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

Caroline Hillairet[1], Olivier Lopez[2]

[1] ENSAE Paris, [2] Sorbonne Université

Joint Research Initiative "Cyber insurance: actuarial modeling"
supported by Risk Fundation, AXA Research Fund

Webinaire Chaire PARI, 16 February 2022

RESEARCH INITIATIVE
Cyber Risk: Actuarial Modeling

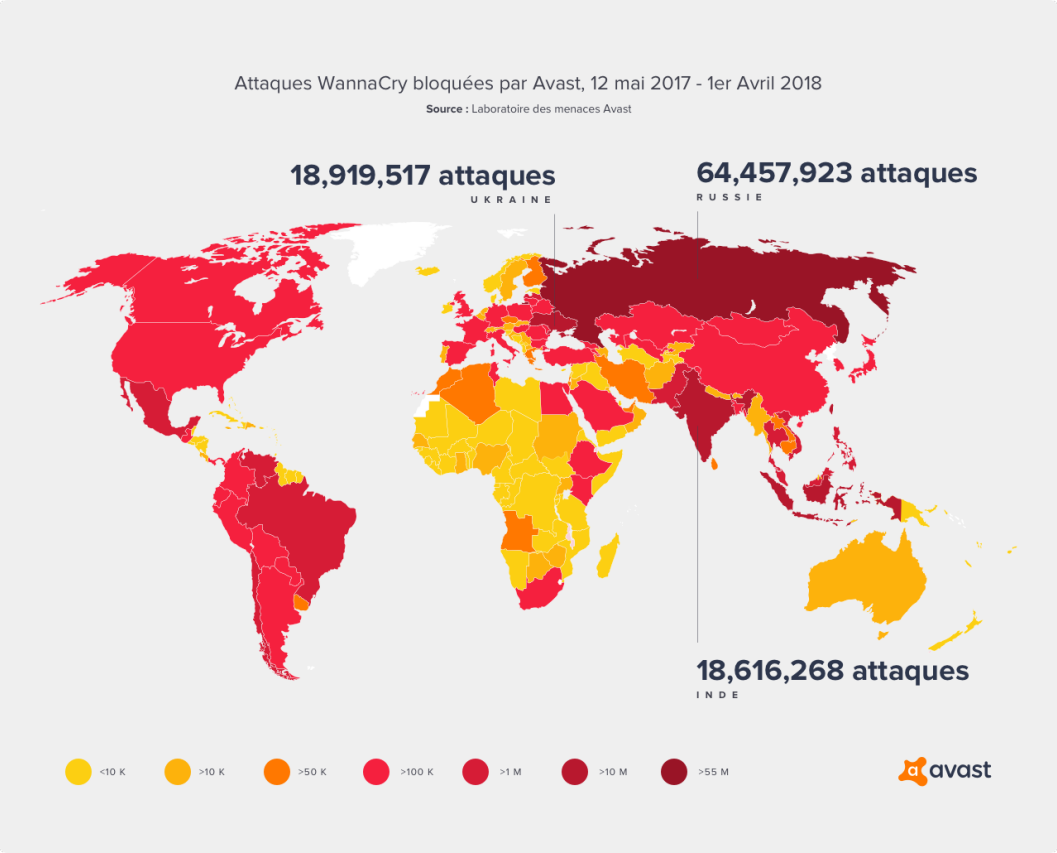
Cyber-risk

- 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.

