Cyber contagion: impact of the network structure on the losses of an insurance portfolio

Caroline Hillairet\textsuperscript{1}, Olivier Lopez\textsuperscript{2}

\textsuperscript{1} ENSAE Paris, \textsuperscript{2} Sorbonne Université
Joint Research Initiative "Cyber insurance: actuarial modeling"
supported by Risk Fundation, AXA Research Fund

Webinaire Chaire PARI, 16 February 2022
Cyber-risk

- Various types of attacks (ransomware, phishing, classic frauds...)
- Strike states, companies, people.
- Huge costs: estimated to 1% of the global GDP.
- In France, number of ransomwares reported to ANSSI multiplied by 3 between 2019 and 2020.
- Report LUCY of AMRAE ("Association pour le Management des Risques et des Assurances de l'Entreprise") published in 2021:
  - Study of all French companies that take out a cyber insurance contract through a broker (2019 and 2020).
  - Volume of premiums increased of 49% between 2019 and 2020.
  - The ratio Claims / Premiums is 167% for 2020, 84% for 2019.
  - 2020 significant increase in Losses is due to the occurrence of 4 XXL claims.
Wannacry

- Ransomware Wannacry: worldwide cyber attack in May 2017.
- Use the vulnerability "EternalBlue".
- Approximately 200,000 infected computers across 150 countries over approximately one week.
- Estimation of the cost: hundreds of million dollars, billions according to some estimations. ($100 millions for the NHS).
Wannacry in maps
Wannacry in maps
Colonial Pipeline

- "Double extorsion": ransomware attack combined with blackmail.
- 4.2% increase of WTI and Brent.
- Authors: the hacker group "Darkside" (Ransomware as a service).
Cyber-risk specificities and mutualization issues

- Similarities with operational risk...
- ... but specificities in the structure of cyber events.

1. The risk is **new** and **evolves fast**.
2. "**Silent cyber**": non-cyber policies may contain guarantees that can be triggered by cyber events if not excluded.
3. **Extreme** events (huge losses can occur)
4. **Accumulation risk**: potential concentration of incidents which leads to loss of mutualization.

These features may endanger mutualization of cyber risk.
Outline

1. Introduction

2. Accumulation of claims and contagion
   - Multi-group SIR model
   - Topology of the network

3. Calibration and applications
   - How to calibrate a Wannacry type scenario
   - Identifying vulnerabilities
   - Impact of the reaction

4. Conclusion
Loss of mutualization : when there is no independence

- Example in insurance: natural catastrophes and portfolios with spatial correlations:

- Cyber risk: how to define proximity?
Contagion models with networks effects

- Multi-group SIR (Susceptible-Infected-Recovered) models with different subpopulations.

\[ B = (\beta_{i,j})_{1 \leq i,j \leq k} \] matrix of infection rates: \( \beta_{i,j} \) materializes how \( j \) contaminates \( i \).

- We also introduce a flexible framework to model the initial attacks that trigger the contagion.

Figure from Magal et al. (2018)
Multigroup compartmental epidemiological model

- Multi-group SIR: consider $K$ subpopulations: $1 \leq i \leq K$

\[
\frac{dS_i(t)}{dt} = - \left( \sum_{j=1}^{K} \beta_{i,j} I_j(t) \right) S_i(t)
\]

\[
\frac{dI_i(t)}{dt} = \left( \sum_{j=1}^{K} \beta_{i,j} I_j(t) \right) S_i(t) - \gamma_i I_i(t)
\]

\[
\frac{dR_i(t)}{dt} = \sum_{i=1}^{K} \gamma_i I_i(t).
\]

- $\mathcal{B}$ matrix of infection rate: $\beta_{i,j}$ materializes how $j$ contaminates $i$.
- Allows to introduce network effects.
Multigroup compartmental epidemiological model

- Multi-group SIR: consider $K$ subpopulations:

\[
\frac{dS_i(t)}{dt} = - \left( \alpha_i(t) + \sum_{j=1}^{K} \beta_{i,j}I_j(t) \right) S_i(t)
\]

\[
\frac{dI_i(t)}{dt} = \left( \alpha_i(t) + \sum_{j=1}^{K} \beta_{i,j}I_j(t) \right) S_i(t) - \gamma_i I_i(t)
\]

\[
\frac{dR_i(t)}{dt} = \sum_{i=1}^{K} \gamma_i I_i(t).
\]

