

Introduction aux processus stochastiques

Projet 1 : Analyse de la propagation d'un virus au sein d'un réseau

Maxime GOFFART
180521

Olivier JORIS
182113

Année académique 2019 - 2020

Table des matières

1	Introduction	3
2	Structure du programme	3
3	Etude du modèle exact	3
3.1	Question 1	3
3.2	Question 2	4
3.3	Question 3	4
3.3.1	Avec la matrice W_{lin}	4
3.3.2	Avec la matrice W_{full}	5
3.4	Question 4	6
4	Etude sur base de simulations	7
4.1	Question 1	7
4.2	Question 2	7
4.3	Question 3	9
4.4	Question 4	10
4.4.1	Comparaison des différentes mesures	14

1 Introduction

Les processus stochastiques permettent d'étudier des phénomènes aléatoires dans divers secteurs : l'économie, la climatologie, la météorologie, la biologie, ...

En particulier dans ce projet, il nous a été demandé d'étudier un phénomène d'actualité : la propagation d'un virus au sein d'un réseau, pouvant être modélisée à l'aide d'une chaîne de Markov. Ainsi, ce projet nous a permis d'appliquer les concepts vus au cours sur un exemple concret et d'actualité.

2 Structure du programme

Nous avons décidé de réaliser le projet en Python après comparaison des avantages et inconvénients des différents langages proposés. Notre code source se compose des fichiers suivants¹ :

- `exact_model.py` : module principal pour l'étude du modèle exact à l'aide de la matrice de transition (section 1 de l'énoncé).
- `simulations_study.py` : module principal pour l'étude sur base de simulations (section 2 de l'énoncé).
- `virus_spread_model.py` : module lié à la propagation du virus au sein de la population.
- `file_management.py` : module qui permet de charger la matrice d'adjacence du graphe W sous forme dense.
- `graphics_generator.py` : module qui permet d'afficher le graphique des proportions des individus.
- `states_manipulator.py` : module qui permet de manipuler les états d'une chaîne de Markov.
- `transition_matrix.py` : module qui permet de manipuler la matrice de transition d'une chaîne de Markov.

Nous avons utilisé différentes bibliothèques externes². En voici la liste :

- `Random` pour la génération de nombres pseudo-aléatoires.
- `Numpy` pour le produit matriciel et la puissance de matrices.
- `Matplotlib` pour l'affichage des graphiques.

3 Etude du modèle exact

3.1 Question 1

Le modèle proposé dans l'énoncé est bien un processus de Markov en temps discret caractérisé par ses 3^N états³. Les états de cette chaîne sont caractérisés par la suite de

1. Chaque méthode est décrite dans le module dans lequel elle est définie.
2. Aucune n'est spécifique aux chaînes de Markov comme mentionné sur eCampus.
3. N étant la taille de la population.

longueur N des catégories⁴ auxquelles appartiennent les individus⁵ à l'instant t . Par exemple : pour $N = 3$, l'état "'S' 'I' 'I'" représente le fait que le premier individu est susceptible d'être infecté et que les deux derniers sont infectés.

Les probabilités de transitions d'un état à un potentiel état de l'instant suivant de la chaîne dépendent à la fois de chacune des catégories des individus à l'instant initial et de leurs interactions avec des personnes infectées⁶.

Les états de cette chaîne qui sont uniquement composés d'individus de la catégorie 'R' ou⁷ de la catégorie 'S' sont absorbants car la propagation du virus n'est plus possible s'il n'y a plus d'infectés. Il en va de même pour les états dans lesquels les infectés n'ont de contact avec personne⁸. Cette chaîne n'est donc ni irréductible, ni régulière, ni périodique.

3.2 Question 2

Soit X une variable aléatoire, qui suit une distribution géométrique, représentant la guérison d'un individu une fois infecté. On a :

$$X \sim \mathcal{G}(\mu)$$

avec μ la probabilité de guérison.

Le temps moyen nécessaire à un individu pour guérir une fois infecté est donné par l'espérance de X . On a :

$$E(X) = \frac{1}{\mu}$$

3.3 Question 3

Pour répondre à cette question, le programme peut-être lancé à l'aide de la commande suivante :

```
$ python3 exact_model.py matrixSize fileName
```

où la première option spécifie la taille de la matrice W utilisée et la deuxième représente le nom du fichier dans lequel cette matrice est encodée.

3.3.1 Avec la matrice W_{lin}

Ainsi, en lançant le programme avec cette commande :

```
$ python3 exact_model.py 6 6x6_lin.txt
```

On obtient ce graphique répondant à la première partie de la question avec les paramètres μ et β fixés aux valeurs suggérées :

-
- 4. S, R ou I.
 - 5. Les individus étant indexés de 1 à N .
 - 6. Modélisées par le graphe W .
 - 7. Il s'agit d'un "ou" inclusif.
 - 8. Cela correspond à une ligne remplie de 0 dans le graphe W .

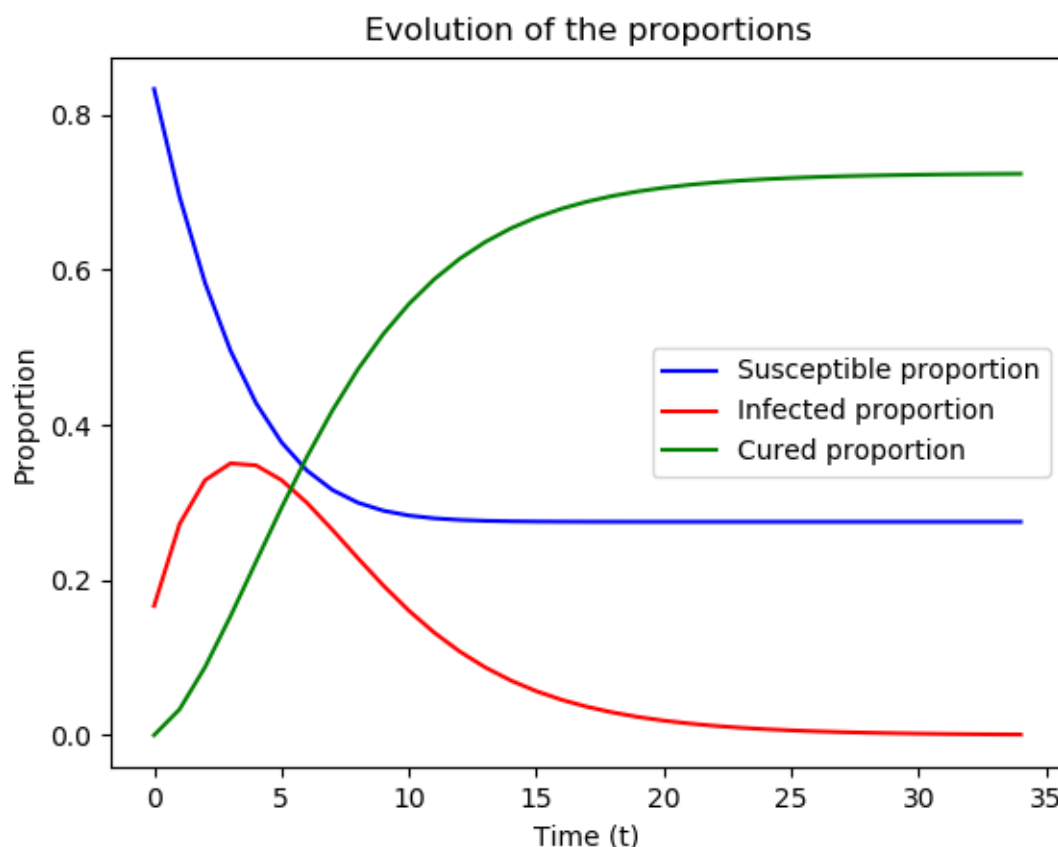


FIGURE 1 – Evolution des proportions d’individus dans chacune des catégories au cours du temps suivant un modèle représenté par une matrice W_{lin} .