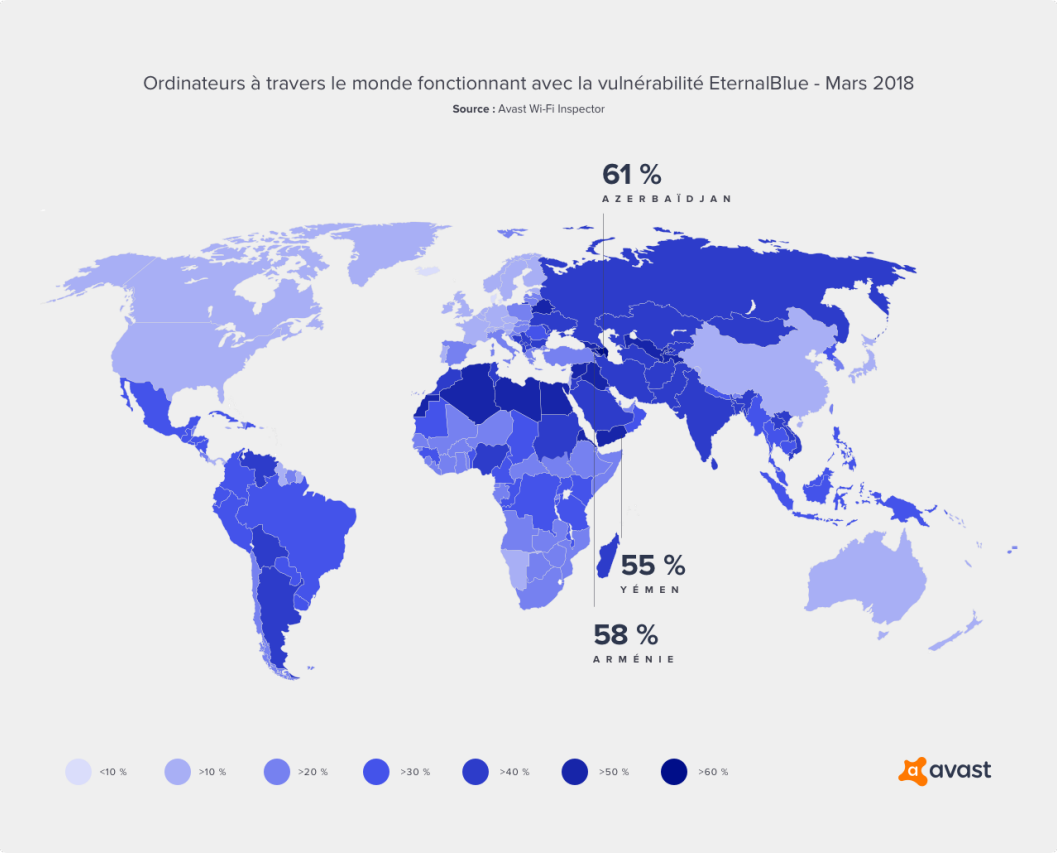
Wannacry

- Ransomware Wannacry : worldwide cyber attack in May 2017.
- Use the vulnerability "EternalBlue".
- Approximatively 200 000 infected computers across 150 countries over approximatively 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

The Colonial Pipeline Cyberattack | What to Know | Pipeline Resumes | $5 Million Ransom | DarkSide Says Shutting Down | Weaknesses in U.S. Cybersecurity

Cyberattack Forces a Shutdown of a Top U.S. Pipeline

The operator, Colonial Pipeline, said it had halted systems for its 5,500 miles of pipeline after being hit by a ransomware attack.

- "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 content 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

RESEARCH INITIATIVE
Cyber Risk: Actuarial Modeling

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.

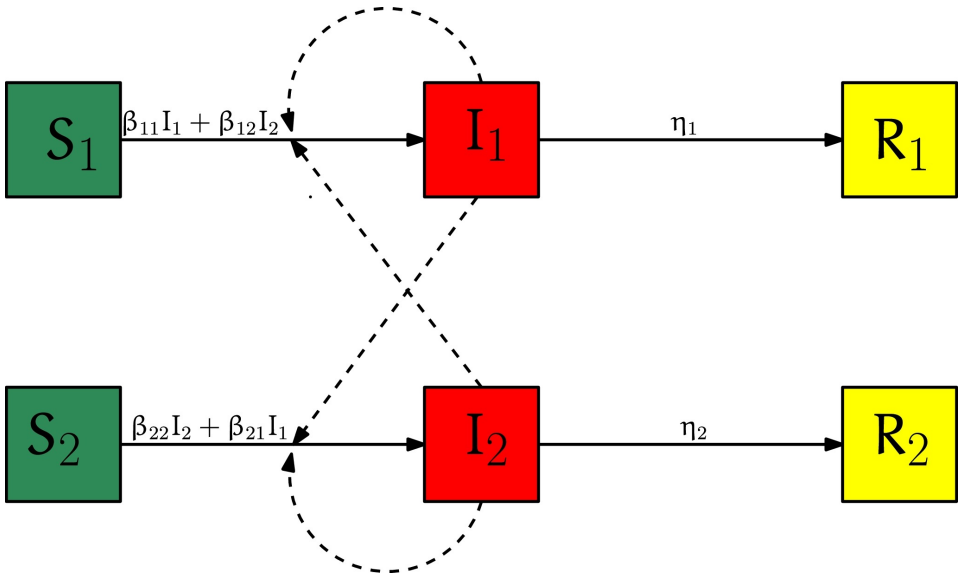


Figure from Magal et al. (2018)

- $\mathcal{B} = (\beta_{i,j})_{1 \leq i,j \leq k}$ 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.

