cyber rejoint l'intérêt de créer un cadre d'évaluation dont nous avons présenté le nôtre dans la section (3.1.3). Toujours dans le premier axe, l'étude invite l'ACPR à conduire une étude afin de "mieux évaluer et comprendre le phénomène, insuffisamment documenté, des couvertures non-affirmatives". En effet, lorsque nous avons tenté d'appliquer notre modèle au cas du cyber silencieux pour la première fois, nous avons manqué de documentation sur la façon dont étaient gérées les expositions silencieuses. Dans ce contexte, nous nous sommes tournés vers les pratiques du marché britannique pour adapter leur cadre d'évaluation à notre étude. Un dernier point soulevé par l'institut des actuaires concerne l'identification des données. En effet, une bonne méthodologie passerait par l'identification des données relatives à l'accumulation : "informations permettant de relier un incident à un groupe d'incidents ou son lien avec d'autres risques [...]. Elles

permettent également de mettre en exergue les ≪chaînes de contamination ≫". Cette catégorie de données permettrait la création d'un réseau plus précis pour modéliser les liens entre assurés.

Un dernier point tout de même important de ce rapport, est la position prise par le ministère de l'Economie concernant le paiement des rançons à la suite d'une attaque cyber. En effet, il est favorable à un paiement des rançons sous condition d'un dépôt de plainte sous 48h. Cette nouvelle disposition sera incluse dans le projet de loi d'orientation et de programmation du ministère de l'intérieur (LOPMI), voir Vie Publique, 2022.

L'étude que nous avons développé dans ce mémoire est soutenue par les intérêts visant à développer l'assurance du risque cyber. De plus, ce risque est toujours une priorité au niveau national et la stratégie présentée dans Direction générale du Trésor, 2022 permettrait "l'affirmation de la place de Paris comme un pôle d'expertise cyber".

Conclusion

Le risque cyber représente un enjeu de grande envergure. En effet, toutes les entreprises, individus ou institutions sont exposées à ce risque.

L'une des particularités du cyber est le risque d'accumulation qui peut générer des évènements comme Wannacry en 2017. De plus, chez les assureurs, la couverture du risque cyber peut se faire de manière affirmative, par une garantie explicite qui couvre les évènements cyber, ou non affirmative, lorsqu'une garantie d'assurance non-vie ne les exclut pas. Le problème des couvertures silencieuses (non-affirmatives) fait porter à des garanties non-cyber les risques cyber. Dans ce contexte, nous avons modélisé le risque d'accumulation cyber pour des garanties non cyber.

Pour ce faire nous avons commencé par étudier les modèles d'épidémiologie avec une structure de réseau afin de prendre en compte les liens qui peuvent exister entre les assurés. En parallèle, nous avons défini une méthode qui permet d'évaluer les expositions silencieuses dans les portefeuilles d'assurance. Ainsi, nous avons construit un modèle qui permet à partir d'un scénario, de générer les pertes probables lors d'un évènement d'accumulation.

Afin d'illustrer notre modélisation, nous avons appliqué à un portefeuille fictif, un scénario qui paralyse le système informatique des assurés et active uniquement la garantie perte d'exploitation. L'interconnexion entre les assurés a été modélisée à l'aide d'un réseau et dépend du secteur d'activité de chacun d'entre eux.

Ainsi, nous avons remarqué que la prise en compte de ce scénario chez les assureurs permet une diminution de la perte moyenne. Cette prise en compte consiste à diminuer le temps de rétablissement des systèmes informatique des assurés en cas d'attaque. De plus, nous avons également montré, comment les proportions des assurés par secteur influencent la diffusion du virus et les coûts probables. De cette façon, nous considérons qu'une étude sur les structures de dépendance entre assurés permettrait de mieux modéliser la propagation d'un virus informatique, mais permettrait également de mieux gérer les risques dérivant de l'interdépendance entre les assurés (entreprises en sous-traitance, collaborations pour des projets de grande envergure, etc).

La modélisation que nous avons faite de ce risque demande d'être complétée par d'autres études. En effet, la construction des "taux silent" pourrait être approchée par des méthodes de machine learning. De même, le choix du réseau peut être justifié par une analyse des interactions entre assurés en exploitant les données chez les assureurs. Finalement, la calibration du modèle épidémiologique peut être explorée par des méthodes bayésiennes afin de répliquer des évènements historiques comme NotPetya.

Bibliographie

Pennec & Michau (s. d.). Les clauses d'exclusion de garantie dans les contrats d'assurance. Pennec Michau - Avocats Associés. URL : https://www.pennec-michau.com/2017/12/01/decembre-2017-fiche-n2-a-savoir-a-conseiller-clauses-dexclusion-de-garantie-contrats-dassurance/ (visité le 01/08/2022).

ACPR (2019). Communiqué de presse: La distribution des garanties contre les risques cyber par les assureurs. France Regulation. Paris : Banque de France. URL : https://acpr.banque-france.fr/sites/default/files/medias/documents/20191112_cp_bilan_cyber_assurance.pdf.

Allen, L. J. (2017). A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. *Infectious Disease Modelling* 2.2, p. 128-142.

AMRAE (2022). Lumière sur la cyberassurance. Repport. LUCY.

ANSSI (2020). Rapport Annuel 2020, p. 60.

ANSSI (2021). état de la menace rançongiciel 2021. Rapp. tech. ANSSI. URL : https://www.cert.ssi.gouv.fr/uploads/CERTFR-2021-CTI-001.pdf.

ANSSI (2022). panorama de la menace informatique 2021. Rapp. tech. ANSSI. URL : https://www.cert.ssi.gouv.fr/uploads/20220309_NP_WHITE_ANSSI_panorama-menace-ANSSI.pdf.

ANSSI et BSI (2021). Fourth edition of the Franco-German Common Situational Picture. Rapp. tech. ANSSI et BSI. URL : https://www.ssi.gouv.fr/uploads/2021/11/anssi_bsi_csp_2021.pdf.

Bang-Jensen, J. et Gutin, G. Z. (2009). Digraphs: Theory, Algorithms and Applications. Springer.

Bessy-Roland, Y. (2019). Modélisation stochastique individuelle de sinistres cyber. Mémoire d'actuariat. EURIA, Université de Bretagne Occidentale.

Bessy-Roland, Y., Boumezoued, A. et Hillairet, C. (2021). Multivariate Hawkes process for cyber insurance. *Annals of Actuarial Science* 15.1, p. 14-39.

BFMTV (2022). "Nous vivons la fin de l'abondance": les mots graves d'Emmanuel Macron en conseil des ministres. URL : https://www.bfmtv.com/politique/nous-vivons-la-fin-de-l-abondance-les-mots-graves-d-emmanuel-macron-en-conseil-des-ministres_AV-202208240240.html.

Bitdefender (2016). Ransomware A Victim's Perspective. Working paper. A study on US et European Internet Users.

Bonnac, A. (2020). Tarification du cyber-risque pour les collectivités locales : une modélisation inspirée par le modèle pendémie. Mémoire d'actuariat. ISFA, Univ. Claude Bernard Lyon 1.

CAIDA (2021). CAIDA's IPv4 and IPv6 AS Core: Visualizing IPv4 and IPv6 Internet Topology at a Macroscopic Scale in 2020. CAIDA. URL : https://www.caida.org/projects/as-core/2020/.

Cartagena, S., Gosrani, V., Grewal, J. et Pikinska, J. (2020). Silent cyber assessment framework. *British Actuarial Journal* 25, e2.

Chen, Y. (2016). Thinning algorithms for simulating point processes. *Florida State University, Tallahassee, FL.*

Ciecka, J. E. (2008). Edmond Halley's life table and its uses. *J. Legal Econ.* 15, p. 65.

Ciortan, M. (jan. 2019). Spectral graph clustering and optimal number of clusters estimation. Towards Data Science. url : https://towardsdatascience.com/spectral-graph-clustering-and-optimal-number-of-clusters-estimation-32704189afbe (visité le 08/03/2022).

Cloudflare (2022). What are Petya and NotPetya? Learning Article. url : https://www.cloudflare.com/learning/security/ransomware/petya-notpetya-ransomware/.

Cohignac, T. et Kazi-Tani, N. (2020). Laplacian Spectra of Graphs and Cyber-Insurance Protection.

Commission européenne (2022). Politiques de cybersécurité. Bâtir l'avenir numérique de l'Europe. url : https://digital-strategy.ec.europa.eu/fr/policies/cybersecurity-policies.

Cour de Cassation - Chambre civile 1 (2001). Cour de Cassation, Chambre civile 1, du 22 mai 2001, 99-10.849, Publié au bulletin.

Cyberattaque de Colonial Pipeline (2022). Cyberattaque de Colonial Pipeline. Wikipédia - L'encyclopédie libre [en ligne]. url : https://fr.wikipedia.org/wiki/Cyberattaque_de_Colonial_Pipeline.

Daffron, J, Ruffle, S, Andrew, C, Copic, J, Quantrill, K, Smith, A et Leverett, E (2019). Bashe attack: Global infection by contagious malware. *Centre for Risk Studies (University of Cambridge): Cambridge, UK*.

De Brive, G. B. (2022). Modélisation du risque cyber pour un portefeuille d'assurance français. Mémoire d'actuariat. ISFA, Univ. Claude Bernard Lyon 1.

De la Harpe, P. et Gabriel, J.-P. (2010). ≪Daniel Bernoulli, pionnier des modèles mathématiques en médecine≫. url : http://images.math.cnrs.fr/Daniel-Bernoulli-pionnier-des-modeles-mathematiques-en-medecine.

Dietz, K. et Heesterbeek, J. (2002). Daniel Bernoulli's epidemiological model revisited. *Mathematical biosciences* 180.1-2, p. 1-21.

Direction générale du Trésor (2021). Lancement d'une concertation nationale sur l'assurance du risque cyber. url : https://www.tresor.economie.gouv.fr/Articles/2021/07/05/lancement-d-une-concertation-nationale-sur-l-assurance-du-risque-cyber.

Direction générale du Trésor (2022). Remise du rapport sur le développement de l'assurance du risque cyber. Article. Ministère de l'économie, des finances et de la souveraineté industrielle et numérique. url : https://www.tresor.economie.gouv.fr/Articles/2022/09/07/remise-du-rapport-sur-le-developpement-de-l-assurance-du-risque-cyber.

Duparquier, J. (1976). la table de mortalité de Halley. *Annales de démographie historique*, p. 485-503.

EIOPA (2020). EIOPA STRATEGY ON CYBER UNDERWRITING. Note. European Insurance et Occupational Pensions Authority. url : https://www.eiopa.europa.eu/sites/default/files/publications/cyber-underwriting-strategy-february-2020_0.pdf.

Elmokashfi, A., Kvalbein, A. et Dovrolis, C. (2010). On the scalability of BGP: The role of topology growth. *IEEE Journal on Selected Areas in Communications* 28.8, p. 1250-1261.