Sur la figure 1, on observe le fait que la proportion d’individus infectés est décroissante après une partie croissante jusque $t = 3$. En effet, les individus ayant peu de contacts les uns avec les autres, la probabilité qu’un individu soit infecté est faible. De plus, les individus initialement infectés finissent par guérir. On observe, également, qu’une partie de la population reste "susceptible" d’obtenir le virus : elle ne l’a jamais contracté car les relations entre les individus sont limitées.

3.3.2 Avec la matrice W_{full}

De la même façon en lançant le programme à l’aide de cette commande :

```
$ python3 exact_model.py 6 6x6_full.txt
```

On obtient ce graphique répondant à la deuxième partie de la question avec les paramètres μ et β fixés aux valeurs suggérées :

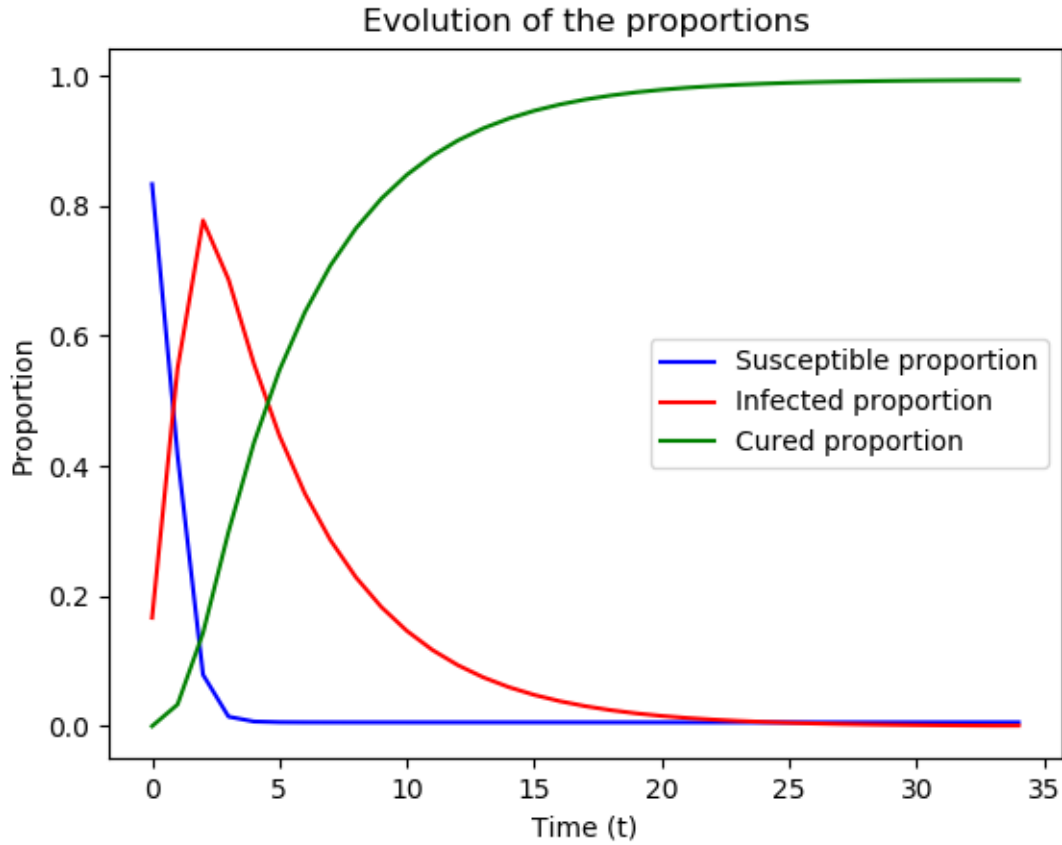


FIGURE 2 – Evolution des proportions d’individus dans chacune des catégorie au cours du temps suivant un modèle représenté par une matrice W_{full} .

Sur la figure 2, on observe, à l’inverse de la situation précédente, que la population connaît un nombre important d’infectés et très rapidement car les gens sont tous en contact. On observe un pic en $t = 2$ avec une proportion de $\pm 77\%$ d’infectés.

Au temps $t = 4$, on peut voir que la proportion d’individus susceptibles est proche de 0, c’est-à-dire que presque toute la population a été infectée à un instant inférieur ou égal à celui-ci. En effet, les individus ayant tous des contacts les uns avec les autres, le virus ne fait que se propager au sein de la population jusqu’à ce que celle-ci soit entièrement infectée.

3.4 Question 4

Lors de l’exécution des deux précédentes commandes, le programme nous retourne également le temps nécessaire à la disparition du virus dans la population. Nous avons obtenu 34 dans le cas où les contacts entre les gens sont limités⁹ et 33 dans l’autre cas¹⁰ avec une précision de 0.001. Nous n’avons pas accès aux états directement dans notre

9. Celui qui utilise la matrice W_{lin} .

10. Celui qui utilise la matrice W_{full} .

programme donc l'algorithme s'exécute jusqu'à atteindre 0 comme proportion d'individus infectés à une certaine précision près. Cette précision peut être modifiée dans le fichier `virus_spread_model.py` en modifiant la "constante" `PRECISION`.

Deux autres exécutions avec une précision égale à l'épsilon machine sur 32 bits ont donné 74 avec la matrice W_{lin} et 73 avec l'autre matrice, W_{full} .

On observe que le premier modèle¹¹ nécessite un temps légèrement supérieur au second avant la disparition du virus. Ceci s'explique par le fait que le virus se répand moins rapidement¹² donc les gens qui sont infectés transmettent le virus à moins d'individus mais de nouvelles infections peuvent survenir plus tard et retardent donc la guérison de la population ce qui mène à la disparition du virus. Comme mentionné précédemment, une partie de la population n'est pas infectée par le virus.

A l'inverse, dans le deuxième cas¹³, le temps avant la disparition du virus est un peu plus court car l'entière de la population est rapidement atteinte donc on peut très vite entrer dans une phase dans laquelle il faut guérir tout le monde.

Cette petite différence de temps entre les 2 est contre intuitive. On s'attendrait à avoir un temps, dans le cas où les relations entre les individus sont représentées par W_{lin} , inférieur au temps dans le cas où les relations sont modélisées par W_{full} .