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:

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

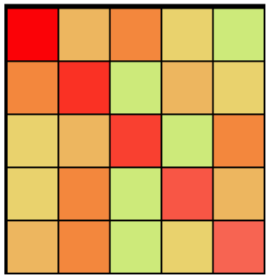
$$\frac{dI_i(t)}{dt} = \eta_i(t)\left(\alpha_i(t) + \sum_{j=1}^{K}\beta_{i,j}I_j(t)\right)S_i(t) - \gamma_i I_i(t)$$

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

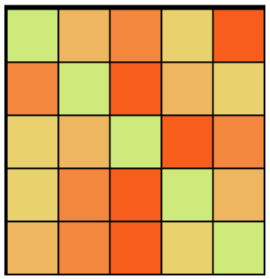
- $\eta_i(t)$ represents how population i tends to protect itself against the threat.
- Example: $\eta_i(t) = 1 - \lambda 1_{I_i(t)\geq s}$, or $\eta_i(t) = 1 - \lambda 1_{\sum_k I_k(t)\geq s}$

Some examples of comparisons that show the impact of the topology of the network

- Two classes of matrices $\mathcal{B}$:
  - "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.



$\rightarrow$ 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

	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

Table 3: Normalized Interaction matrix $\mathbf{B}_0$.

- Calibration of a Wannacry-type scenario $\mathcal{B} = \beta \mathcal{B}_0$
- Disclaimer: this contagion matrix does not completely reflect the true connectivity between sectors.

Outline

Calibration of the strength of the attack

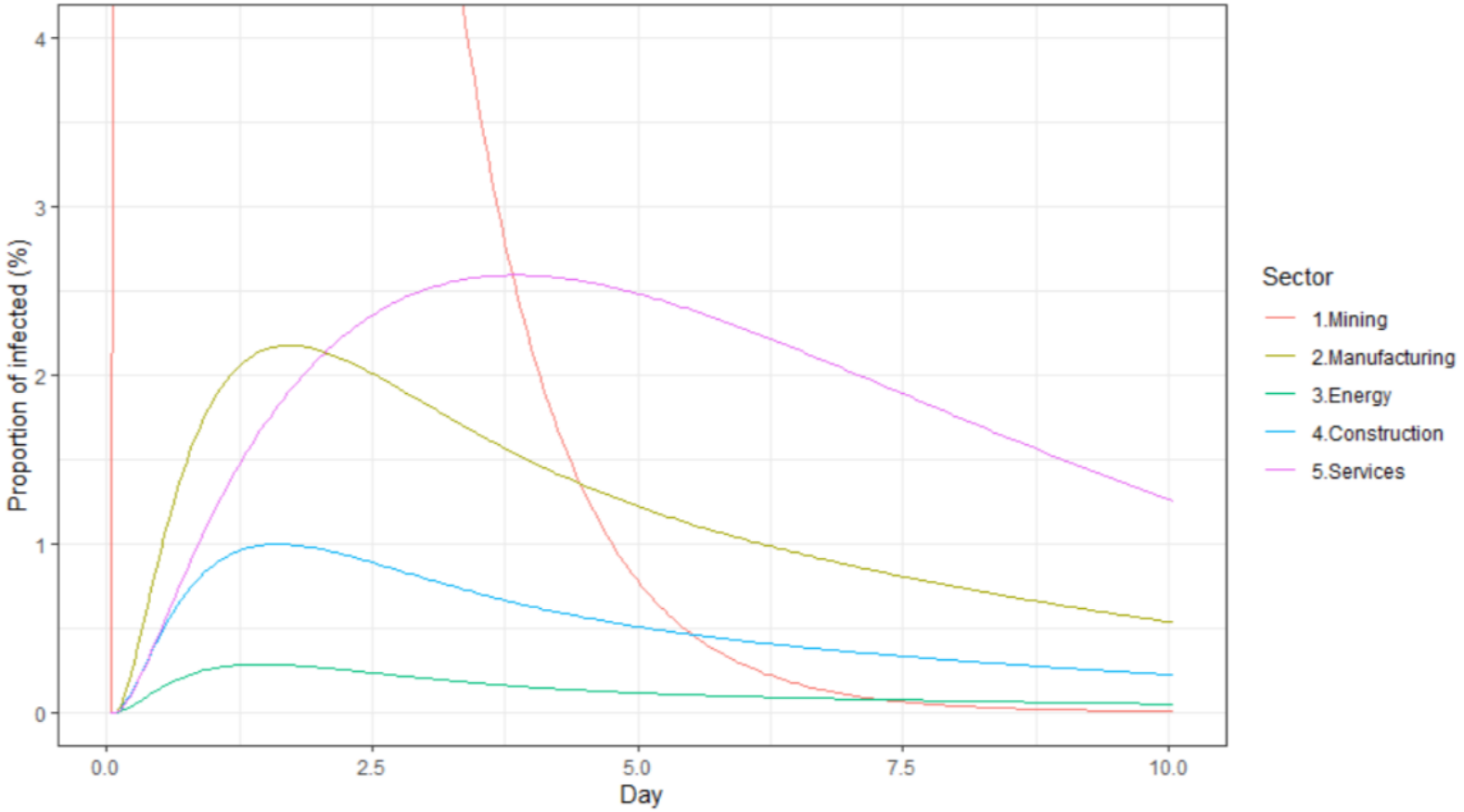
- Difficult calibration of SIR model in cyber epidemic (few available data).
- What is a reasonable β, and what is the exposure N in the global population?
- What do we know on Wannacry :
  - the total number of infected
  - the duration of the epidemic
  - the dynamics of the ransom payments (using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