Euler, L. (1741). Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae*, p. 128-140.

Fahrenwaldt, M. A., Weber, S. et Weske, K. (2018). Pricing of cyber insurance contracts in a network model. *ASTIN Bulletin: The Journal of the IAA* 48.3, p. 1175-1218.

Fan, R. et Chung, K. (1997). Spectral graph theory. T. 92. American Methematical Society.

Forum des compétences (2022). Scénario de risques importants sur les corruptions des données. Working paper. Forum des compétences.

Forum des compétences et EGERIE (2022). Scénarios d'attaques fondamentaux et cartographie des risques cyber. Working paper. Forum des compétences.

France Assureurs (2022). France Assureurs appelle à faire de la lutte contre les menaces cyber une priorité nationale. Communiqué de Presse. url : https://www.franceassureurs.fr/wp-content/uploads/220421_FRANCE_ASSUREURS_CP-LIVRE-BLANC-DONNEES.pdf.

France Diplomatie (2022). La France et la cybersécurité. Lutter contre la criminalité organisée. url : https://www.diplomatie.gouv.fr/fr/politique-etrangere-de-la-france/securite-desarmement-et-non-proliferation/lutter-contre-la-criminalite-organisee/la-france-et-la-cybersecurite/.

Fédération française des assureurs (2022). cartographie prospective 2022 de l'assurance, p. 24.

Gkantsidis, C., Mihail, M. et Zegura, E. (2003). Spectral Analysis of Internet Topologies. *IEEE INFOCOM*.

Hagberg, A. A., Schult, D. A. et Swart, P. J. (2008). Exploring Network Structure, Dynamics, and Function using NetworkX. *Proceedings of the 7th Python in Science Conference*. Sous la dir. de Varoquaux, G., Vaught, T. et Millman, J. Pasadena, CA USA, p. 11 -15.

Hillairet, C. et Lopez, O. (2021). Propagation of cyber incidents in an insurance portfolio: counting processes combined with compartmental epidemiological models. *Scandinavian Actuarial Journal* 2021.8, p. 671-694.

Hillairet, C. et Lopez, O. (2022). Cyber-assurance : enjeux, modélisations et leviers de mutualisation. *Opinions et débats* 24, p. 64.

Hillairet, C., Lopez, O., d'Oultremont, L. et Spoorenberg, B. (oct. 2021). Cyber contagion: impact of the network structure on the losses of an insurance portfolio. working paper or preprint. url : https://hal.sorbonne-universite.fr/hal-03388840.

Hiscox (2021). Don't let cyber be a game of chance. Repport. Cyber Readiness Report.

Institut Louis Bachelier (2022). Cyber-risk: actuarial modeling. Research Initiative. url : https://sites.google.com/view/cyber-actuarial/home.

Kermack, W. O. et McKendrick, A. G. (1927). A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 115.772, p. 700-721.

Kiss, I. Z., Miller, J. S. et Simon, P. (2017). Mathematics of Epidemics on Networks: From Exact to Approximate Models. Springer International Publishing.

Kouakou, L. (2020). Réseaux et assurance : Optimisation d'un modèle de franchise collaborative dans le cadre d'un contrat d'assurance IARD. Mémoire d'actuariat. EURIA, Université de Bretagne Occidentale.

Laplacian matrix (juin 2022). Laplacian matrix. Wikipédia - L'encyclopédie libre [en ligne]. url : https://en.wikipedia.org/wiki/Laplacian_matrix.

Laurent, R. M. (2022). La gestion du risque cyber dans les captives. Mémoire d'actuariat. le CNAM.

Le Parisien (2022). Cyberattaques : le gouvernement légalise l'indemnisation des rançons. url : https://www.leparisien.fr/high-tech/cyberattaques-le-gouvernement-legalise-lindemnisation-des-rancons-07-09-2022-LQASQHV6WFAKNI4FMQ3BU6GZ4Y.php.

LeClair, J. (2015). National Cybersecurity Institute at Excelsior College Before the United States House of Representatives Committee on Small Business Small Business, Big Threat: Protecting Small Businesses from Cyber Attacks. Statement for the Record. Cyber Readiness Report. url : https://docs.house.gov/meetings/SM/SM00/20150422/103276/HHRG-114-SM00-20150422-SD003-U4.pdf.

Lloyd's (2022). Realistic Disaster Scenarios: Scenario Specification. Working paper. Lloyd's of London.

LMA (2021). Cyber War and Cyber Operation Exclusion Clauses. Lloyd's Market Association Bulletin. url : https://www.lmalloyds.com/LMA/News/LMA_bulletins/LMA_Bulletins/LMA21-042-PD.aspx.

Lyon, B. (juin 2022). The Internet 1997 - 2021. The OPTE project. url : https://www.opte.org/the-internet.

Magal, P., Seydi, O. et Webb, G. (2018). Final size of a multi-group SIR epidemic model: Irreducible and non-irreducible modes of transmission. *Mathematical biosciences* 301, p. 59-67.

Marin, J.-M., Pudlo, P., Robert, C. P. et Ryder, R. J. (2012). Approximate Bayesian computational methods. *Statistics and Computing* 22.6, p. 1167-1180.

Marsh (2020). "Silent Cyber" — Frequently Asked Questions. MARSH JTL Specialty. url : https://www.marsh.com/content/dam/marsh/Documents/PDF/US-en/silent-cyber-faq.pdf.

Miller, J. C. et Ting, T. (2020). Eon (epidemics on networks): a fast, flexible python package for simulation, analytic approximation, and analysis of epidemics on networks. *arXiv preprint arXiv:2001.02436*.

Mohurle, S. et Patil, M. (2017). A brief study of wannacry threat: Ransomware attack 2017. *International Journal of Advanced Research in Computer Science* 8.5, p. 1938-1940.

Monteilh, J. (2020a). La modélisation épidémiologique (1/3) : 1760, bernoulli et la variole. url : https://petiteshistoiresdessciences.com/2020/04/17/1760-bernoulli-et-la-variole-la-modelisation-epidemiologique-1ere-partie/.

Monteilh, J. (2020b). La modélisation épidémiologique (2/3) : 1760, bernoulli et la variole. url : https://petiteshistoiresdessciences.com/2020/04/17/1760-bernoulli-et-la-variole-la-modelisation-epidemiologique-1ere-partie/.

Monteilh, J. (2020c). La modélisation épidémiologique (3/3) : 1760, bernoulli et la variole. url : https://petiteshistoiresdessciences.com/2020/04/17/1760-bernoulli-et-la-variole-la-modelisation-epidemiologique-1ere-partie/.

Munich Re (2022). Cyber insurance: Risks and trends 2022. Publication. url : https://www.munichre.com/topics-online/en/digitalisation/cyber/cyber-insurance-risks-and-trends-2022.html.

Negreiro, A. et Del Mar, M. (2022). The NIS2 Directive: A high common level of cybersecurity in the EU. EU Legislation in Progress. Parlement Européen. url : https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2021)689333#:~:text=The%20NIS2%20Directive%3A%20A%20high%20common%20level%20of, common%20level%20of%20cybersecurity%20across%20the%20Member%20States..

OCDE (2018). Origin of value added in final demand.

Parlement Européen et Conseil de l'Union Européenne (2009). Directive 2009/138/CE du 25 novembre 2009 sur l'accès aux activités de l'assurance et de la réassurance et leur exercice (solvabilité II). OJ L. 335/I. url : https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02009L0138-20210630.

PRA (2016). Cyber insurance underwriting risk. Consultation Paper — CP39/16. Bank of England. url : https://www.bankofengland.co.uk/-/media/boe/files/prudential-regulation/consultation-paper/2016/cp3916.

PRA (2017). Cyber insurance underwriting risk. Policy Statement — PS15/17. Bank of England. url : https://www.bankofengland.co.uk/-/media/boe/files/prudential-regulation/policy-statement/2017/ps1517.

PRA et Bank of England (2019). Cyber underwriting risk: follow-up survey results. url : https://www.bankofengland.co.uk/prudential-regulation/letter/2019/cyber-underwriting-risk-follow-up-survey-results.

Prévention des risques majeurs (2022). Risque cyber. Gouvernement. url : https://www.gouvernement.fr/risques/risques-cyber.

RIGAUD, G. (2022). Modèle d'accumulation du risque cyber. Mémoire d'actuariat. Paris Dauphine, Université Paris Sciences et Lettres.

Ruffle, S., Leverett, É., Coburn, A., Copic, J., Kelly, S., Evan, T., Ralph, D., Tuveson, M., Bochmann, O., Pryor, L. et al. (2015). Business blackout: The insurance implications of a cyber attack on the US power grid. *Center Risk Stud., Univ. Cambridge, Cambridge, UK, Tech. Rep.*

Staff Writer (2017). Total WannaCry losses pegged at $4 billion. url : https://www.reinsurancene.ws/total-wannacry-losses-pegged-4-billion/#:~:text=Ransomware%20attacks%20have%20reached%

20a%20new%20peak%20this, according%20to%20Trend%20Micro%E2%80%99s%20security%20and%20threats%20report..

Van Mieghem, P. et Cator, E. (2012). Epidemics in networks with nodal self-infection and the epidemic threshold. *Physical Review E* 86.1, p. 016116.

Van Mieghem, P., Omic, J. et Kooij, R. (2009). Virus Spread in Networks. *IEEE/ACM Transactions on Networking* 17.1, p. 1-14.

Ver Steeg, G. et Galstyan, A. (2012). Information transfer in social media. *Proceedings of the 21st international conference on World Wide Web*, p. 509-518.

Verizon (2022). Data Breach Investigations Report. Rapp. tech. AVerizon. url : https://www.verizon.com/business/resources/reports/2022/dbir/2022-data-breach-investigations-report-dbir.pdf.

Vie Publique (2022). Projet de loi d'orientation et de programmation du ministère de l'intérieur. url : https://www.vie-publique.fr/loi/284424-projet-loi-lopmi-2023-27-orientation-programmation-ministere-interieur.

Vittorio, A. (2022). Merck's $1.4 Billion Insurance Win Splits Cyber From 'Act of War'. Bloomberg Law. url : https://news.bloomberglaw.com/privacy-and-data-security/mercks-1-4-billion-insurance-win-splits-cyber-from-act-of-war.

Vlcek, O. (mai 2018). WannaCry : le bilan un an plus tard. Blog Avast. url : https://blog.avast.com/fr/wannacry-le-bilan-un-an-plus-tard-avast (visité le 17/08/2022).

Xu, M. et Hua, L. (2019). Cybersecurity insurance: Modeling and pricing. *North American Actuarial Journal* 23.2, p. 220-249.

Zelnik-Manor, L. et Perona, P. (2004). Self-Tuning Spectral Clustering. *caltech*.

