O strachu przed liczeniem na papierze

SFINKS'22

mgr inż. Piotr Pasza Storożenko[1,2]

28.06.2022r.

[1]Wydział Matematyki i Nauk Infromacyjnych PW
[2]Wydział Fizyki PW
pstorozenko.github.io

Po co?

Aspirujący naukowcy (studenci) często mogą być przestraszeni koniecznością czy nawet **propozycją** rozwiązania problemu symulacyjnego na papierze.

Studium przypadku

W trakcie pracy magisterskiej

1. Przeanalizowałem ponad 270 GB danych z Twittera (ponad 350mln tłitów).

Studium przypadku

W trakcie pracy magisterskiej

1. Przeanalizowałem ponad 270 GB danych z Twittera (ponad 350mln tłitów).
2. Poszukiwałem szeregów czasowych popularności hasztagów w czasie.

Studium przypadku

W trakcie pracy magisterskiej

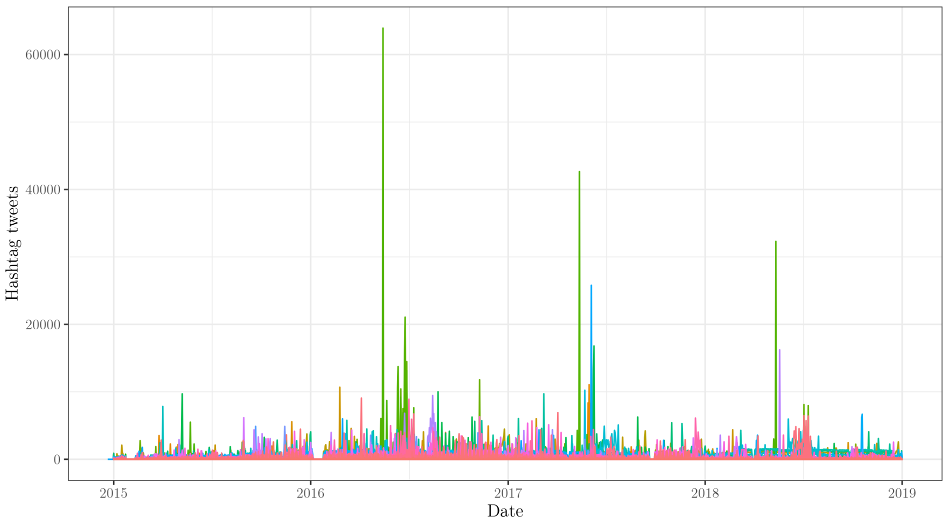
1. Przeanalizowałem ponad 270 GB danych z Twittera (ponad 350mln tłitów).
2. Poszukiwałem szeregów czasowych popularności hasztagów w czasie.
3. Zaproponowałem model symulacji agentowej rozszerzonego modelu SIR.

