

# Probabilistyczne modele propagacji w grafach.

Bartosz Łabuz

19 listopada 2025

# Spis treści

<b>1 Wstęp</b>	<b>3</b>
1.1 Motywacja i zastosowania . . . . .	3
1.2 Cel pracy . . . . .	3
1.3 Zakres pracy . . . . .	3
<b>2 Podstawy matematyczne</b>	<b>5</b>
2.1 Notacja . . . . .	5
2.2 Rodziny grafów . . . . .	5
2.3 Rozkłady prawdopodobieństwa . . . . .	6
2.4 Fakty, sumy i nierówności . . . . .	8
<b>3 Modele propagacji losowej</b>	<b>12</b>
3.1 Model SI . . . . .	12
3.2 Model SIR . . . . .	13
3.3 Model SIS . . . . .	15
<b>4 Analiza modelu SI</b>	<b>17</b>
4.1 Dwa wierzchołki, jedna krawędź . . . . .	17
4.2 Trójkąt . . . . .	17
4.3 Całkowita infekcja pewna . . . . .	19
4.4 Grafy ścieżkowe . . . . .	19
4.5 Grafy gwiazdne . . . . .	21
4.6 Ograniczenia na czas zarażenia . . . . .	24
4.7 Grafy cykliczne . . . . .	27
4.8 Grafy pełne . . . . .	31
4.9 Drzewa . . . . .	34
<b>5 Rozkłady wymierające</b>	<b>37</b>
5.1 Rozkład umierający geometryczny . . . . .	37
5.2 Rozkład umierający dwumianowy . . . . .	38
5.3 Rozkład umierający ujemny dwumianowy . . . . .	39

<b>6</b>	<b>Analiza modelu SIR</b>	<b>41</b>
6.1	Dwa wierzchołki, jedna krawędź . . . . .	41
6.2	Pewne wygaśnięcie . . . . .	42
6.3	Grafy ścieżkowe . . . . .	43
6.4	Grafy gwiazdne . . . . .	44
<b>7</b>	<b>Analiza modelu SIS</b>	<b>47</b>
7.1	Pewne wygaśnięcie . . . . .	47
7.2	Dwa wierzchołki, jedna krawędź . . . . .	47
7.3	Grafy pełne . . . . .	49

# Rozdział 1

## Wstęp

### 1.1 Motywacja i zastosowania

Propagację wirusów podczas epidemii ludzkość obserwowała już od starożytności. W dzisiejszych czasach, wraz z rozwojem internetu i mediów społeczeństwowych, mamy możliwość doświadczyć również dynamicznej propagacji informacji. Aby efektywnie rozprzestrzenić informacje, nie można robić tego “na ślepo”, lecz trzeba wykorzystać wiedzę teoretyczną. Najbardziej naturalną metodą matematycznej reprezentacji relacji międzyludzkich są grafy: wierzchołkami grafu są ludzie, a krawędzie określają, czy dane osoby mają ze sobą kontakt. Połączenie teorii grafów z rachunkiem prawdopodobieństwa pozwala stworzyć dokładny i praktyczny model propagacji informacji.

### 1.2 Cel pracy

Celem niniejszej pracy jest

- teoretyczna analiza procesów losowej propagacji w grafach,
- wyznaczenie rozkładu prawdopodobieństwa propagacji na wybranych rodzinach grafów,
- symulacja propagacji w środowisku komputerowym w celu zweryfikowania wyników teoretycznych.

### 1.3 Zakres pracy

Praca obejmuje:

- wstęp teoretyczny z zakresu teorii grafów i rachunku prawdopodobieństwa,
- opis badanych modeli propagacji: SI, SIR, SIS,
- implementację symulacji w Pythonie,
- analizę wyników i wnioski dotyczące wpływu struktury grafu na propagacje.