Comment expliquer cette erreur ? Une source d'erreurs potentielles est que les probabilités deviennent très faibles car on utilise des puissances d'une matrice remplie de nombres compris entre 0 et 1 dès lors des erreurs de calculs peuvent survenir.

4 Etude sur base de simulations

4.1 Question 1

L'hypothèse d'indépendance posée par les auteurs de l'article n'est en général pas vérifiée car la probabilité qu'un individu se trouve dans un état à un instant t dépend en partie de son interaction avec les autres personnes de la population et de leur état à un instant précédent. Ainsi, si un individu dans un état *susceptible* a des contacts avec des individus dans un état *infectieux*, alors son état à un instant $t' > t$ sera impacté par ces derniers. L'indépendance n'est donc vérifiée que dans le cas où un individu n'a de contact avec personne.

4.2 Question 2

Pour répondre à cette question, le programme peut-être lancé à l'aide de la commande suivante :

```
1 $ python3 simulations_study.py populationSize fileName
```

11. Celui dont la matrice d'adjacence est W_{lin} .

12. Car les gens sont moins en contact les uns avec les autres.

13. Celui dont la matrice d'adjacence est W_{full} .

où la première option spécifie la taille de la matrice W utilisée et la deuxième option correspond au nom du fichier dans lequel cette matrice est encodée.

Ainsi, en lançant le programme avec cette commande :

```
$ python3 simulations_study.py 6 6x6_lin.txt
```

Le programme demande à l'utilisateur les paramètres que celui-ci souhaite utiliser pour les simulations, on choisit ainsi les valeurs suggérées dans l'énoncé. Ensuite, le script retourne le graphique de la figure 3 ainsi que le temps moyen de disparition des individus infectieux y correspondant¹⁴. On observe ainsi que l'étude sur base du modèle exact et celle sur base de simulations se ressemblent pour le cas où la matrice d'adjacence correspond à W_{lin} . Les différences observées peuvent être justifiées par le fait que les simulations n'utilisent pas la matrice de transition donc la précision est plus faible par rapport au modèle exact.

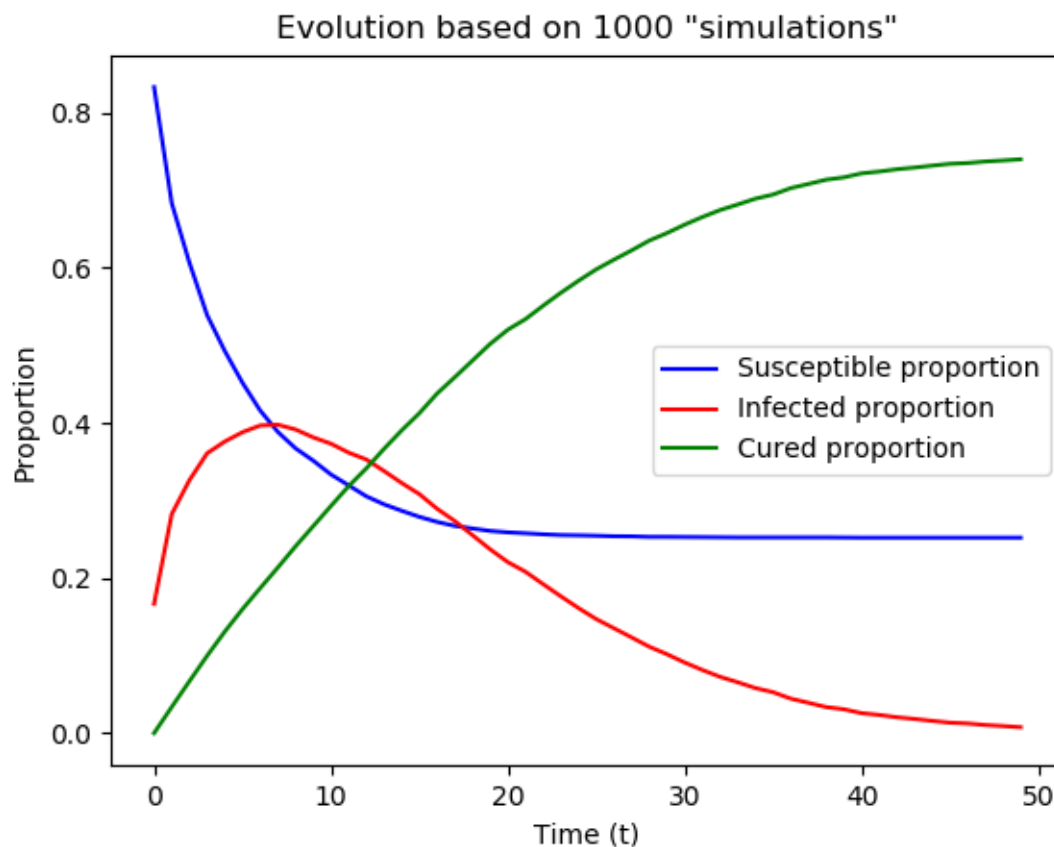


FIGURE 3 – Evolution des proportions d'individus dans chacune des catégories au cours du temps suivant un modèle représenté par une matrice W_{lin} de taille 6×6 .

De la même façon en lançant le programme à l'aide de cette commande¹⁵

14. Etant ici d'environ 45 unités.

15. Et en choisissant les paramètres suggérés dans l'énoncé.


```
1 $ python3 simulations_study.py 6 6x6_full.txt
```

On obtient le graphique de la figure 4 ainsi que le temps moyen de disparition des individus infectieux y correspondant¹⁶. On observe que l'étude sur base du modèle exact et celle sur base de simulations se ressemblent également dans ce cas de figure¹⁷. On peut donc penser que cette simulation approxime de manière satisfaisante la réalité et permettrait ainsi de modéliser cette même simulation avec une population plus nombreuse¹⁸.