Zhuo, R., Huffaker, B., Claffy, K. et Greenstein, S. (2020). GDPR and internet interconnection. voxeu. url : https://voxeu.org/article/gdpr-and-internet-interconnection.

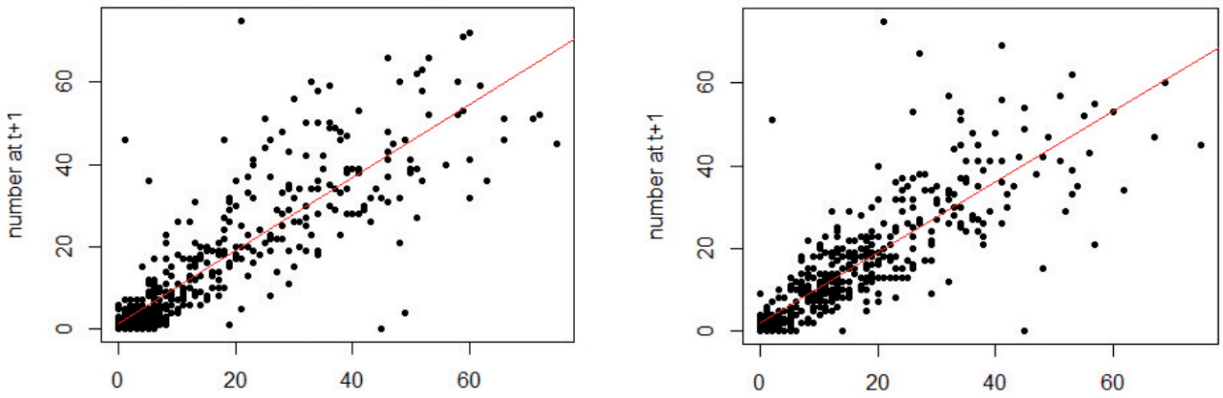
Annexe A

Compléments relatifs aux éléments présentés dans le mémoire.

A.1 Auto-corrélation des évènements Cyber

A.1.1 Mise en évidence de l'auto-corrélation

Comme nous l'avons présenté dans la section (1.2.2.1) du chapitre 1, la base de données PRC à permis de mettre en évidence l'auto-corrélation des évènements cyber. Une façon d'illustrer ce phénomène entre les évènements est de tracer les auto-corrélogrammes. Ainsi, en traçant sur un axe le nombre d'évènements survenus durant le mois t et sur l'autre, le nombre d'évènements survenus durant le mois $t+1$, nous pouvons illustrer la dépendance linéaire entre les deux.



(a) Régression par type d'organisation

(b) Régression par type d'attaque

FIGURE A.1 : Régression du nombre d'événements pendant le mois t+1 en fonction du nombre d'événements pendant le mois t issue des données de la base PRC (HILLAIRET et LOPEZ, 2022).

Sur la Figure (A.1), nous pouvons observer la présence d'une forte auto-corrélation des évènements selon s'ils sont classés par type d'organisation visée ou par type d'attaque.

A.1.2 Les processus de Hawkes à noyau exponentiel

Comme nous l'avons rapidement évoqué dans la section (1.2.2.1) du chapitre 1, les processus de comptage où l'intensité du processus ponctuel associé est une constante (processus de Poisson homogènes) et où l'intensité est une fonction du temps (processus de Poisson non-homogènes) sont les plus communs. Nous allons maintenant présenter un processus de comptage où le processus ponctuel associé dépend non seulement du temps t mais aussi de tout le passé du processus ponctuel. Pour prendre en compte le "passé", ou l'histoire du processus ponctuel, nous utiliserons la filtration naturelle du processus N définie par $\mathcal{F}_t^N = \sigma\{N(s); s \leq t\}$. En d'autres termes, $\mathcal{F}_t^N$ définit une collection de sous-ensembles de N, cette collection contiendra toutes les combinaisons possibles de $N(s)$ pour tous les $s \leq t$.

Il est important de garder à l'esprit que $\lambda(.)$ est maintenant un processus aléatoire puisqu'il dépend des réalisations de $N(.)$.

Un processus de Hawkes univarié à décroissance exponentielle est défini par un processus de comptage N tel que :

1. N(t) = 0
2. L'intensité du processus $\lambda(.)$ est définie par :

$$\lambda(t|\mathcal{F}_{t-}^N) = \mu + \int_0^t \alpha e^{-\beta(t-s))}\, dN(s) = \mu + \sum_{\{k:t_k<t\}} \alpha e^{-\beta(t-t_k)}$$

   with $\mu > 0$ et $0 < \alpha < \beta$.

3. $\mathbb{P}[N(t+h) - N(t) = 1|\mathcal{F}_{t-}^N] = \lambda(t|\mathcal{F}_{t-}^N)h + o(h)$
4. $\mathbb{P}[N(t+h) - N(t) \geq 2|\mathcal{F}_{t-}^N] = o(h)$

Les conditions (3) et (4) peuvent être réécrites avec la fonction $t \to H_t(\omega)$:

$$(3) \iff \mathbb{P}[H_{\Delta t} = 1|\mathcal{F}_{t-}^H] = \mathbb{P}\left[(H_{t+\Delta t} - H_t) = 1|\mathcal{F}_{t-}^H\right] = \lambda(t|\mathcal{F}_{t-}^H)\Delta t + o(\Delta t)$$

$$(4) \iff \mathbb{P}[H_{\Delta t} \geq 2|\mathcal{F}_{t-}^H] = \mathbb{P}\left[(H_{t+\Delta t} - H_t) \geq 2|\mathcal{F}_{t-}^H\right] = o(\Delta t)$$

Où $o(.)$ est une fonction telle que : $lim_{t\to 0}\frac{o(t)}{t} = 0$.

Regardons de plus près l'intensité $\lambda(t|\mathcal{F}_{t-}^N)$. Le premier paramètre μ représente l'intensité minimale du processus, nous pouvons remarquer que si nous retirons la somme dans l'expression de $\lambda(t|\mathcal{F}_{t-}^N)$ nous avons un processus de Poisson homogène. Le second paramètre de l'intensité est cette somme : $\sum_{:t_k<t\}} \alpha e^{-\beta(t-t_k)}$. Nous pouvons commencer par remarquer que la somme porte sur tous les événements passés t_k jusqu'au temps t, de plus c'est une somme strictement positive puisque la fonction exponentielle est positive, de plus $\mu > 0$ et $0 < \alpha < \beta$ par hypothèse. Ce qui signifie que nous allons ajouter une certaine intensité en fonction des événements passés. Si on regarde de près, l'exponentielle de la somme va ajouter une intensité en fonction de l'éloignement du temps t par rapport aux événements passés. Les points les plus éloignés ont moins d'impact sur l'intensité que les points les plus proches :

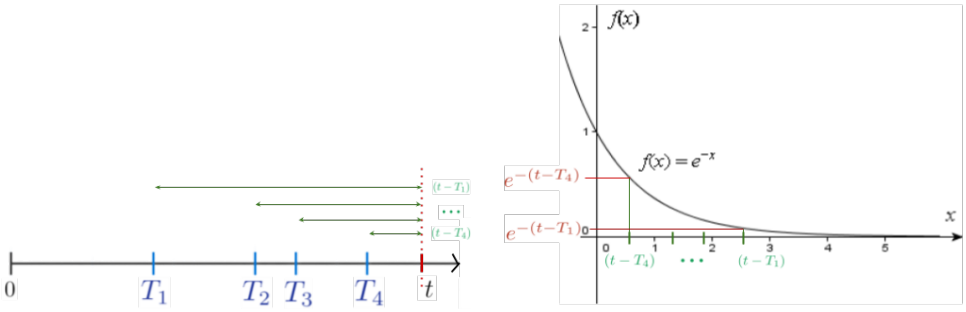


FIGURE A.2 : Représentation de $t - T_k$ et leur impact sur la fonction exponentielle

Les paramètres α et β nous permettent de calibrer la décroissance de l'exponentielle et la taille des sauts. Enfin, l'intensité d'un processus de Hawkes varie en fonction du nombre d'événements passés et de leur distance. A partir d'un événement t_k l'intensité va décroître exponentiellement jusqu'à l'arrivée de l'événement t_{k+1}, à ce moment $(t = t_{k+1})$, l'intensité va sauter avec une taille de α puisque nous aurons une arrivée supplémentaire dans la somme : La figure suivante montre comment l'intensité dépend des événements passés.

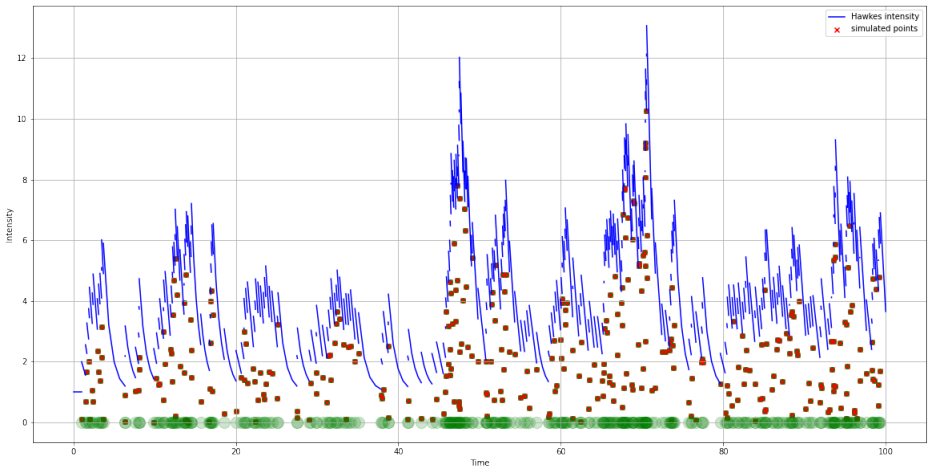


FIGURE A.3 : L'intensité du processus de Hawkes avec l'intensité : $\mu = 1 \quad \alpha = 1 \quad \beta = 1.2$

Simulation : Un processus de Hawkes peut être simulé à l'aide d'un algorithme d'acceptation-rejet similaire à celui utilisé pour le processus de Poisson non homogène. L'algorithme que nous utilisons est celui publié par Ogata en 1981 (figure A.5), l'idée est de simuler divers processus de Poisson non homogènes en utilisant une méthode de thinning où l'intensité du sup $\overline{\lambda} := Sup_{0 \geq t \geq T} \lambda(t)$ est mise à jour à chaque étape.