# Rozdział 2

## Podstawy matematyczne

### 2.1 Notacja

Przez  $\mathbb{N}$  oznaczamy zbiór  $\{0, 1, 2, \dots\}$ , a przez  $\mathbb{N}_+ = \{1, 2, 3, \dots\}$ . Moc zbioru  $A$  oznaczamy  $|A|$ . Logarytm naturalny z  $x$  oznaczamy  $\log(x)$ . Dla  $n \in \mathbb{N}_+$  przez  $H_n = 1 + \frac{1}{2} + \dots + \frac{1}{n}$  oznaczamy  $n$ 'tą liczbę harmoniczną. Jeśli  $f : \mathbb{R} \rightarrow \mathbb{R}$  jest funkcją to przez  $f(\pm\infty)$  oznaczamy  $\lim_{x \rightarrow \pm\infty} f(x)$ . Dla symbol Pochhammer'a oznaczamy  $(x)_n = x(x+1)\dots(x+n-1)$ .

Niech  $G = (V, E)$  będzie grafem prostym nieskierowanym. Zbiór sąsiadów  $v \in V$  oznaczamy  $N(v)$ . Odległość między  $u$  i  $v$  oznaczamy  $d(u, v)$  dla  $u, v \in V$ . Ekscentryczność  $v \in V$  oznaczamy  $\epsilon(v) = \max_{u \in V} d(u, v)$ .

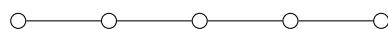
Jeśli  $\mathbb{P}$  jest miarą prawdopodobieństwa na przestrzeni  $\Omega$  to prawdopodobieństwo zdarzenia  $A$  oznaczamy  $\mathbb{P}[A]$ . Dla zmiennej losowej  $X : \Omega \rightarrow \mathbb{R}$  jej wartość oczekiwana oznaczamy  $\mathbb{E}[X]$  a jej wariancje  $\text{Var}[X]$ . Funkcję masy prawdopodobieństwa oznaczamy  $\mathbb{P}[X = t]$  a dystrybuante oznaczamy  $F_X(t) = \mathbb{P}[X \leq t]$  dla  $t \in \mathbb{R}$ .

### 2.2 Rodziny grafów

Niech  $n \in \mathbb{N}_+$ . Poniżej znajdują się opisy rodzin grafów na których będziemy rozważać propagację oraz ich wizualizację dla  $n = 5$ .

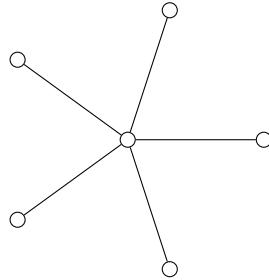
#### Graf ścieżkowy $P_n$

Graf ścieżkowy ma zbiór wierzchołków  $V = \{1, 2, \dots, n\}$  oraz zbiór krawędzi  $E = \{\{i, i+1\} : i \in \{1, 2, \dots, n-1\}\}$ . Oznaczamy go przez  $P_n$ .



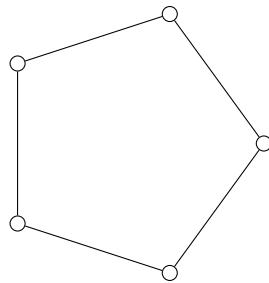
#### Graf gwiazda $S_n$

Graf gwiazda ma zbiór wierzchołków  $V = \{0, 1, \dots, n\}$  oraz zbiór krawędzi  $E = \{\{0, i\} : i \in \{1, 2, \dots, n\}\}$ . Oznaczamy go przez  $S_n$ .



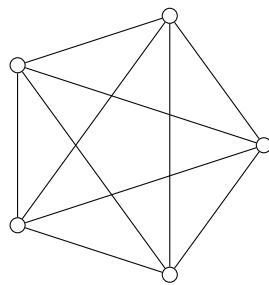
### Graf cykliczny $C_n$

Graf cykliczny ma zbiór wierzchołków  $V = \{1, 2, \dots, n\}$  oraz zbiór krawędzi  $E = \{\{i, i + 1\} : i \in \{1, 2, \dots, n - 1\}\} \cup \{\{n, 1\}\}$ . Oznaczamy go przez  $C_n$ .