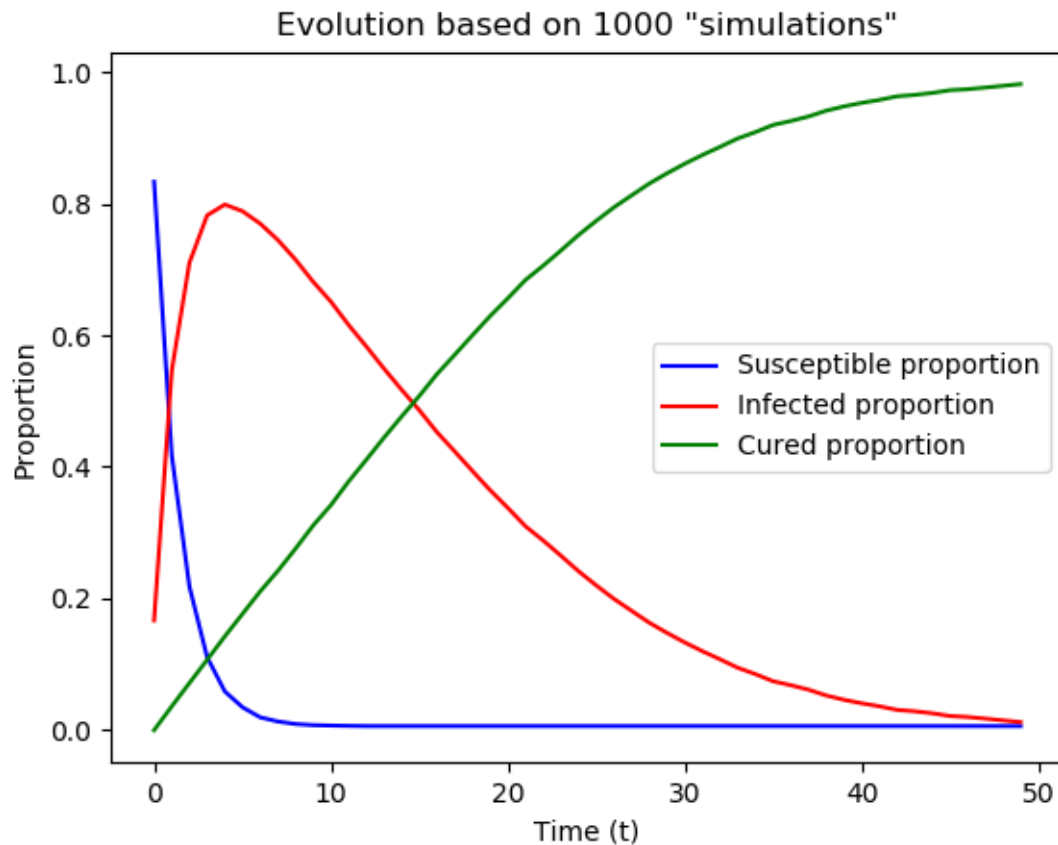


FIGURE 4 – Evolution des proportions d'individus dans chacune des catégories au cours du temps suivant un modèle représenté par une matrice W_{full} de taille 6×6 .

4.3 Question 3

En lançant le programme à l'aide de la commande suivante¹⁹

```
1 $ python3 simulations_study.py 2000 Wbig_dense.txt
```

16. Valant ici environ 50 unités.

17. Comme précédemment, les différences peuvent s'expliquer par le fait que les simulations n'utilisent pas la matrice de transition donc la précision est plus faible.

18. En prenant un nombre de simulations suffisamment grand pour que les calculs soient stables.

19. En choisissant un nombre suffisant de simulations pour avoir des résultats stables.

Le programme retourne le graphique de la figure 5 ainsi que le temps moyen de disparition des individus infectieux y correspondant²⁰ avec les probabilités β et μ fixées aux valeurs suggérées.

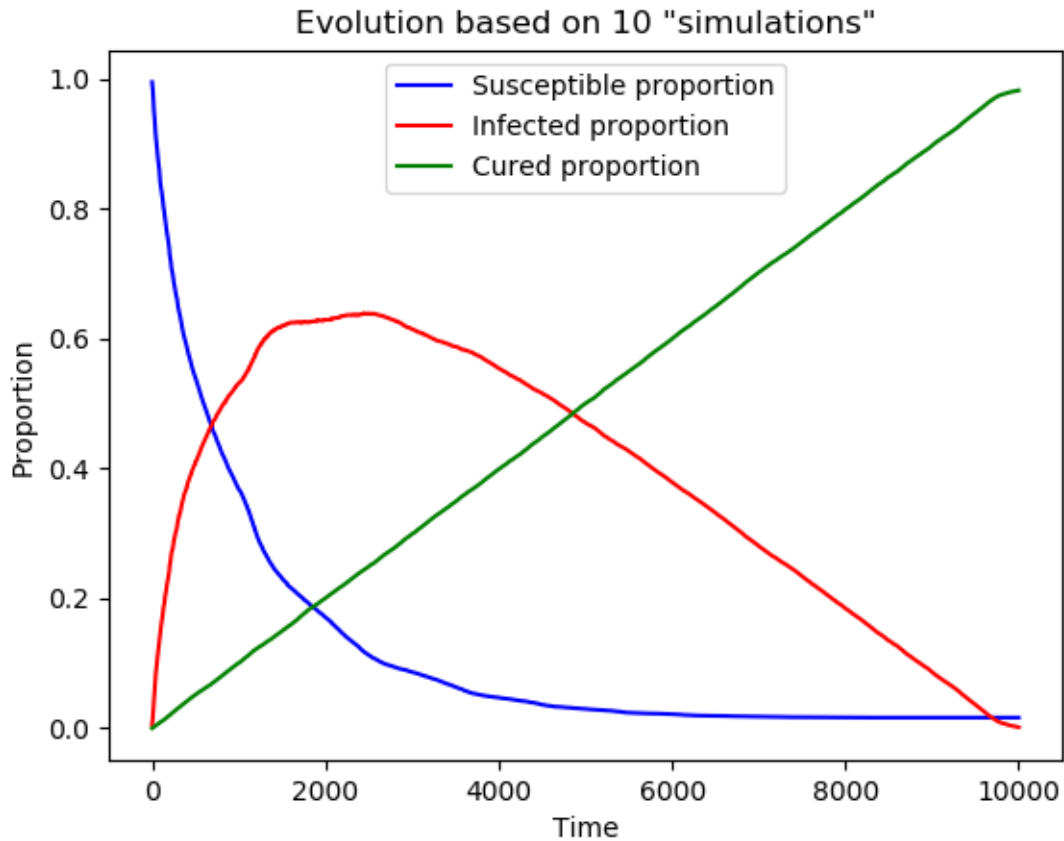


FIGURE 5 – Evolution des proportions d'individus dans chacune des catégories au cours du temps suivant un modèle représenté par une matrice d'adjacence W_{big} de taille 2000×2000 générée aléatoirement selon un modèle "scale-free" quand aucune mesure de sécurité n'est prise.

4.4 Question 4

Si le graphique de la figure 5 représente l'évolution de l'épidémie dans le cas où aucune mesure n'est prise. Etudions l'impact des mesures suivantes :

(a) Réduire la probabilité de transmission de la maladie via différents moyens revient à diminuer la probabilité de contamination β . Ainsi, en réduisant la probabilité β à 0.2, on obtient le graphique de la figure 6 qui permet de visualiser que cette mesure permet

20. Etant ici d'environ 10 000 unités.

d'avoir une proportion de la population qui ne se voit pas infectée par le virus. De plus, le temps de disparition de ce dernier diminue²¹.