Comparison with a "biological" pandemic

- For Wannacry: no incubation period
- Parameter γ 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.

RESEARCH INITIATIVE
Cyber Risk: Actuarial Modeling

Measuring the vulnerability

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

Targeted sector	β	α	Total infected	Peak
Mining	1.845×10^{-5}	3.5	714'347	89'984
Manufacturing	1.845×10^{-5}	0.078	587'338	70'815
Energy	1.845×10^{-5}	1.077	450'824	50'759
Services	1.845×10^{-5}	0.0049	256'833	27'483
Construction	1.845×10^{-5}	0.009	223'744	26'233

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

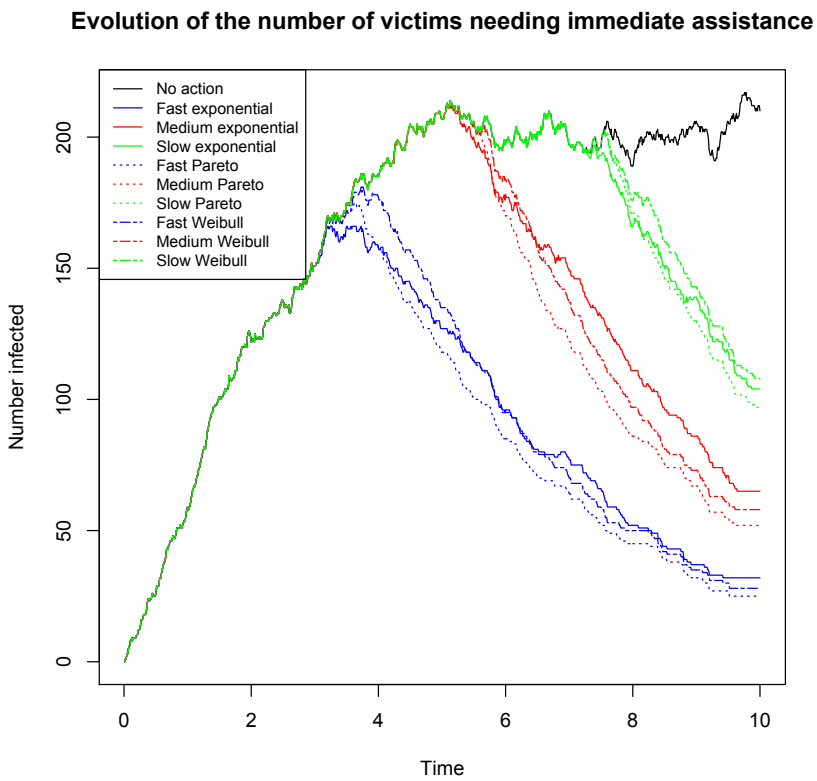
Impact of the protection

- 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.