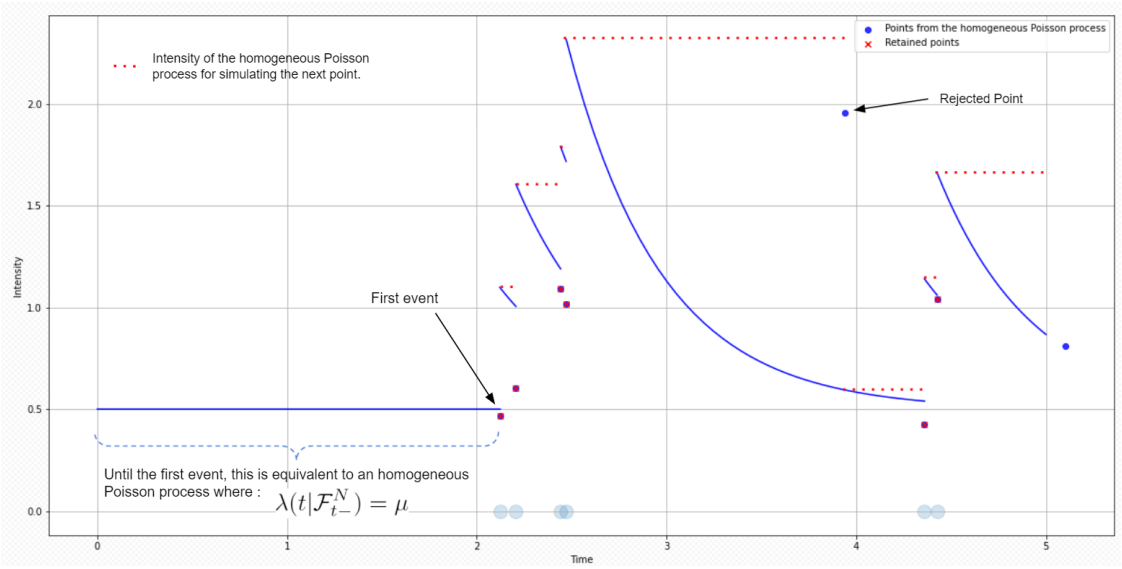


FIGURE A.4 : Méthode de thinning pour un procédé Hawkes où : $\mu = 0.5 \quad \alpha = 0.6 \quad \beta = 2$

Dans la figure précédente, nous pouvons voir comment l'intensité du processus de Hawkes varie : au début, l'histoire du processus est "vide" (ce qui implique que la somme dans le deuxième terme de l'expression est égale à zéro), ce qui signifie que l'intensité est juste une constante μ. Après le premier événement, l'intensité fait un bond de α et commence à décroître exponentiellement en prenant en mémoire tous les événements précédents. Pour simuler ce type de processus, nous commençons par simuler le premier événement comme un processus de Poisson homogène, les points suivants sont simulés en utilisant le thinning où le processus de Poisson homogène associé (ligne pointillée rouge) est égal au maximum de l'intensité au dernier point simulé (accepté ou non). L'algorithme suivant décrit la procédure précise :

Input: μ, α, β, T
Initialize $\mathcal{T} = \emptyset$, $s = 0$, $n = 0$;
while $s < T$ **do**
  Set $\bar{\lambda} = \lambda(s^+) = \mu + \sum_{\tau \in \mathcal{T}} \alpha e^{-\beta(s-\tau)}$;
  Generate $u \sim$ `uniform(0,1)`;
  Let $w = -\ln u/\bar{\lambda}$; `// so that` $w \sim$ `exponential(`$\bar{\lambda}$`)`
  Set $s = s + w$; `// so that` s `is the next candidate point`
  Generate $D \sim$ `uniform(0,1)`;
  if $D\bar{\lambda} \leq \lambda(s) = \mu + \sum_{\tau \in \mathcal{T}} \alpha e^{-\beta(s-\tau)}$ **then** `// accepting with prob.` $\lambda(s)/\bar{\lambda}$
    $n = n + 1$; `// updating the number of points accepted`
    $t_n = s$; `// naming it` t_n
    $\mathcal{T} = \mathcal{T} \bigcup \{t_n\}$; `// adding` t_n `to the ordered set` $\mathcal{T}$
  end
end
if $t_n \leq T$ **then**
  return $\{t_k\}_{k=1,2,\ldots,n}$
else
  return $\{t_k\}_{k=1,2,\ldots,n-1}$
end

FIGURE A.5 : Algorithme d'Ogata (CHEN, 2016)

A.2 Compléments pour la représentation des graphes

Dans cette annexe nous apportons des compléments théoriques sur la représentation des graphes avec une courte ouverture sur la méthode du *clusturing* spectral. Dans les configurations présentées dans la section 2.1.3.4 la matrice d'incidence est appelée :

- "matrice d'incidence sommets-arcs" pour les graphes orientés et est défini par :

$$I_{ij} = \begin{cases} -1 & \text{si l'arc } a_o^j \in A \text{ sort du nœud } s_i \in S, \\ 1 & \text{si l'arc } a_o^j \in A \text{ entre dans le nœud } s_i \in S, \\ 0 & \text{sinon.} \end{cases}$$

- "matrice d'incidence sommets-arête" pour les graphes orienté et est définit par :

$$I_{ij} = \begin{cases} 1 & \text{si l'arête } a_n^j \in A \text{ est reliée au nœud } s_i \in S, \\ 0 & \text{sinon.} \end{cases}$$

La matrice d'incidence pour le graphe G_n décrit dans la Figure (2.9) serait donc la matrice symétrique $I^n = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$ et pour le graphe G_o nous aurions $I^o = \begin{bmatrix} -1 & 1 & -1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 1 & -1 & 0 & 0 & -1 \end{bmatrix}$.

Finalement, nous introduisons une matrice essentielle à l'analyse spectrale des graphes, la **matrice Laplacienne**. L'analyse spectrale a d'importantes applications dans l'analyse des réseaux, citons par exemple l'analyse de la topologie du réseaux internet (Gkantsidis et al., 2003) qui conclut notamment à un plus grand développement du réseau nord américain que ceux des autres continents. Une autre application de l'analyse spectrale des graphes est celle du clusturing. En effet, il est bien connu que la méthode des K-means peut être limitée en fonction de la topologie induite par les données, l'analyse spectrale permet une meilleure discrimination des données et permet d'apporter certains éléments pour l'estimation du nombre de clusters (Ciortan, 2019).

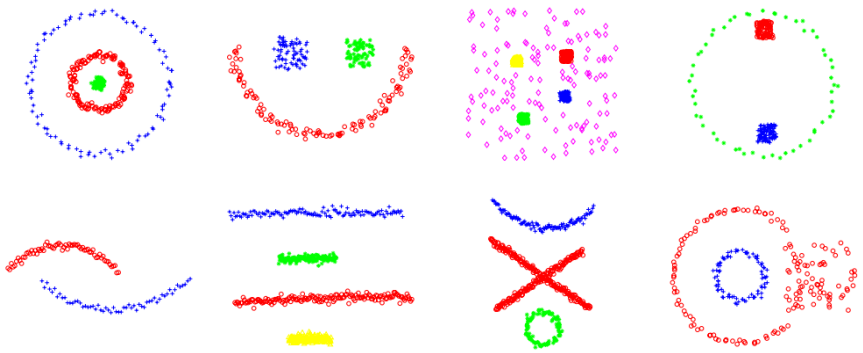


Figure A.6 : Résultats de clusturing spectral par initialisation automatisée (Zelnik-Manor et Perona, 2004).

Il existe plusieurs méthodes algébriques pour pour trouver la matrice Laplacienne qui dépendent entre autre du type de graphe à étudier. Nous en donnons les plus courantes (Laplacian matrix, 2022).

- Pour un graphe non orienté, nous pouvons définir la matrice Laplacienne par :

  $L = D - R$ où D est la matrice des degrés du graphe et R la matrice d'adjacence.

- Pour un graphe orienté, nous pouvons définir la matrice Laplacienne par :

  $L^+ = D^+ - R$ où D^+ est la matrice des demi-degrés extérieurs du graphe et R la matrice d'adjacence.

  $L^- = D^- - R$ où D^- est la matrice des demi-degrés intérieurs du graphe et R la matrice d'adjacence.

Pour le clusturing spectral il est également utile d'avoir la matrice Laplacienne symétrique et normalisée comme donnée si dessous :

$$L^{sym} = \sqrt{D^{NP}} L \sqrt{D^{NP}} \quad \text{où, } D^{NP} \text{ est la matrice inverse de Moore-Penrose.}$$

Ainsi nous avons pour les exemples des Figures (2.8) et (2.9) :

$$L^n = D^n - R^n = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} - \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 3 & -1 & -1 & -1 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 2 & -1 \\ -1 & 0 & -1 & 2 \end{bmatrix}$$

$$L^{o+} = D^{o+} - R^o = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ -1 & 0 & 2 & -1 \\ -1 & 0 & -1 & 1 \end{bmatrix}$$

$$L^{o-} = D^{o-} - R^o = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & -1 \\ -1 & 0 & -1 & 1 \end{bmatrix}$$

A.3 Brève analyse du modèle SIS sur réseau

De la même façon que pour l'analyse du modèle SIR, nous proposons dans cette annexe quelques éléments d'analyse pour le modèle SIS sur réseau. Nous commencerons donc par observer comment les paramètres influencent la diffusion du virus avant d'étudier l'impact du réseau sur la diffusion.

A.3.1 Impact des paramètres sur la diffusion du virus

Comme nous avons pu le voir dans la section (2.3), les taux d'infection et de contagion influencent directement la vitesse de diffusion du virus au sein du réseau.

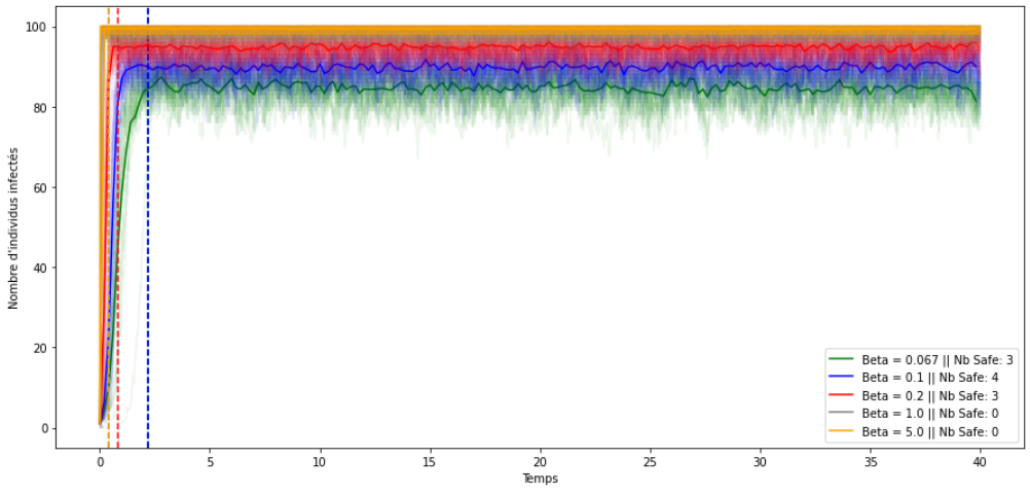


FIGURE A.7 : Évolution du nombre d'infectés en fonction du taux de contagion β.