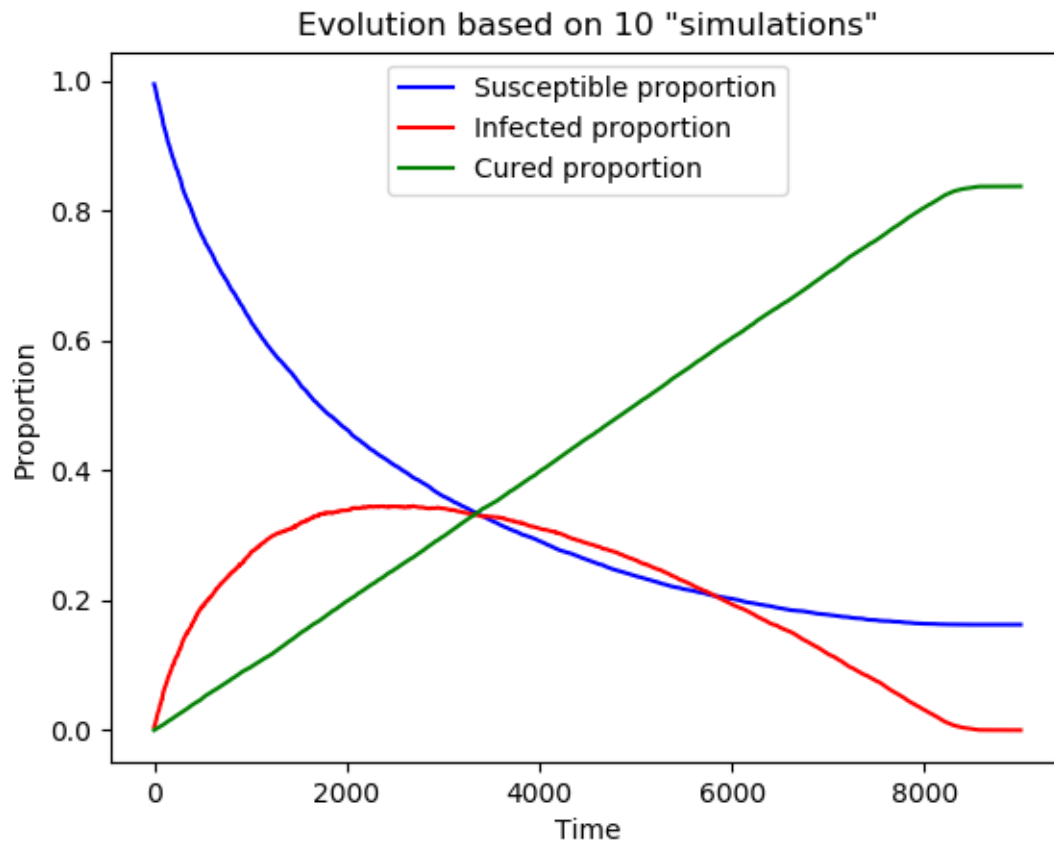


FIGURE 6 – Evolution des proportions d'individus dans chacune des catégories au cours du temps suivant un modèle représenté par une matrice d'adjacence W_{big} de taille 2000×2000 générée aléatoirement selon un modèle "scale-free" avec une probabilité de transmission de la maladie réduite : $\beta = 0.2$.

(b) Réduire les interactions entre les individus revient à diminuer le nombre de 1 dans la matrice W_{big} . En imposant un maximum d'interactions égal à 6 par personne par rapport à la matrice initiale, on obtient le graphique de la figure 7 montrant que cette mesure permet également d'accélérer la stabilisation de la situation²² et de réduire la proportion d'infectés.

21. Valant ici environ 8300

22. Le temps de stabilisation valant ici envrion 8000.

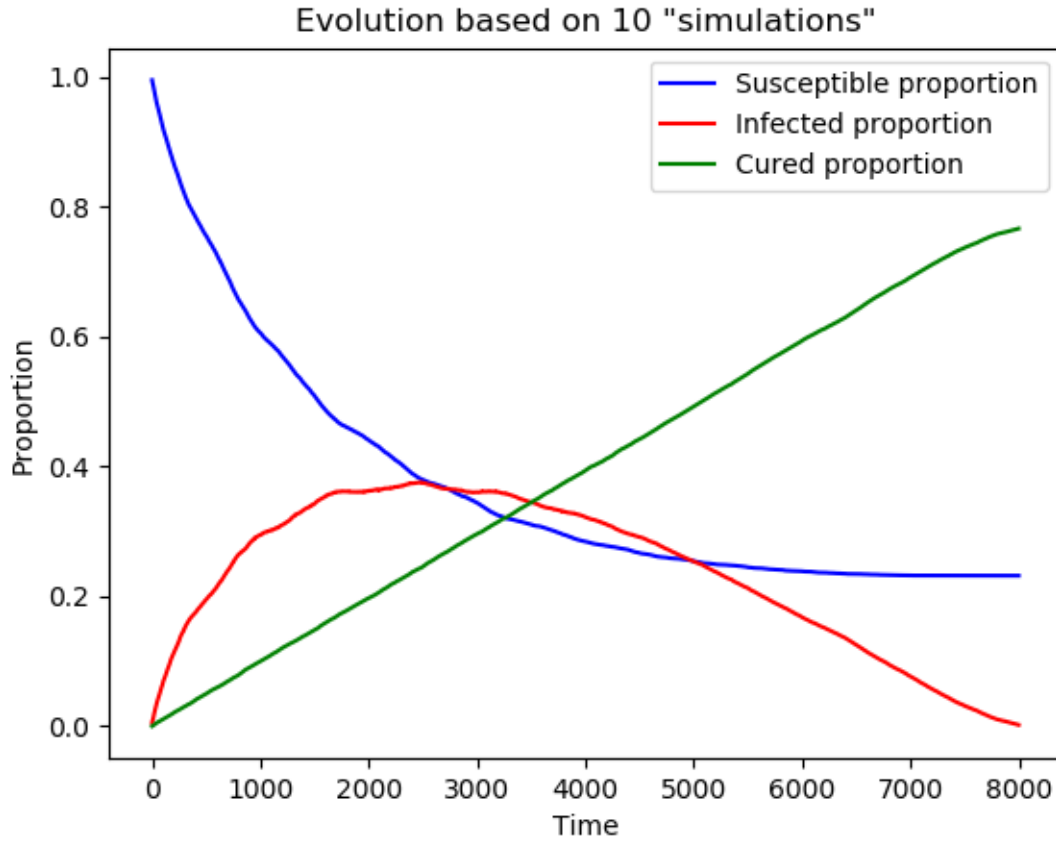


FIGURE 7 – Evolution des proportions d'individus dans chacune des catégories au cours du temps suivant un modèle représenté par une matrice d'adjacence W_{big} de taille 2000×2000 générée aléatoirement selon un modèle "scale-free" avec des mesures de confinement prises telles qu'une personne a au maximum 6 interactions physiques avec d'autres personnes du modèle initial.

(c) Vacciner un certain pourcentage de la population correspond à ajouter cette proportion d'individus dans l'état 'R' à l'instant $t = 0$. Le graphique de la figure 8 représente la situation où 30% de la population est vaccinée. Celui-ci permet également d'avoir une proportion d'individus qui n'est jamais infectée et de réduire le temps de disparition du virus²³.