- $\alpha_i(t)$ represents an intensity of attacks in class $i$.
- Example: single initial burst $\alpha_i(t) = \alpha 1_{0 \leq t < 1}$ for some $i$. 
Multigroup compartmental epidemiological model

- Multi-group SIR: consider $K$ subpopulations:

\[
\frac{dS_i(t)}{dt} = -\eta_i(t) \left( \alpha_i(t) + \sum_{j=1}^{K} \beta_{i,j} I_j(t) \right) S_i(t)
\]

\[
\frac{dI_i(t)}{dt} = \eta_i(t) \left( \alpha_i(t) + \sum_{j=1}^{K} \beta_{i,j} I_j(t) \right) S_i(t) - \gamma_i I_i(t)
\]

\[
\frac{dR_i(t)}{dt} = \sum_{i=1}^{K} \gamma_i l_i(t).
\]

- $\eta_i(t)$ represents how population $i$ tends to protect itself against the threat.
- Example: $\eta_i(t) = 1 - \lambda 1_{l_i(t) \geq s}$, or $\eta_i(t) = 1 - \lambda 1_{\sum_k l_k(t) \geq s}$
Some examples of comparisons that show the impact of the topology of the network

- Two classes of matrices $B$:
  - "Clustered" : the transmission is essentially intern to a class.
  - "Non-clustered" : the transmission is stronger from one class to another than within a given class.

→ the "Non-clustered" situation is worse, since the outbreak rapidly spreads from one class to any others.
Calibration of connections on OECD data

- Calibration of the model based on macroeconomic data: OECD data to identify the dependence/connectivity between some sectors of activity (more details in d’Oultremont, Lopez, Spoorenberg (2021) and Hillairet et al. (2021)).

- Contagion matrix

<table>
<thead>
<tr>
<th></th>
<th>Mining</th>
<th>Manufacturing</th>
<th>Energy</th>
<th>Construction</th>
<th>Services</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mining</td>
<td>0.0634</td>
<td>0.2927</td>
<td>0.0449</td>
<td>0.1427</td>
<td>0.1255</td>
<td>0.6692</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>0.0063</td>
<td>0.0527</td>
<td>0.0027</td>
<td>0.0108</td>
<td>0.0351</td>
<td>0.1076</td>
</tr>
<tr>
<td>Energy</td>
<td>0.0135</td>
<td>0.0370</td>
<td>0.0571</td>
<td>0.0150</td>
<td>0.0452</td>
<td>0.1679</td>
</tr>
<tr>
<td>Construction</td>
<td>0.0019</td>
<td>0.0068</td>
<td>0.0007</td>
<td>0.0141</td>
<td>0.0091</td>
<td>0.0326</td>
</tr>
<tr>
<td>Services</td>
<td>0.0003</td>
<td>0.0042</td>
<td>0.0004</td>
<td>0.0017</td>
<td>0.0161</td>
<td>0.0227</td>
</tr>
<tr>
<td>Total</td>
<td>0.0855</td>
<td>0.3934</td>
<td>0.1057</td>
<td>0.1844</td>
<td>0.2309</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3: Normalized Interaction matrix $B_0$.

- Calibration of a Wannacry-type scenario $B = \beta B_0$
- Disclaimer: this contagion matrix does not completely reflect the true connectivity between sectors.
Outline

1. Introduction

2. Accumulation of claims and contagion
   - Multi-group SIR model
   - Topology of the network

3. Calibration and applications
   - How to calibrate a Wannacry type scenario
   - Identifying vulnerabilities
   - Impact of the reaction

4. Conclusion
Calibration of the strength of the attack

- Difficult calibration of SIR model in cyber epidemic (few available data).
- What is a reasonable $\beta$, and what is the exposure $N$ in the global population?
- What do we know on Wannacry:
  - the total number of infected
  - the duration of the epidemic
  - the dynamics of the ransom payments (using bitcoins addresses used to pay the ransoms).
Comparison with a "biological" pandemic

- For Wannacry: no incubation period
- Parameter $\gamma$ is high, at least if the damages are visible: the infected are rapidly moved to quarantine
- The duration of the pandemic is short (Wannacry: 10 days approximately).
- $R_0$: small, here $R_0 \approx 1.04$.
- Damages (not taken into account in the diffusion model): the recovery may be much longer (several months or even years)
Example of epidemic dynamics of Wannacry type

Evolution of the proportion of infected - Attack on Mining.