### Graf pełny $K_n$

Graf pełny ma zbiór wierzchołków  $V = \{1, 2, \dots, n\}$  oraz zbiór krawędzi  $E = \{\{i, j\} : i, j \in \{1, 2, \dots, n\} \wedge i \neq j\}$ . Oznaczamy go przez  $K_n$ .



## 2.3 Rozkłady prawdopodobieństwa

Rozkład Bernoulliego  $\text{Ber}(p)$

Próba Bernoulliego to doświadczenie losowe, którego wynik może być jednym z dwóch: sukces z prawdopodobieństwem  $p \in (0; 1)$  albo porażka z prawdopodobieństwem  $1 - p$ . Zmienna losowa  $X$  ma rozkład Bernoulliego jeżeli przyjmuje wartość 1 w przypadku sukcesu i 0 w przypadku porażki próby Bernoulliego. Oznaczamy  $X \sim \text{Ber}(p)$ .

### Rozkład dwumianowy $\text{Bin}(n, p)$

Zmienna losowa  $X$  ma rozkład dwumianowy, jeżeli opisuje liczbę sukcesów w  $n$  próbach Bernoulliego gdzie  $n \in \mathbb{N}_+$ , a każda próba ma prawdopodobieństwo sukcesu  $p \in (0; 1)$ . Wtedy:

$$\mathbb{P}[X = k] = \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \{0, 1, \dots, n\}.$$

Wartość oczekiwana i wariancja wynoszą:

$$\mathbb{E}[X] = np, \quad \text{Var}[X] = np(1-p).$$

Oznaczamy  $X \sim \text{Bin}(n, p)$ .

### Rozkład geometryczny $\text{Geo}(p)$

Zmienna losowa  $X$  ma rozkład geometryczny, jeżeli opisuje liczbę prób Bernoulliego potrzebnych do uzyskania pierwszego sukcesu gdzie każda próba ma prawdopodobieństwo sukcesu  $p \in (0; 1)$ . Wtedy:

$$\mathbb{P}[X = k] = p(1-p)^{k-1}, \quad k \in \mathbb{N}_+.$$

Dystrybuanta jest równa:

$$\mathbb{P}[X \leq t] = 1 - (1-p)^t.$$

Wartość oczekiwana i wariancja wynoszą:

$$\mathbb{E}[X] = \frac{1}{p}, \quad \text{Var}[X] = \frac{1-p}{p^2}.$$

Oznaczamy  $X \sim \text{Geo}(p)$ .

### Rozkład ujemny dwumianowy $\text{NegBin}(m, p)$

Zmienna losowa  $X$  ma rozkład ujemny dwumianowy, jeżeli opisuje liczbę prób Bernoulliego potrzebnych do uzyskania  $m$  sukcesów gdzie  $m \in \mathbb{N}_+$ , a każda próba ma prawdopodobieństwo sukcesu  $p \in (0; 1)$ . Wtedy:

$$\mathbb{P}[X = k] = \binom{k-1}{m-1} p^m (1-p)^{k-m}, \quad k \geq m.$$

Wartość oczekiwana i wariancja wynoszą:

$$\mathbb{E}[X] = \frac{m}{p}, \quad \text{Var}[X] = \frac{m(1-p)}{p^2}.$$

Oznaczamy  $X \sim \text{NegBin}(m, p)$ .

**Rozkład normalny  $\mathcal{N}(\mu, \sigma^2)$**

Zdefiniujmy funkcje

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}, \quad \Phi(t) = \int_{-\infty}^t \varphi(x) dx.$$

Niech  $\mu \in \mathbb{R}$  oraz  $\sigma > 0$ . Zmienna losowa  $X$  ma rozkład normalny, jeżeli jej funkcja gęstości wyraża się wzorem:

$$f_X(t) = \frac{1}{\sigma} \cdot \varphi\left(\frac{t - \mu}{\sigma}\right), \quad t \in \mathbb{R}.$$

Dystrybuanta jest równa:

$$\mathbb{P}[X \leq t] = \Phi\left(\frac{t - \mu}{\sigma}\right), \quad t \in \mathbb{R}.$$

Wartość oczekiwana i wariancja wynoszą:

$$\mathbb{E}[X] = \mu, \quad \text{Var}[X] = \sigma^2.$$

Oznaczamy  $X \sim \mathcal{N}(\mu, \sigma^2)$ .