Studium przypadku

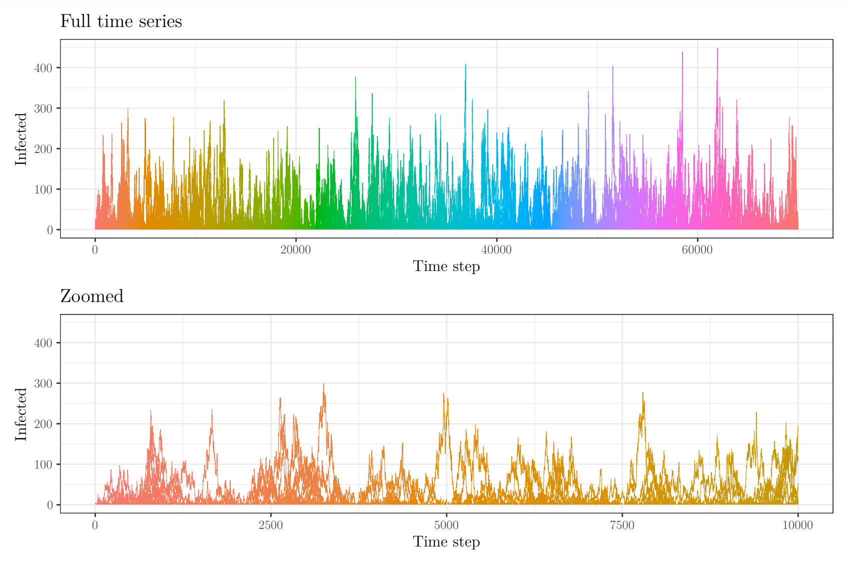
W trakcie pracy magisterskiej

1. Przeanalizowałem ponad 270 GB danych z Twittera (ponad 350mln tłitów).
2. Poszukiwałem szeregów czasowych popularności hasztagów w czasie.
3. Zaproponowałem model symulacji agentowej rozszerzonego modelu SIR.
4. Model zdawał się mieć najciekawsze właściwości dla specyficznego parametru $R_0 = 1$ – średnio tyle samo jest nowych zarażeń co wyzdrowień...

Przykładowy hashtag



Przykładowy wynik symulacji agentowej



Popchnięty przez promotra

Pewnie w tym miejscu bym skończył pracę nad magisterką.

Popchnięty przez promotra

Pewnie w tym miejscu bym skończył pracę nad magisterką.
Gdyby nie mój promotor dr inż. Grzegorz Siudem.

Popchnięty przez promotra

Pewnie w tym miejscu bym skończył pracę nad magisterką.

Gdyby nie mój promotor dr inż. Grzegorz Siudem.

Który zaproponował, żebym powalczył z modelem na papierze. ¯_(ツ)_/¯

Problem

- Skoro rozpatrujemy $R_0 = 1$ to średnio **tyle samo** agentów jest zarażanych co zdrowieje.

Problem

- Skoro rozpatrujemy $R_0 = 1$ to średnio **tyle samo** agentów jest zarażanych co zdrowieje.
- Uprośćmy model do **jednego** agenta który może zachorować albo wyzdrowieć w danym kroku czasowym.

Problem

- Skoro rozpatrujemy $R_0 = 1$ to średnio **tyle samo** agentów jest zarażanych co zdrowieje.
- Uprośćmy model do **jednego** agenta który może zachorować albo wyzdrowieć w danym kroku czasowym.
- Zaczynamy z jednym agentem, gdy epidemia wygaśnie, patrzymy **ile maksymalnie agentów było chorych**.

Problem

Co nas interesuje?
Prawdopodobieństwo, że zaczynając od 1 zarażonego dojdziemy do N i wrócimy do 0 **nie przebijając** N.

Problem

Co nas interesuje?
Prawdopodobieństwo, że zaczynając od 1 zarażonego dojdziemy do N i wrócimy do 0 **nie przebijając** N.

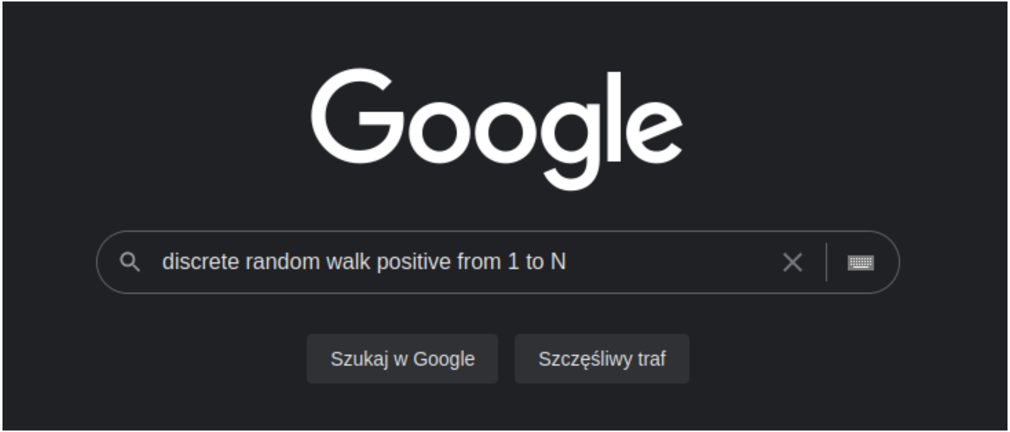
Chodząc po parku pomyślałem, że problem można rozbić na dwa etapy.