The value of the peak for the Mining sector is at 70% (after 10 hours).
How can we use these models?

- **"Ranking" of sectors of activity**: one can identify which sector is more "systemic" than others in the sense that, if attacked, it will lead to a higher number of victims.

- **Quantifying the "peak"**: helps to identify how many "tech" assistance will be required by the policyholders during such a crisis. **Saturation risk**: if a large number of policyholders are simultaneously victims of the attack, the insurer may not be able to assist them all (causing then an increase of the costs).

- **Diversification**: design a portfolio that may resist to such contagious episode.

- **Identify the benefits of protection**:
  - of a given sector: protecting some key sectors may help to prevent the infection from spreading, even if this sector is not directly attacked.
  - from the reaction of the targets.
Measuring the vulnerability

- Evolution of the proportion of infected in each sector, depending on the bombing $\alpha_i$.
- Measuring the vulnerability of the different sectors

<table>
<thead>
<tr>
<th>Targeted sector</th>
<th>$\beta$</th>
<th>$\alpha$</th>
<th>Total infected</th>
<th>Peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mining</td>
<td>$1.845 \times 10^{-5}$</td>
<td>3.5</td>
<td>714’347</td>
<td>89’984</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>$1.845 \times 10^{-5}$</td>
<td>0.078</td>
<td>587’338</td>
<td>70’815</td>
</tr>
<tr>
<td>Energy</td>
<td>$1.845 \times 10^{-5}$</td>
<td>1.077</td>
<td>450’824</td>
<td>50’759</td>
</tr>
<tr>
<td>Services</td>
<td>$1.845 \times 10^{-5}$</td>
<td>0.0049</td>
<td>256’833</td>
<td>27’483</td>
</tr>
<tr>
<td>Construction</td>
<td>$1.845 \times 10^{-5}$</td>
<td>0.009</td>
<td>223’744</td>
<td>26’233</td>
</tr>
</tbody>
</table>

Comparison of the sectors through different attack scenarios. The sectors are ordered from the one leading to the highest epidemic, to the lowest.
Impact of the protection

- Measuring the impact of the protection \( \eta_i(t) = 1 - \lambda_1 \sum_{k} I_k(t) \geq s \):
  ratios \# victims if reaction / \# victims without reaction.

<table>
<thead>
<tr>
<th>( \lambda = 10% )</th>
<th>( s = 10'000 )</th>
<th>( s = 50'000 )</th>
<th>( s = 100'000 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Collateral</td>
<td>Total</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>94.60%</td>
<td>96.99%</td>
<td>95.82%</td>
</tr>
<tr>
<td>Energy</td>
<td>99.81%</td>
<td>99.98%</td>
<td>99.84%</td>
</tr>
<tr>
<td>Construction</td>
<td>98.51%</td>
<td>99.40%</td>
<td>98.87%</td>
</tr>
<tr>
<td>Services</td>
<td>73.10%</td>
<td>77.62%</td>
<td>80.40%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \lambda = 30% )</th>
<th>( s = 10'000 )</th>
<th>( s = 50'000 )</th>
<th>( s = 100'000 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Collateral</td>
<td>Total</td>
</tr>
<tr>
<td>Mining</td>
<td>99.39%</td>
<td>99.95%</td>
<td>99.49%</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>84.99%</td>
<td>91.44%</td>
<td>88.33%</td>
</tr>
<tr>
<td>Energy</td>
<td>99.42%</td>
<td>99.93%</td>
<td>99.52%</td>
</tr>
<tr>
<td>Construction</td>
<td>95.66%</td>
<td>98.24%</td>
<td>96.70%</td>
</tr>
<tr>
<td>Services</td>
<td>43.66%</td>
<td>51.37%</td>
<td>57.35%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \lambda = 50% )</th>
<th>( s = 10'000 )</th>
<th>( s = 50'000 )</th>
<th>( s = 100'000 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Collateral</td>
<td>Total</td>
</tr>
<tr>
<td>Mining</td>
<td>98.97%</td>
<td>99.90%</td>
<td>99.14%</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>76.87%</td>
<td>86.55%</td>
<td>81.92%</td>
</tr>
<tr>
<td>Energy</td>
<td>99.03%</td>
<td>99.88%</td>
<td>99.21%</td>
</tr>
<tr>
<td>Construction</td>
<td>92.99%</td>
<td>97.14%</td>
<td>94.66%</td>
</tr>
<tr>
<td>Services</td>
<td>30.04%</td>
<td>38.29%</td>
<td>45.65%</td>
</tr>
</tbody>
</table>
Illustration: simulation of a Wannacry-type scenario