23. Le temps de stabilisation de la situation passant ici à environ 6000.

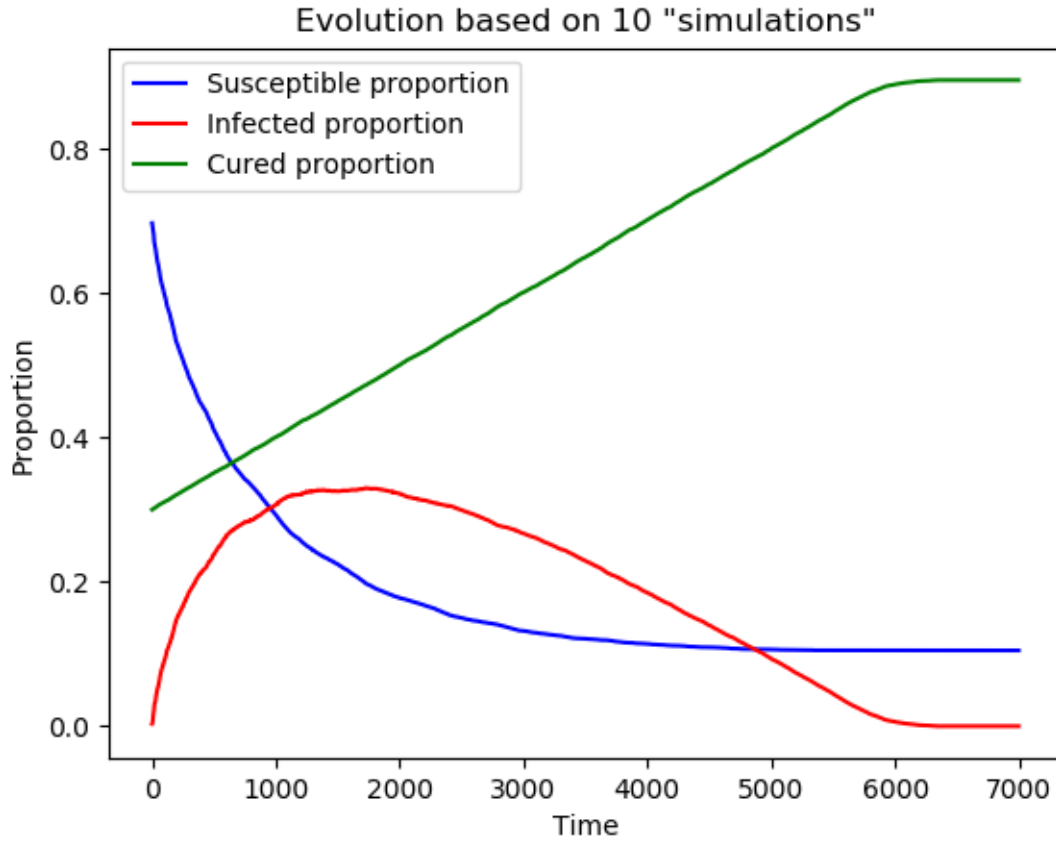


FIGURE 8 – Evolution des proportions d’individus dans chacune des catégories au cours du temps suivant un modèle représenté par une matrice d’adjacence W_{big} de taille 2000×2000 générée aléatoirement selon un modèle "scale-free" avec une proportion de 30% d’individus vaccinés.

(d) Traiter les patients avec un médicament qui permettrait d’accélérer la guérison revient à augmenter la probabilité de guérison μ . En augmentant la probabilité de guérison à 0.5, on voit sur la figure 9 que la principale amélioration est la vitesse de disparition du virus²⁴. On remarque également qu’une proportion d’individus ne se fait pas infecter.

²⁴. Le temps de disparition du virus étant ici égal à environ 3600.

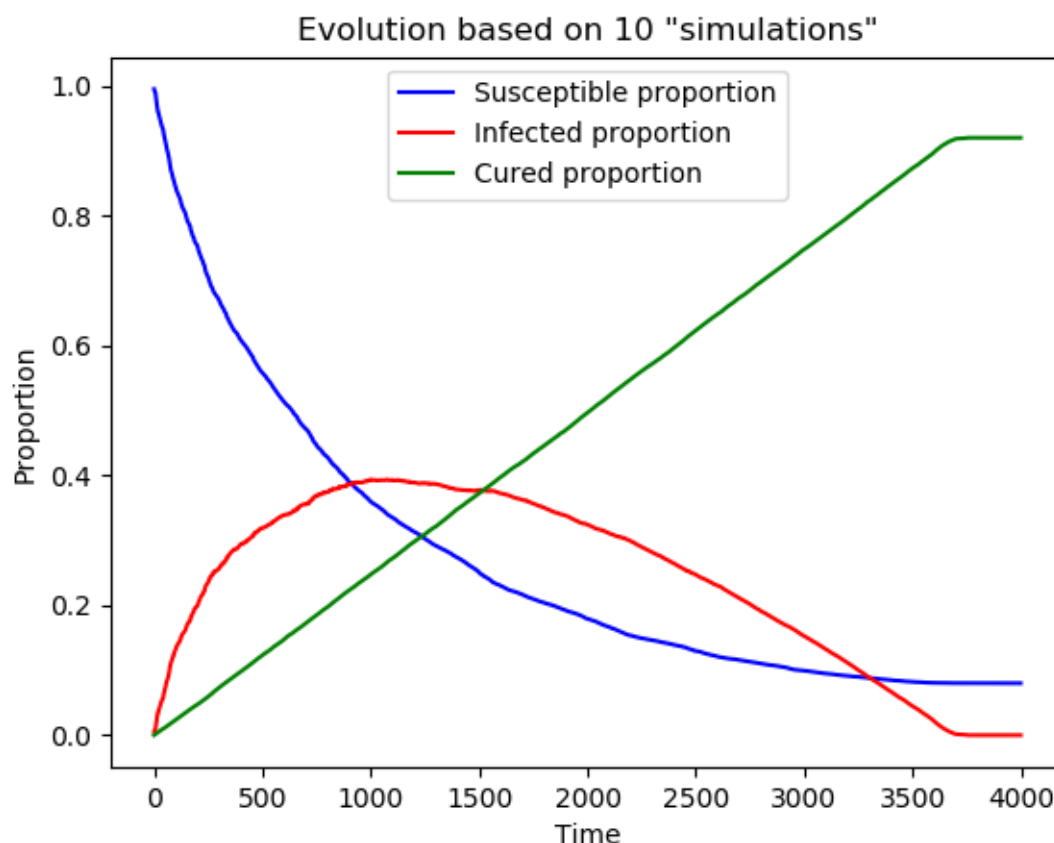


FIGURE 9 – Evolution des proportions d’individus dans chacune des catégories au cours du temps suivant un modèle représenté par une matrice d’adjacence W_{big} de taille 2000×2000 générée aléatoirement selon un modèle "scale-free" avec une accélération de guérison : $\mu = 0.5$.

4.4.1 Comparaison des différentes mesures

Si on fixe un nombre de lits disponibles pour une hospitalisation à 13, soit 6,5 lits pour 1000 habitants²⁵ et que l’on suppose que 5% des personnes infectées ont besoin d’être hospitalisées. Cela revient ainsi à veiller à avoir une proportion d’infectés ne dépassant pas $\frac{13}{2000 \times 0,05} = 0.13$. Voici les valeurs minimales des différentes mesures à prendre :

(a) Pour cette mesure, il faut arriver à réduire la probabilité de transmission à 0.09 dans quel cas on obtient le graphique de la figure 10.