Sur la figure (A.7), nous illustrons comment le virus se propage en fonction du taux d'infection. Pour cela nous avons fixé le taux de rétablissement à 1. De plus, le graphe considéré est un graphe complet ou tous les nœuds (100 nœuds) sont reliés entre eux. Ainsi, nous observons comment plus le taux de contagion est élevé, plus la diffusion du virus est rapide. Les droites pointillées représentent les instants où la courbe moyenne se stabilise et nous désignons par "NB_safe" les trajectoires qui atteignent un état qui stoppe la diffusion du virus (plus d'infectés).

A.3.2 Impact de la condition initiale sur un réseau internet

De la même façon que nous l'avons fait dans la section (2.3.3), nous introduisons un réseau internet en utilisant l'approche décrite dans (ELMOKASHFI et al., 2010). Nous illustrons sur la figure (A.8) le réseau en question.Ainsi nous allons observer comment se comporte la diffusion du virus en fonction de l'initialisation du virus. Le nœud 0 se trouve au centre du réseau tandis que le nœud 87 se trouve au sein d'une petite compagnie.

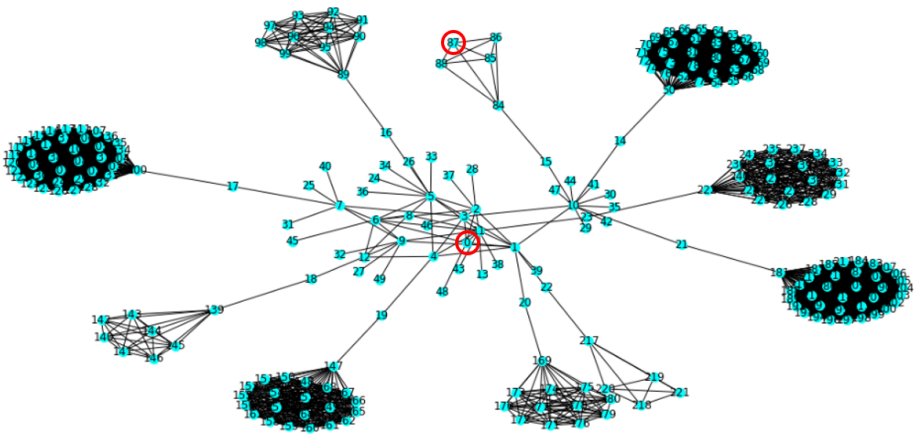
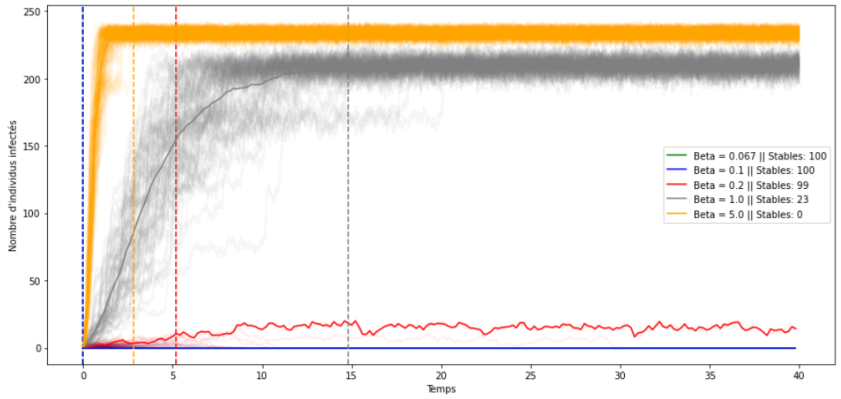
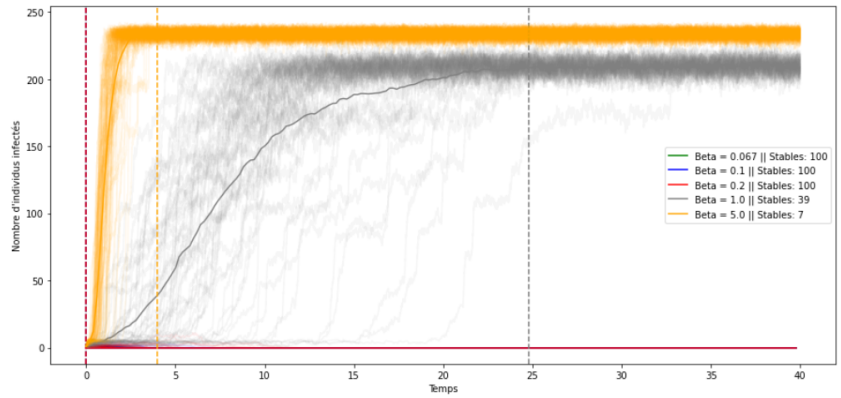


FIGURE A.8 : Réseau internet utilisé pour l'étude de la condition initiale.

Ainsi nous allons observer comment se comporte la diffusion du virus en fonction de l'initialisation du virus. Le nœud 0 se trouve au centre du réseau tandis que le nœud 87 se trouve au sein d'une petite compagnie.



(a) Initialisation au nœud 0



(b) Initialisation au nœud 87

FIGURE A.9 : Évolution du nombre d'infectés en fonction de la condition initale et du taux de contagion.

Nous désignons par “Stables” les trajectoires qui atteignent un état où le virus n’est plus en mesure de se diffuser (élimination du virus au sein du graphe). Nous remarquons sur le figures (A.9) comment les instants dans lesquels les droites se stabilisent se décalent en fonction de la condition initiale. Ainsi, nous pouvons dire qu’une initialisation au nœud 0, c’est a dire au centre du réseau, favorise la diffusion du virus (voir la courbe de taux de contagion $\beta = 1$).

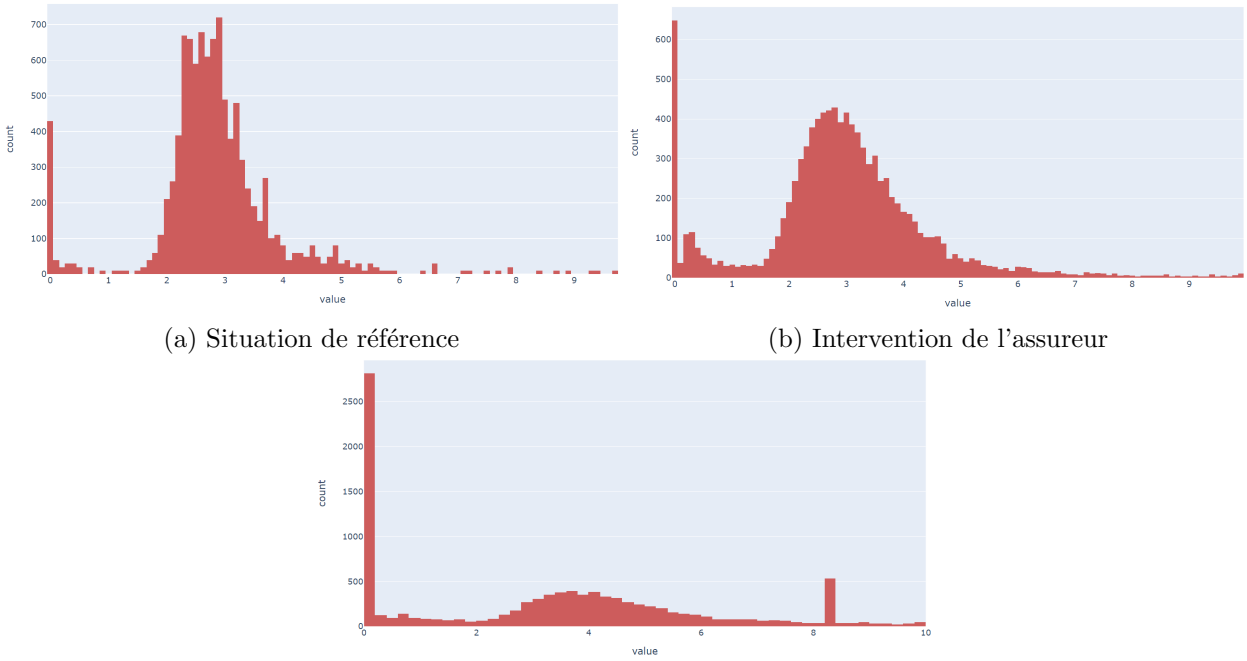
A.4 Analyse des pics des trajectoires au cours du temps

De façon générale, dans les sections (3.2.2 et 3.2.3) nous nous sommes intéressés aux pertes engendrées par la diffusion du virus sur la garantie perte d'exploitation. Cependant pour les assureur qui décident d'assister les assurés en cas d'infection, il se peut que les capacités d'intervention se saturent du fait du grand nombre d'infections. Dans cette annexe nous allons donc nous intéresser à la distribution des pics des pertes au cours du temps. Sur l'ensemble des figures suivantes, nous illustrons les distributions des pics. C'est à dire, avec toutes les trajectoires simulées (10 000 simulations), nous allons tracer l'histogramme des temps auxquels surviennent les pics des pertes engendrées par le virus.

La situation de référence correspond à la situation du portefeuille et des paramètres de diffusion de la section (3.2.2.2). C'est à dire une equi-distribution des assurés au sein des différents secteurs, un taux de contagion de 0.01 et un taux de rétablissement de 1. L'intervention de l'assureur correspond à la situation lorsque l'assureur décide d'intervenir en assistant les assurés infecté à rétablir au plus vite leur situation, voir section (3.2.3.1). Les paramètres du modèle sont donc de 0.01 pour le taux d'infection et de 1.5 pour le taux de rétablissement, avec une équi-distribution des assurés au sein des secteurs. Finalement, le portefeuille restructuré correspond au portefeuille tel que présenté dans la section (3.2.3.2). Dans ce dernier cas de figure, le nombre d'assurés dans le secteur *Mining* est réduit de moitié et les paramètres de diffusion sont les mêmes que dans la situation de référence.

A.4.1 Vue d'ensemble

Nous commençons par tracer les distribution des pics d'infection au cours du temps, tous secteurs confondus, sur la figure (A.10).