- Simulation on a portfolio of 10 000 policyholders.
- We model the reaction of the policyholders (i.e. their capacity to protect themselves once they are informed on the ongoing threat).
- Different forms and types of responses (in blue fast reaction, in red medium reaction, in green slow reaction)

![Chart showing the evolution of the number of victims needing immediate assistance](chart.png)
Extension: estimation of the matrix $B_0$ from portfolio data

- Examples of models used in ecology: PLN models to describe vectors of count data:
  - $N = (N^{(1)}, ..., N^{(i)})$,
  - $N^{(i)}|Z \sim \mathcal{P}(\lambda_i \exp(Z^{(i)}))$,
  - where $Z$ is a latent vector with $Z \sim \mathcal{N}(0, \Sigma)$.

- Under some assumptions (one of them being the symmetry of $B_0$), one can make connections between $B_0$ and $\Sigma$.

- Illustration on public data:
Outline

1. Introduction

2. Accumulation of claims and contagion
   - Multi-group SIR model
   - Topology of the network

3. Calibration and applications
   - How to calibrate a Wannacry type scenario
   - Identifying vulnerabilities
   - Impact of the reaction

4. Conclusion
Conclusion

- Impact of cyber pandemic on a insurance portfolio: policyholders are exposed to the risk whose intensity is estimated at a global level.
- Flexible model with network effects.
- Calibration requires macro-level data.
- Add prevention and response module (reactivity in detecting the incident and to implement countermeasures).
- Help for an insurance company to quantify
  - how much should be the response capacity in order to sufficiently reduce the impact of a given scenario
  - how much is won if one increases the capacity of response (to fix the price of a partnership with a cyber security firm).
Conclusion: what about the cost?

- The cost of an attack is an **heavy-tailed** random variable.
- Extreme value theory: classification of distributions with respect to their **tail index** $\gamma$, which is a **nonlinear** characteristic of the distribution.
- **High tail index**: risk management through mean / variance optimization is not possible anymore.
- **Consequences**: if $\gamma$ is too high, the insurer will:
  - exclude the risk
  - or introduces very restrictive **limits** to the compensation.
Conclusion: Heavy-tailed distributions and heterogeneity

- Consider a population in which we have a mixture of two heavy tail distributions (indexes $\gamma_1$ and $\gamma_2$, with $\gamma_1 << \gamma_2$).

- Suppose that we do not have the ability to determine if a given policyholder belongs to population 1 or to population 2 (and, in fact, we do not even know if there are 2 or 3 or more subpopulations).

- From a statistical point of view, the estimated tail index for the total population will be close to $\max(\gamma_1, \gamma_2)$.

- If, on the other hand, one is able to identify these populations, one can offer better coverage for population 1, without deteriorating too much the situation for population 2.
References

- Joint Research Initiative "Cyber-insurance: actuarial modeling" : https://sites.google.com/view/cyber-actuarial/home?authuser=0

- On contagion scenarii:
References

- **On extreme events:**
  

- **On frequency modeling and clustering effect:**
  