25. Source : https://www.indexmundi.com/fr/belgique/lits_d_hopitaux_par_habitant.html.

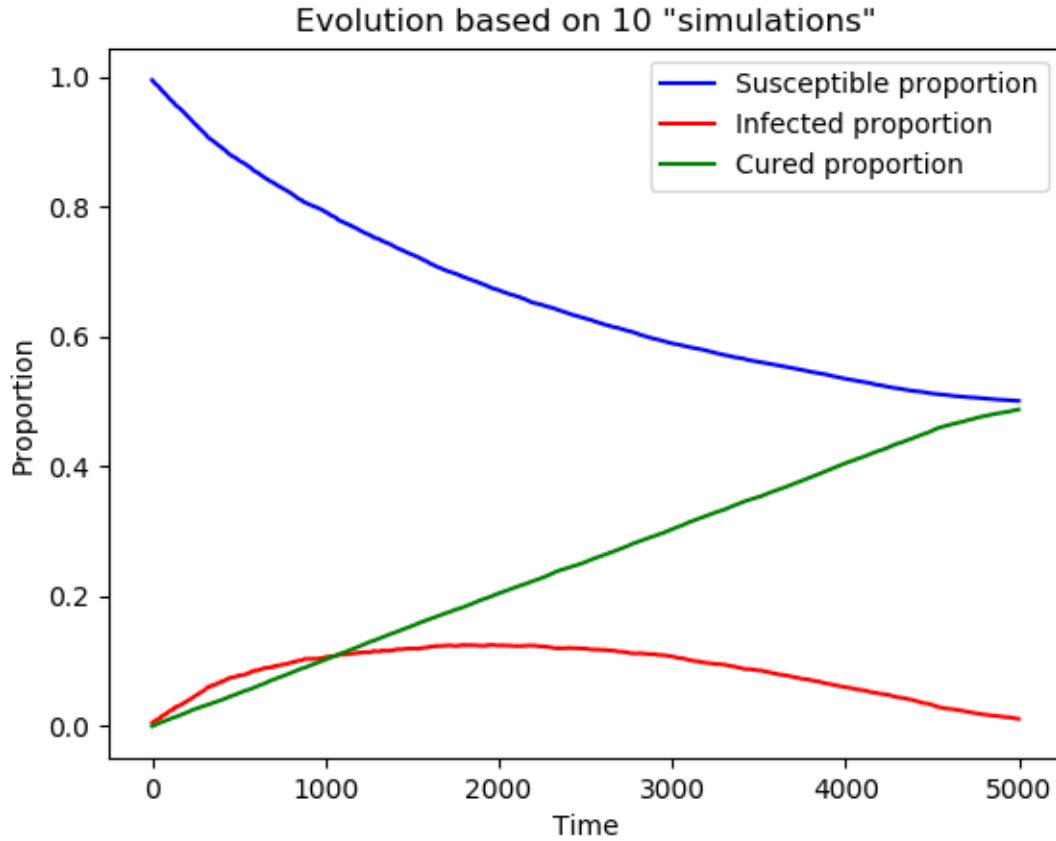


FIGURE 10 – Evolution des proportions d'individus dans chacune des catégories au cours du temps suivant un modèle représenté par une matrice d'adjacence W_{big} de taille 2000×2000 générée aléatoirement selon un modèle "scale-free" avec une probabilité de transmission de la maladie réduite : $\beta = 0.09$.

(b) Il faut ici réduire les contacts entre individus de façon à avoir maximum 3 contacts par individus²⁶ et le résultat correspondant est représenté sur le graphique de la figure 11.

26. En partant de la matrice W initiale, c'est-à-dire que si une personne avait moins de 3 interactions de nouvelles ne sont pas créées.

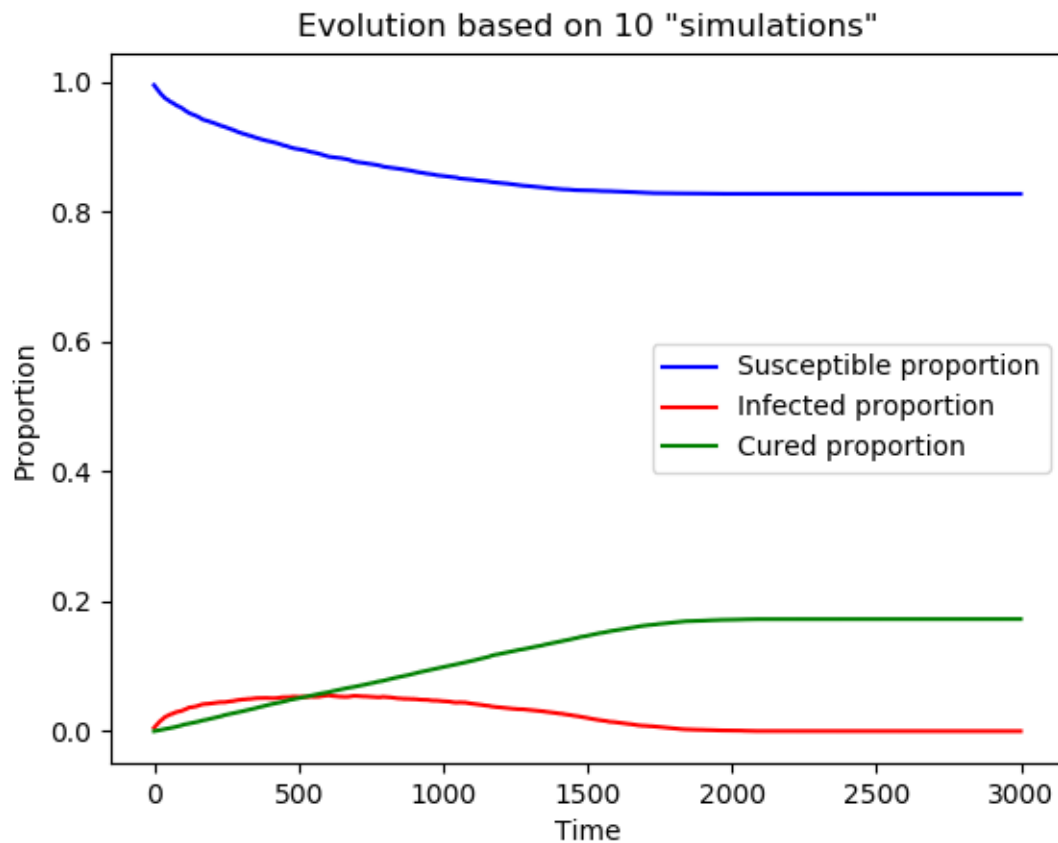


FIGURE 11 – Evolution des proportions d'individus dans chacune des catégories au cours du temps suivant un modèle représenté par une matrice d'adjacence W_{big} de taille 2000×2000 générée aléatoirement selon un modèle "scale-free" avec des mesures de confinement prises telles qu'une personne a au maximum 3 interactions physiques avec d'autres personnes du modèle initial.

(c) Il faut qu'initialement environ 55% de la population soit vaccinée pour respecter la limite de lits d'hôpitaux disponibles, le graphique de la figure 12 représente cette situation.