1. Policzmy prawdopodobieństwo dojścia do N zakażonych od 1.
2. Policzmy prawdopodobieństwo zejścia z N do 0.

Zdjęcie ławki

Pierwsze wpisanie w Google, rekonstrukcja czerwiec 2022

Drugie wpisanie w Google, rekonstrukcja czerwiec 2022



Problem Ruiny Gracza (The Gambler's Ruin)

Okazuje się, że problem

> *Policzmy prawdopodobieństwo dojścia do N zakażonych od 1 agenta.*

Problem Ruiny Gracza (The Gambler's Ruin)

Okazuje się, że problem

> *Policzmy prawdopodobieństwo dojścia do N zakażonych od 1 agenta.*

jest znany pod postacią

> *Policzmy prawdopodobieństwo uzyskania N monet zaczynając od 1 monety.*

jako *The Gambler's Ruin* i jego rozwiązanie to $p_N \sim 1/N$.

Korzystając z tego rozwiązania możemy policzyć prawdopodobieństwo dojścia

1. od 1 do N zakażonych (N stopni),
2. od N do 0 zakażonych ($N+1$ stopni),

jako:

Problem Ruiny Gracza (The Gambler's Ruin)

Korzystając z tego rozwiązania możemy policzyć prawdopodobieństwo dojścia

1. od 1 do N zakażonych (N stopni),
2. od N do 0 zakażonych ($N+1$ stopni),

jako:

$$p_N = \frac{1}{N} \cdot \frac{1}{N+1} = \frac{1}{N(N+1)} \tag{1}$$

Z racji, że $\sum p_N = 1$, możemy o nim myśleć jak o prawdopodobieństwie.

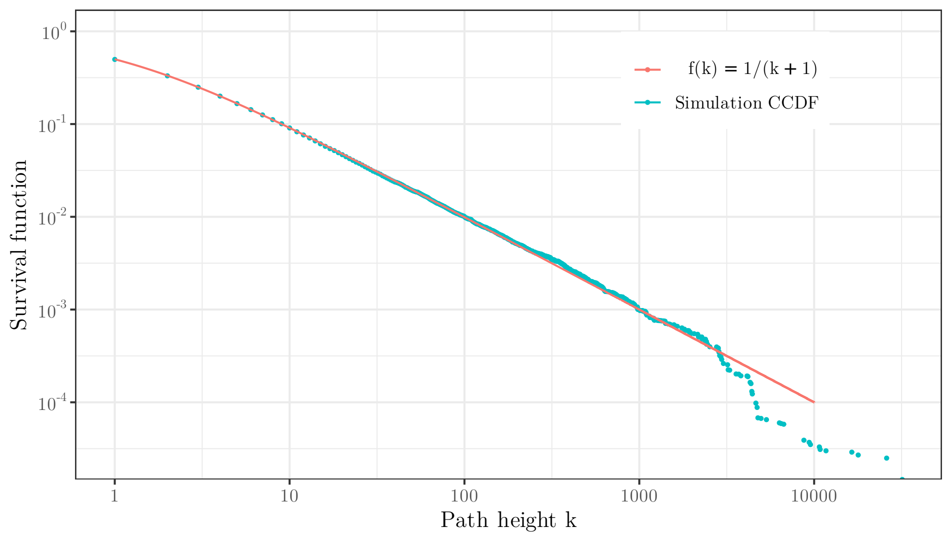
Jak to się ma do symulacji?

Z symulacji najłatwiej uzyskać tzw. **funkcję przeżycia** $S(N)$ - prawdopodobieństwo uzyskania szczytu o wysokości maksymalnie N.
Czyli:

$$S(N) = 1 - \sum_{k=1}^{N} p(k) = 1 - \sum_{k=1}^{N} \frac{1}{k(k+1)} = \text{<matematyka>} = \frac{1}{N+1} \tag{2}$$

Jak to się ma do symulacji?

Jak to się ma do symulacji?



Wnioski

Problemy są **rozwiązywalne** częściej niż się studentom wydaje!

Wnioski

Problemy są **rozwiązywalne** częściej niż się studentom wydaje!

Wiele z nich jest już **rozwiązanych**, tylko pod innymi nazwami.

Wnioski

Problemy są **rozwiązywalne** częściej niż się studentom wydaje!

Wiele z nich jest już **rozwiązanych**, tylko pod innymi nazwami.

Generalnie warto wziąć , i zamknąć się bez / na pół godziny z .

Dziękuję za uwagę!

Piotr Pasza Storożenko