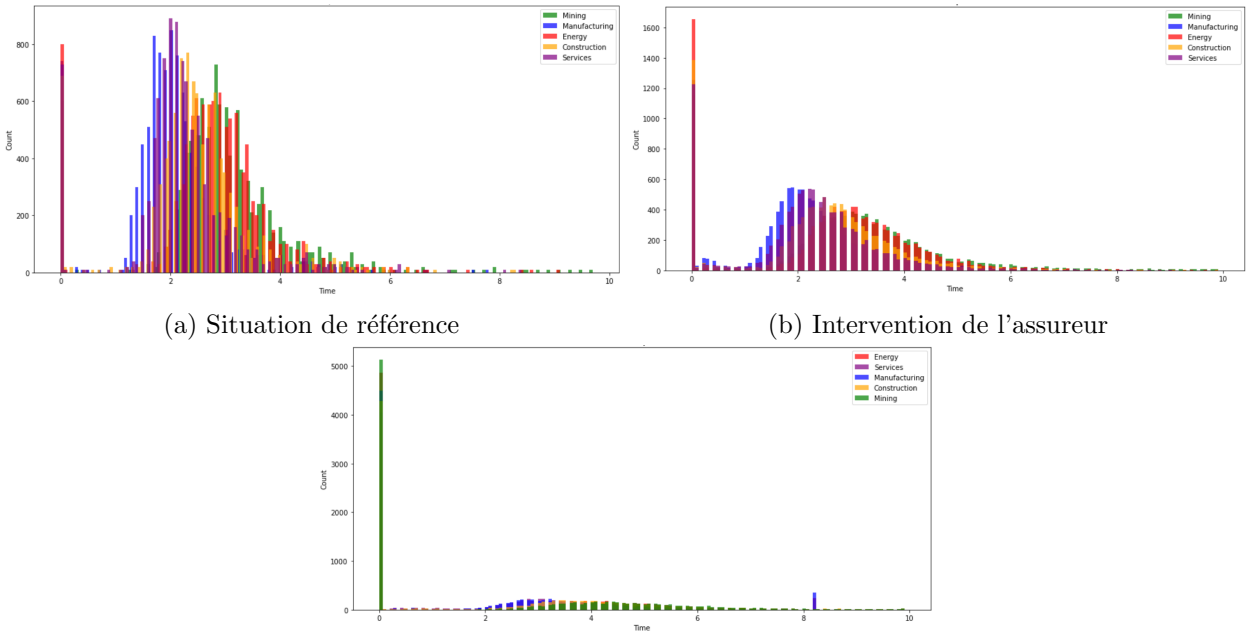
(a) Situation de référence

(b) Intervention de l'assureur

(c) Portefeuille restructuré

FIGURE A.10 : Distribution des pics des pertes au cours du temps.

Ainsi, nous observons comment le nombre de trajectoires ayant des pertes nulles augmentent lorsque nous prenons en compte le risque d'accumulation. Cependant nous pouvons remarquer sur la figure (A.11) que la distribution des pics dépend des secteurs.



(a) Situation de référence

(b) Intervention de l'assureur

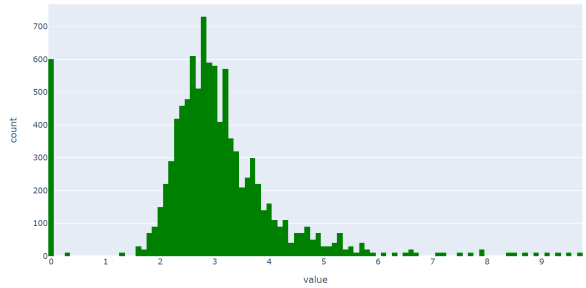
(c) Portefeuille restructuré

FIGURE A.11 : Distribution des pics des pertes au cours du temps par secteurs.

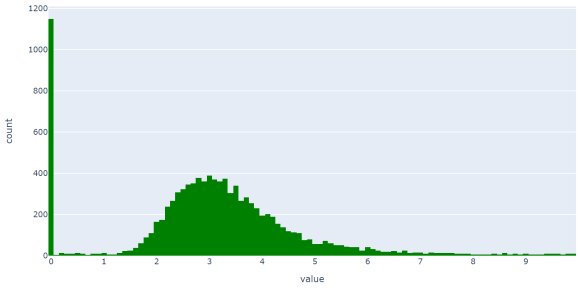
Dans la section suivante, nous présenterons les distributions secteurs par secteurs afin de donner plus de détails sur la façon dont elles évoluent en fonction des hypothèses.

A.4.2 Distributions secteur par secteur

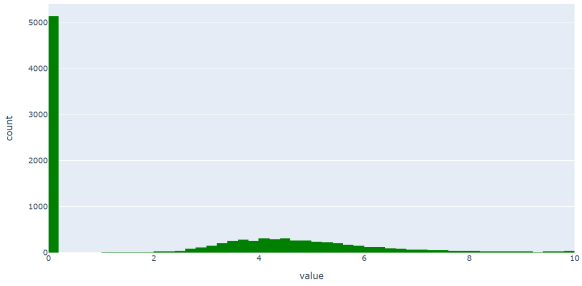
Dans cette section nous ne commenterons pas les différentes figures. Cependant nous pouvons remarque de façon générale, que les leviers à dispositions des assureurs (taux de rétablissement et structure du portefeuille) permettent de réduire les pics d'infections au cours du temps.



(a) Situation de référence

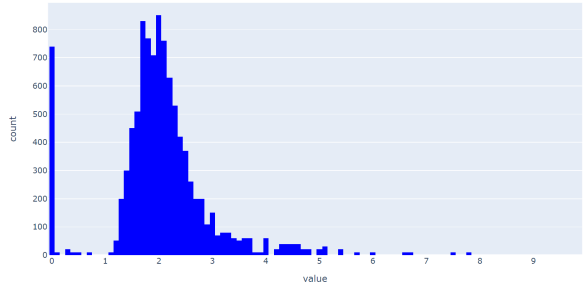


(b) Intervention de l'assureur

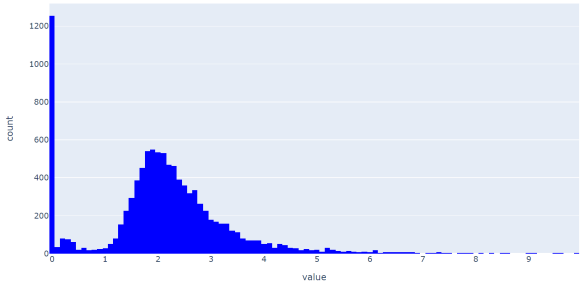


(c) Portefeuille restructuré

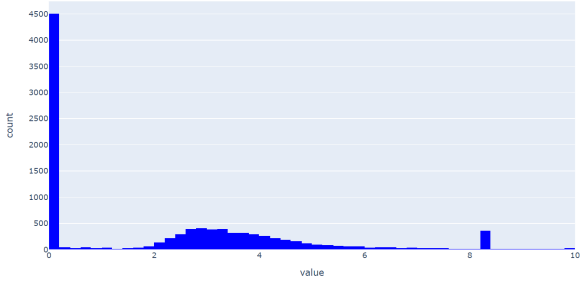
FIGURE A.12 : Distribution des pics des pertes au cours du temps pour le secteur *Mining*.



(a) Situation de référence

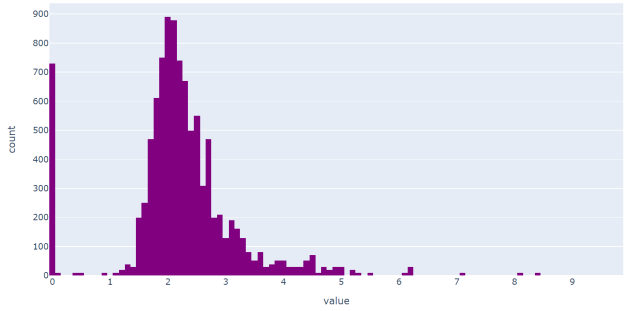


(b) Intervention de l'assureur

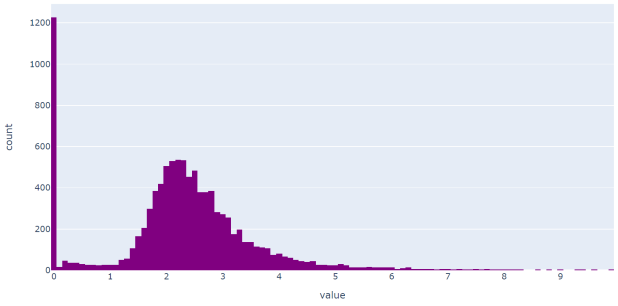


(c) Portefeuille restructuré

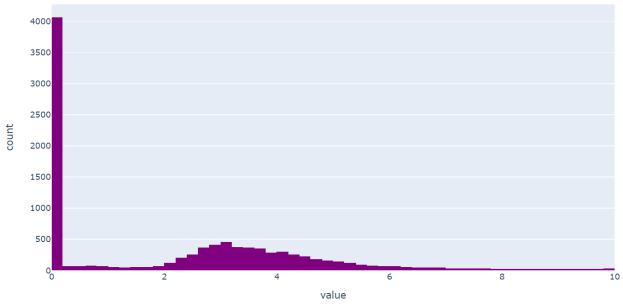
FIGURE A.13 : Distribution des pics des pertes au cours du temps pour le secteur *Manufacturing*.



(a) Situation de référence

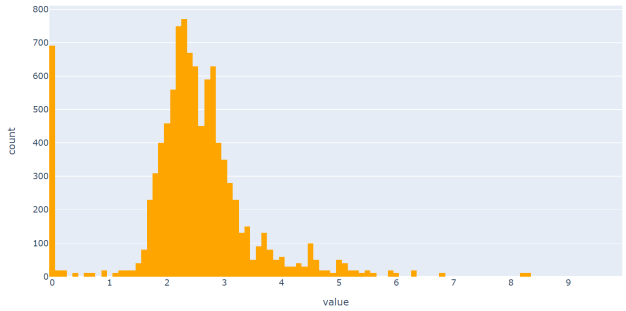


(b) Intervention de l'assureur

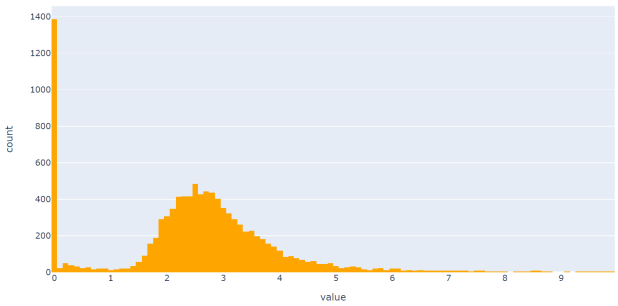


(c) Portefeuille restructuré

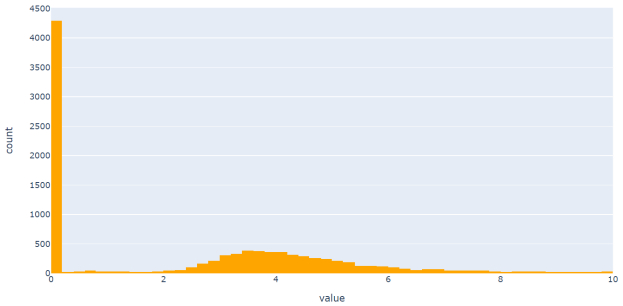
FIGURE A.14 : Distribution des pics des pertes au cours du temps pour le secteur *Services*.