$\lambda = 10\%$	$s = 10'000$		$s = 50'000$		$s = 100'000$	
	Total	Collateral	Total	Collateral	Total	Collateral
Mining	99.80%	99.99%	99,83%	99.99%	99.87%	99.99%
Manufacturing	94.60%	96.99%	95.82%	97.82%	97.09%	98.63%
Energy	99.81%	99.98%	99.84%	99.98%	99.88%	99.99%
Construction	98.51%	99.40%	98.87%	99.59%	99.18%	99.74%
Services	73.10%	77.62%	80.40%	84.36%	86.79%	90.05%
$\lambda = 30\%$	$s = 10'000$		$s = 50'000$		$s = 100'000$	
	Total	Collateral	Total	Collateral	Total	Collateral
Mining	99.39%	99.95%	99.49%	99.97%	99.61%	99.98%
Manufacturing	84.99%	91.44%	88.33%	93.78%	91.83%	96.07%
Energy	99.42%	99.93%	99.52%	99.95%	99.64%	99.97%
Construction	95.66%	98.24%	96.70%	98.79%	97.62%	99.23%
Services	43.66%	51.37%	57.35%	64.73%	69.97%	76.61%
$\lambda = 50\%$	$s = 10'000$		$s = 50'000$		$s = 100'000$	
	Total	Collateral	Total	Collateral	Total	Collateral
Mining	98.97%	99.90%	99.14%	99.93%	99.35%	99.96%
Manufacturing	76.87%	86.55%	81.92%	90.19%	87.25%	93.77%
Energy	99.03%	99.88%	99.21%	99.92%	99.40%	99.95%
Construction	92.99%	97.14%	94.66%	98.03%	96.14%	98.75%
Services	30.04%	38.29%	45.65%	54.04%	60.53%	68.54%

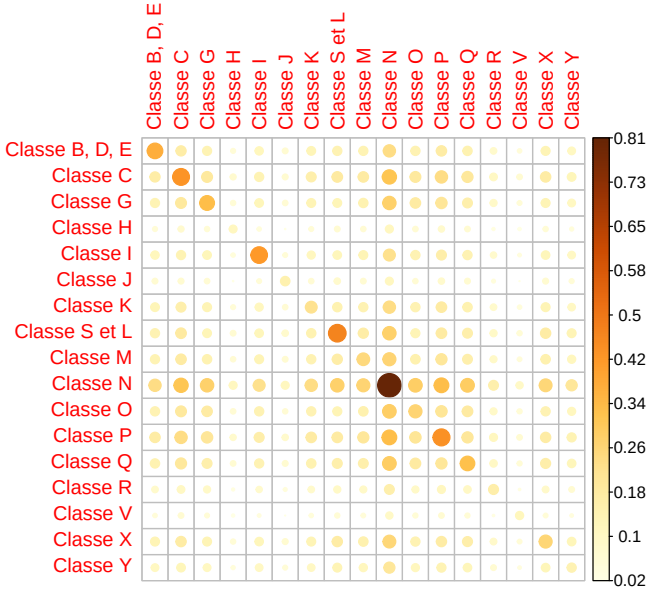
Illustration: simulation of a Wannacry-type scenario

- 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)



Extension: estimation of the matrix B_0 from portfolio data

- Examples of models used in ecology: PLN models to describe vectors of count data:
  - $N = (N^{(1)}, ..., N^{(j)})$,
  - $N^{(i)}|Z \sim \mathcal{P}(\lambda_i \exp(Z^{(i)}))$,
  - 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 connexions between B_0 and Σ.
- Illustration on public data:



Outline

RESEARCH INITIATIVE
Cyber Risk: Actuarial Modeling

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).

RESEARCH INITIATIVE
Cyber Risk: Actuarial Modeling

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 γ, which is a nonlinear characteristic of the distribution.
- High tail index: risk management through mean / variance optimization is not possible anymore.
- Consequences: if γ is too high, the insurer will:
  - exclude the risk
  - or introduces very restrictive limits to the compensation.

RESEARCH INITIATIVE
Cyber Risk: Actuarial Modeling

Conclusion: Heavy-tailed distributions and heterogeneity

- Consider a population in which we have a mixture of two heavy tail distributions (indexes γ_1 and γ_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:

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

Hillairet C., Lopez O., D'Oultremont L., Spoorenberg B. (2022) **Cyber contagion: impact of the network structure on the losses of an insurance portfolio** Preprint

RESEARCH INITIATIVE Cyber Risk: Actuarial Modeling

References

- On extreme events:

Farkas, S., Lopez, O., Thomas, M., (2021) **Cyber claims analysis through Generalized Pareto Regression Trees with applications to insurance pricing and reserving.**, in *Insurance: Mathematics and Economics, Elsevier, vol. 98(C), pages 92-105*

- On frequency modeling and clustering effect:

Bessy-Roland Y., Boumezoued A., Hillairet C., (2020) **Multivariate Hawkes process for Cyber Risk Insurance** in *Annals of Actuarial Science>Volume 15 Issue 1*