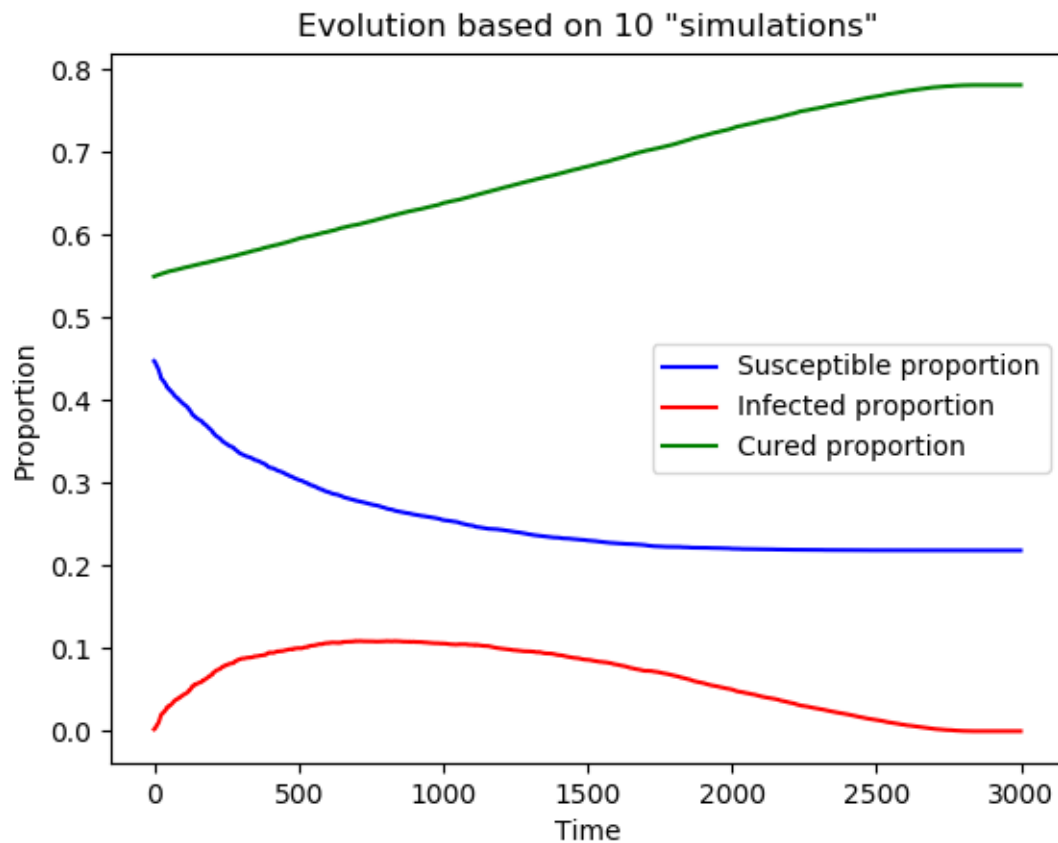


FIGURE 12 – Evolution des proportions d’individus dans chacune des catégories au cours du temps suivant un modèle représenté par une matrice d’adjacence W_{big} de taille 2000×2000 générée aléatoirement selon un modèle "scale-free" avec une proportion initiale d’individus vaccinés de 55%.

(d) Enfin, en faisant passer la probabilité μ de guérison à 1, le nombre de lits d’hopitaux est quand même dépassé. Ceci peut être observé sur la figure 13.

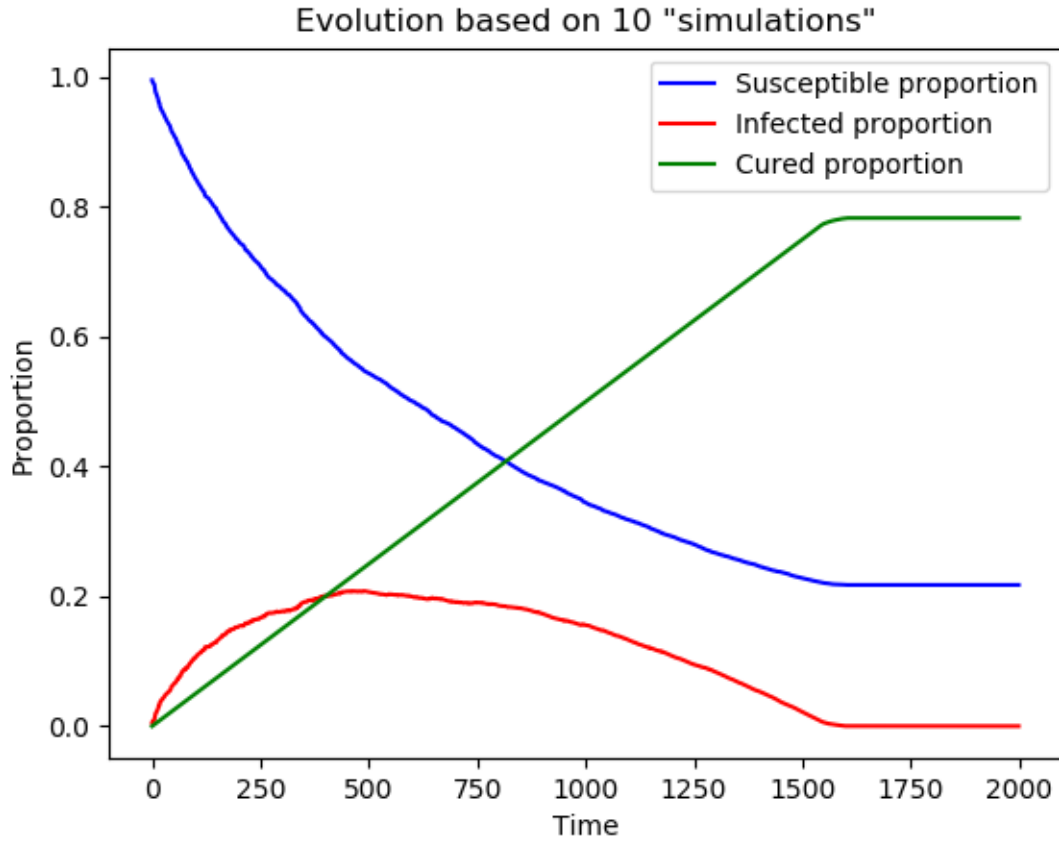


FIGURE 13 – Evolution des proportions d’individus dans chacune des catégories au cours du temps suivant un modèle représenté par une matrice d’adjacence W_{big} de taille 2000×2000 générée aléatoirement selon un modèle "scale-free" avec une accélération de la guérison du virus : $\mu = 1$.

On peut enfin comparer le temps de disparition du virus selon les différentes mesures prises. On obtient ainsi ce classement de durée décroissante : $(a) > (c) > (b) > (d)$. On conclut donc que la mesure (d) a pour principal effet de diminuer la durée de l’épidémie mais pas spécialement la proportion d’individus infectés, cette mesure pourrait donc être prise quand la maladie n’est pas importante et qu’il n’y a donc besoin d’hospitalisation que pour une proportion négligeable de la population. Les autres mesures²⁷ ont à la fois l’avantage de diminuer la proportion d’individus infectés et le temps de disparition du virus.

27. Pouvant être combinées les unes avec les autres.