(a) Situation de référence

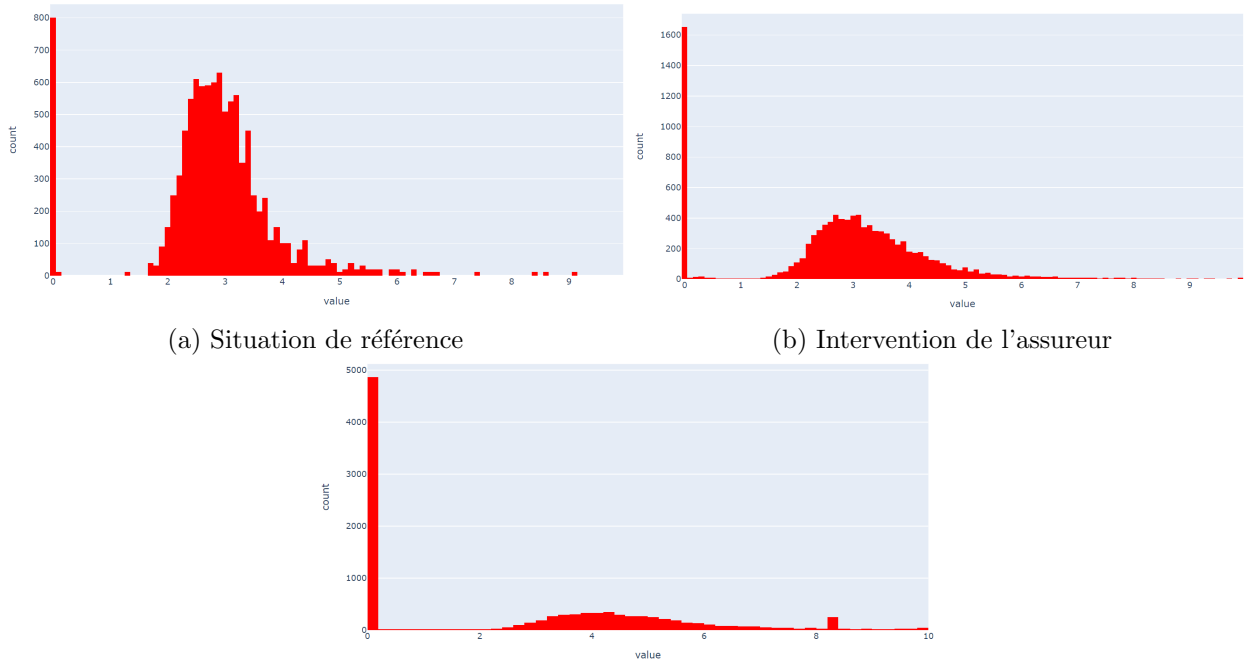


(b) Intervention de l'assureur



(c) Portefeuille restructuré

FIGURE A.15 : Distribution des pics des pertes au cours du temps pour le secteur *Construction*.



(a) Situation de référence

(b) Intervention de l'assureur

(c) Portefeuille restructuré

FIGURE A.16 : Distribution des pics des pertes au cours du temps pour le secteur *Energy*.

L'objectif de cette annexe était d'illustrer la richesse de la modélisation du risque d'accumulation par des modèles de réseaux. Une analyse plus fine pourrait permettre de hiérarchiser l'intervention des assureurs au cours du temps sur les entreprises infectées en fonction des coûts qu'elles sont susceptibles d'engendrer.

A.5 Pseudo-codes des algorithmes de Gillespie et event drive

Nous présentons dans la suite les pseudo-codes associés aux différents algorithmes présentés en fin de Chapitre 2. Sur la Figure (A.17) nous proposons le pseudo-code pour l'algorithme de Gillespie issu de Kiss et al., 2017.

Input: Network G, per-edge transmission rate τ, recovery rate γ, set of index node(s) initial_infecteds, maximum time t_{max}.
Output: Lists times, S, I, and R giving number in each state at each time.

```
function Gillespie_network_epidemic(G, τ, γ, initial_infections, t_max)
    times, S, I, R ← [0], [|G|-len(initial_infections)], [len(initial_infections)], [0]
    infected_nodes ← initial_infections
    at_risk_nodes ← uninfected nodes with infected neighbours
    for each node u in at_risk_nodes do
        infection_rate[u] = τ × number of infected neighbours
    total_infection_rate ← ∑_{u∈at_risk_nodes} infection_rate[u],
    total_recovery_rate ← γ × len(infected_nodes)
    total_rate ← total_transmission_rate + total_recovery_rate
    time ← exponential_variate(total_rate)
    while time< t_max and total_rate> 0 do
        r = uniform_random(0,total_rate)
        if r <total_recovery_rate then
            u = random.choice(infected_nodes)
            remove u from infected_nodes
            reduce infection_rate[v] for u's susceptible neighbours v
        else
            choose u from at_risk_nodes with probability infection_rate[u] / total_infection_rate.
            remove u from at_risk_nodes
            add u to infected_nodes
            for susceptible neighbours v of u do
                if v not in at_risk_nodes then
                    add v to at_risk_nodes
                update infection_rate[v]
        update times, S, I, and R
        update total_recovery_rate, total_infection_rate, and total_rate
        time ← time + exponential_variate(total_rate)
    return times, S, I, R
```

Figure A.17 : Pseudo-code de l'agorithme de Gillespie, issu de Kiss et al., 2017.

A présent nous proposons le pseudo code pour l'algorithme event-drive permettant de simuler la propagation d'un virus au sein d'un graphe. L'algorithme principal est présenté sur la Figure (A.18) et les fonctions auxiliaires nécessaires à sa réalisation sont présentés sur la Figure (A.19).

```
Input: Network G, per-edge transmission rate τ, recovery rate γ, set of index node(s)
  initial_infecteds, and maximum time t_max.
Output: Lists times, S, I, and R giving number in each state at each time.

  function fast_SIR(G,τ, γ, initial_infecteds, t_max)
      times, S, I, R ← [0], [|G|], [0], [0]
      Q ← empty priority queue
      for u in G.nodes do
          u.status ← susceptible
          u.pred_inf_time ← ∞
      for u in initial_infecteds do
          Event ← {node: u, time: 0, action: transmit}
          u.pred_inf_time ← 0
          add Event to Q                                   ▷ ordered by time
      while Q is not empty do
          Event ← earliest remaining event in Q
          if Event.action is transmit then
              if Event.node.status is susceptible then
                  process_trans_SIR(G, Event.node, Event.time, τ, γ, times, S, I, R, Q, t_max)
          else
              process_rec_SIR(Event.node, Event.time, times, S, I, R)
      return times, S, I, R
```

FIGURE A.18 : Pseudo-code pour l'algorithme event-drive du modèle SIR sur une structures de réseau, issu de KISS et al., 2017.

```
function process_trans_SIR(G, u, t, τ, γ, times, S, I, R, Q, t_max)
    append times, S, I, and R with t, S.last−1, I.last+1, and R.last
    u.status ← infected
    u.rec_time ← t+exponential_variate(γ)
    if u.rec_time< t_max then
        newEvent ← {node: u, time: u.rec_time, action: recover}
        add newEvent to Q
    for v in G.neighbours(u) do
        find_trans_SIR(Q, t, τ, u, v, t_max)
function find_trans_SIR(Q, t, τ, source, target, t_max)
    if target.status is susceptible then
        inf_time ← t+exponential_variate(τ)
        if inf_time < minimum(source.rec_time, target.pred_inf_time, t_max) then
            newEvent ← {node: target, time: inf_time, action: transmit}
            add newEvent to Q
            target.pred_inf_time ← inf_time
function process_rec_SIR(u, t, times, S, I, R)
    append times, S, I, and R with t, S.last, I.last−1, and R.last+1
    u.status ← recovered
```

FIGURE A.19 : Fonctions auxiliaires à l'algorithme du SIR sur réseau de le Figure (A.18), issu de KISS et al., 2017.

L'algorithme de Gillespie s'adapte bien au modèle SIR et SIS tandis que l'event-drive demande quelques petites modifications pour être adapté au modèle SIS. En effet, sur le modèle SIR, un individus infecté ne peut pas réinfecter la personne qui l'a infecté, alors que dans le modèle SIS, une fois le noeud infecteur rétablis, il peut se refaire infecté, y compris par un noeud qu'il à lui même infecté.

A.6 Annexe 1 de la directive du Parlement Européen et Conseil de l'Union Européenne, 2009

La prochaine Annexe rappelle les différentes branches d'assurance en Non-Vie. Elle est issue de l'annexe de la directive du Parlement Européen et Conseil de l'Union Européenne, 2009.

Dans les prochaines pages, le lecteur trouvera l'annexe en question telle que présentée dans la directive.

▼B

ANNEXE I

CLASSIFICATION PAR BRANCHE D'ASSURANCE NON-VIE

A. **Classification des risques par branches d'assurance**

1. *Accidents (y compris les accidents de travail et les maladies professionnelles):*

   — prestations forfaitaires;

   — prestations indemnitaires;

   — combinaisons;

   — personnes transportées.

2. *Maladie:*

   — prestations forfaitaires;

   — prestations indemnitaires;

   — combinaisons.

3. *Corps de véhicules terrestres (autres que ferroviaires)*

   Tout dommage subi par:

   — véhicules terrestres automoteurs;

   — véhicules terrestres non automoteurs.

4. *Corps de véhicules ferroviaires*

   Tout dommage subi par les véhicules ferroviaires.

5. *Corps de véhicules aériens*

   Tout dommage subi par les véhicules aériens.

6. *Corps de véhicules maritimes, lacustres et fluviaux*

   Tout dommage subi par:

   — véhicules fluviaux;

   — véhicules lacustres;

   — véhicules maritimes.

7. *Marchandises transportées (y compris les marchandises, bagages et tous autres biens)*

   Tout dommage subi par les marchandises transportées ou bagages, quel que soit le moyen de transport.

8. *Incendie et éléments naturels*

   Tout dommage subi par les biens (autres que les biens compris dans les branches 3, 4, 5, 6 et 7) lorsqu'il est causé par:

   — incendie;

   — explosion;

   — tempête;

   — éléments naturels autres que la tempête;

   — énergie nucléaire;

   — affaissement de terrain.

▼B

9. *Autres dommages aux biens*

Tout dommage subi par les biens (autres que les biens compris dans les branches 3, 4, 5, 6 et 7) lorsque ce dommage est causé par la grêle ou la gelée, ainsi que par tout événement, tel le vol, autre que ceux compris dans la branche 8.

10. *R.C. véhicules terrestres automoteurs*

Toute responsabilité résultant de l'emploi de véhicules terrestres automoteurs (y compris la responsabilité du transporteur).

11. *R.C. véhicules aériens*

Toute responsabilité résultant de l'emploi de véhicules aériens (y compris la responsabilité du transporteur).

12. *R.C. véhicules maritimes, lacustres et fluviaux*

Toute responsabilité résultant de l'emploi de véhicules fluviaux, lacustres et maritimes (y compris la responsabilité du transporteur).

13. *R.C. générale*

Toute responsabilité autre que celles mentionnées sous les branches 10, 11 et 12.

14. *Crédit:*

— insolvabilité générale;

— crédit à l'exportation;

— vente à tempérament;

— crédit hypothécaire;

— crédit agricole.

15. *Caution:*

— caution directe;

— caution indirecte.

16. *Pertes pécuniaires diverses:*

— risques d'emploi;

— insuffisance de recettes (générale);

— mauvais temps;

— pertes de bénéfices;

— persistance de frais généraux;

— dépenses commerciales imprévues;

— perte de la valeur vénale;

— pertes de loyers ou de revenus;

— autres pertes commerciales indirectes;

— autres pertes pécuniaires non commerciales;

— autres pertes pécuniaires.

17. *Protection juridique*

Protection juridique.

▼B

18. *Assistance*

Assistance aux personnes en difficulté au cours de déplacements, d'absences de leur domicile ou de leur résidence habituelle.

B. **Appellation d'agréments donnés simultanément pour plusieurs branches d'assurance**

Lorsque l'agrément porte à la fois:

a) sur les branches 1 et 2, il est donné sous l'appellation «Accidents et maladie»;

b) sur les branches 1 (quatrième tiret), 3, 7 et 10, il est donné sous l'appellation «Assurance automobile»;

c) sur les branches 1 (quatrième tiret), 4, 6, 7 et 12, il est donné sous l'appellation «Assurance maritime et transport»;

d) sur les branches 1 (quatrième tiret), 5, 7 et 11, il est donné sous l'appellation «Assurance aviation»;

e) sur les branches 8 et 9, il est donné sous l'appellation «Incendie et autres dommages aux biens»;

f) sur les branches 10, 11, 12 et 13, il est donné sous l'appellation «Responsabilité civile»;

g) sur les branches 14 et 15, il est donné sous l'appellation «Crédit et caution»;

h) sur toutes les branches, il est donné sous l'appellation choisie par l'État membre intéressé, qui la communique aux autres États membres et à la Commission.

A.7 Exemples de clauses d'exclusions

Dans cette annexe nous présentons deux clauses d'exclusions pour le risque cyber :

WAR, HI-JACKING AND OTHER PERILS EXCLUSION CLAUSE (AVIATION)

This Policy does not cover claims caused by

(a) War, invasion, acts of foreign enemies, hostilities (whether war be declared or not), civil war, rebellion, revolution, insurrection, martial law, military or usurped power or attempts at usurpation of power.

(b) Any hostile detonation of any weapon of war employing atomic or nuclear fission and/or fusion or other like reaction or radioactive force or matter.

(c) Strikes, riots, civil commotions or labour disturbances.

(d) Any act of one or more persons, whether or not agents of a sovereign Power, for political or terrorist purposes and whether the loss or damage resulting therefrom is accidental or intentional.

(e) Any malicious act or act of sabotage.

(f) Confiscation, nationalisation, seizure, restraint, detention, appropriation, requisition for title or use by or under the order of any Government (whether civil military or de facto) or public or local authority.

(g) Hi-jacking or any unlawful seizure or wrongful exercise of control of the Aircraft or crew in Flight (including any attempt at such seizure or control) made by any person or persons on board the Aircraft acting without the consent of the Insured.

Furthermore this Policy does not cover claims arising whilst the Aircraft is outside the control of the Insured by reason of any of the above perils. The Aircraft shall be deemed to have been restored to the control of the Insured on the safe return of the Aircraft to the Insured at an airfield not excluded by the geographical limits of this Policy, and entirely suitable for the operation of the Aircraft (such safe return shall require that the Aircraft be parked with engines shut down and under no duress).

AVN48B
01.10.96

FIGURE A.20 : AVN 48B – War, Hi-Jacking and Other Perils Exclusion Clause (Source : Acurra International Limited).

1. Electronic Data Exclusion

Notwithstanding any provision to the contrary within the Policy or any endorsement thereto, it is understood and agreed as follows:-

a) This Policy does not insure loss, damage, destruction, distortion, erasure, corruption or alteration of ELECTRONIC DATA from any cause whatsoever (including but not limited to COMPUTER VIRUS) or loss of use, reduction in functionality, cost, expense of whatsoever nature resulting therefrom, regardless of any other cause or event contributing concurrently or in any other sequence to the loss.

ELECTRONIC DATA means facts, concepts and information converted to a form useable for communications, interpretation or processing by electronic and electromechanical data processing or electronically controlled equipment and includes programmes, software and other coded instructions for the processing and manipulation of data or the direction and manipulation of such equipment.

COMPUTER VIRUS means a set of corrupting, harmful or otherwise unauthorised instructions or code including a set of maliciously introduced unauthorised instructions or code, programmatic or otherwise, that propagate themselves through a computer system or network of whatsoever nature. COMPUTER VIRUS includes but is not limited to 'Trojan Horses', 'worms' and 'time or logic bombs'.

b) However, in the event that a peril listed below results from any of the matters described in paragraph a) above, this Policy, subject to all its terms, conditions and exclusions, will cover physical damage occurring during the Policy period to property insured by this Policy directly caused by such listed peril.

Listed Perils

Fire
Explosion

2. Electronic Data Processing Media Valuation

Notwithstanding any provision to the contrary within the Policy or any endorsement thereto, it is understood and agreed as follows:-

Should electronic data processing media insured by this Policy suffer physical loss or damage insured by this Policy, then the basis of valuation shall be the cost to repair, replace or restore such media to the condition that existed immediately prior to such loss or damage, including the cost of reproducing any ELECTRONIC DATA contained thereon, providing such media is repaired, replaced or restored. Such cost of reproduction shall include all reasonable and necessary amounts, not to exceed ..... any one loss, incurred by the Assured in recreating, gathering and assembling such ELECTRONIC DATA. If the media is not repaired, replaced or restored the basis of valuation shall be the cost of the blank media. However this Policy does not insure any amount pertaining to the value of such ELECTRONIC DATA to the Assured or any other party, even if such ELECTRONIC DATA cannot be recreated, gathered or assembled.

NMA2914

FIGURE A.21 : NMA2914 (Source : Insurance Endorsements).

A.8 Etapes détaillées du cadre d'évaluation de l'IFoA

Nous présentons dans cette annexe, les étapes détaillées du cadre d'évaluation de l'IFoA (Cartagena et al., 2020). En effet, dans la section (3.1.1) nous ne présentions que les principales étapes.

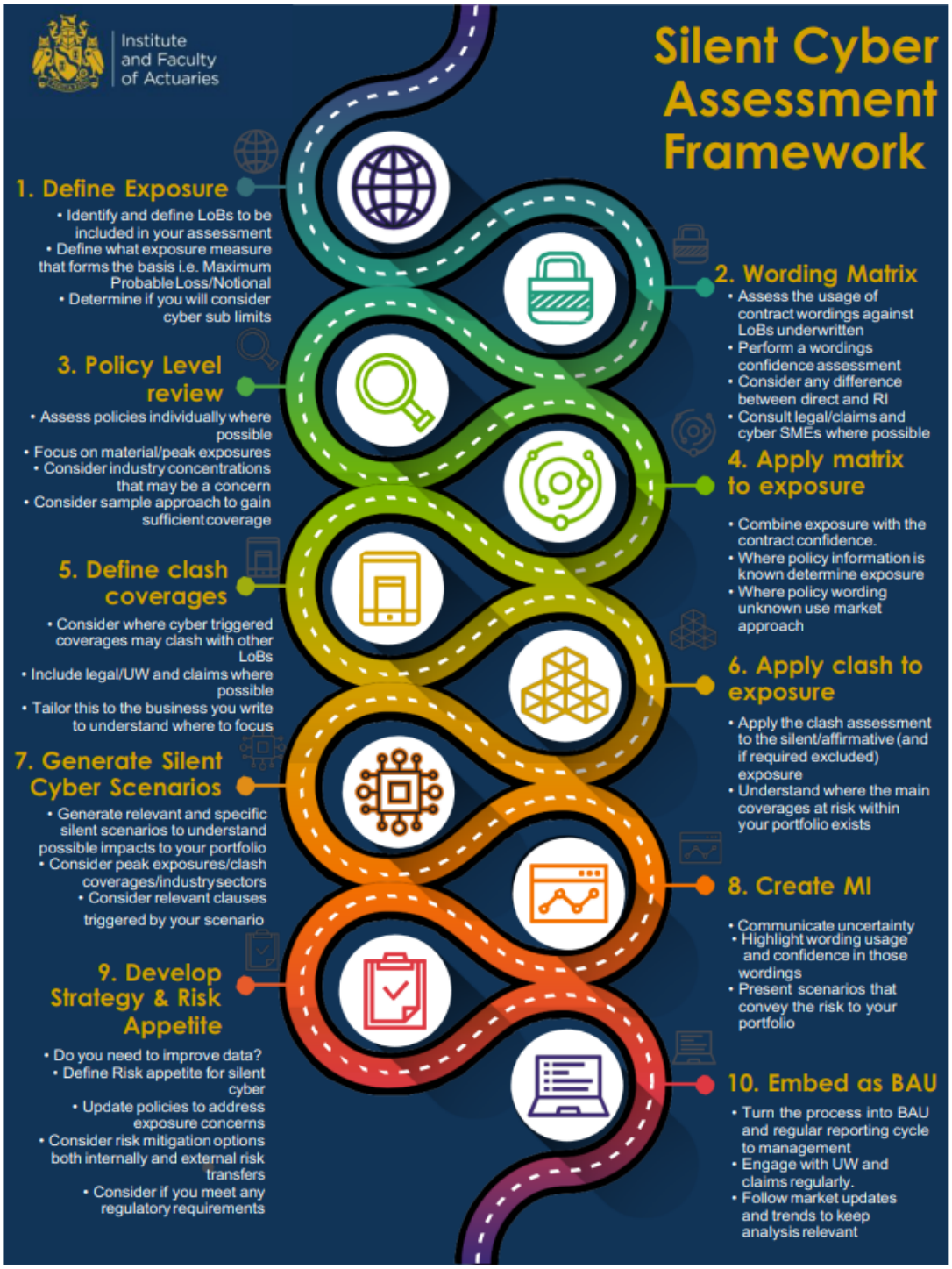


Figure A.22 : Étapes détaillées pour le cadre d'évaluation de l'IFoA (Cartagena et al., 2020).