Jeśli  $\mu = 0$  oraz  $\sigma = 1$  to mówimy, że  $X$  ma rozkład standardowy normalny.  $\varphi$  oraz  $\Phi$  są odpowiednio PDF jak i CDF takiego rozkładu.

## 2.4 Fakty, sumy i nierówności

**Fakt 1.** Niech  $X_1, X_2, \dots, X_n : \Omega \rightarrow \mathbb{R}$  będą *IID* o CDF równej  $F_X$ . Zdefiniujmy zmienną losową  $Y = \max\{X_1, X_2, \dots, X_n\}$ . Wtedy

$$F_Y(t) = F_X^n(t).$$

**Fakt 2.** Niech  $X_1, X_2, \dots, X_n : \Omega \rightarrow \mathbb{R}$  będą *IID* o CDF równej  $F_X$ . Zdefiniujmy zmienną losową  $Y = \min\{X_1, X_2, \dots, X_n\}$ . Wtedy

$$F_Y(t) = 1 - (1 - F_X(t))^n.$$

**Fakt 3.** Niech  $X \sim \text{Bin}(n, p)$  oraz  $Y \sim \text{Bin}(m, p)$  będą niezależnymi zmiennymi losowymi. Wtedy

$$X + Y \sim \text{Bin}(n + m, p).$$

**Fakt 4.** Niech  $X_1, X_2, \dots, X_m \sim \text{Geo}(p)$  będą IID. Zdefiniujmy zmienną losową  $Y = X_1 + \dots + X_m$ . Wtedy

$$Y \sim \text{NegBin}(m, p).$$

**Fakt 5.** Niech  $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$  oraz  $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$  będą niezależnymi zmiennymi losowymi. Wtedy

$$X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

**Fakt 6.** Niech  $X \sim \text{Bin}(n, p)$ . Oznaczmy  $q = 1 - p$ . Wtedy

$$\mathbb{E}[z^X] = (q + pz)^n, \quad \mathbb{E}[Xz^X] = npz(q + pz)^{n-1}.$$

**Fakt 7.** Niech  $X \sim \text{Bin}(n, p)$  oraz  $Y \sim \text{Bin}(m, r)$  będą niezależnymi zmiennymi losowymi. Zdefiniujmy zmienną losową  $Z = X + Y$ . Wtedy

$$\mathbb{P}[Z = k] = \sum_{j=\max\{0, k-m\}}^{\min\{n, k\}} \binom{n}{j} \binom{m}{k-j} p^j (1-p)^{n-j} r^{k-j} (1-r)^{m-(k-j)}.$$

**Fakt 8.** Niech  $x, y > 0$ . Wtedy

$$\text{B}(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

**Fakt 9.** Niech  $\mathbf{X} = (X_t)_{t \in \mathbb{N}}$  będzie dyskretnym i jednorodnym łańcuchem Markova z skończoną liczbą stanów. Niech  $\mathcal{A}$  oznacza zbiór stanów absorbowących. Wtedy jeśli dla każdego stanu  $Q \notin \mathcal{A}$  jest  $k > 0$ , że dla dowolnego  $t \in \mathbb{N}$  mamy

$$\mathbb{P}[X_{t+k} \in \mathcal{A} | X_t = Q] > 0$$

to zachodzi

$$\mathbb{P}[\min\{t \in \mathbb{N} : X_t \in \mathcal{A}\} < \infty] = 1.$$

**Suma 1.** Niech  $n \in \mathbb{N}$  oraz  $x, y \in \mathbb{R}$ . Wtedy

$$\sum_{k=0}^n \binom{n}{k} x^k y^{n-k} = (x + y)^n.$$

**Suma 2.** Niech  $n \in \mathbb{N}$  oraz  $x, y \in \mathbb{R}$ . Wtedy

$$\sum_{k=0}^n k \cdot \binom{n}{k} x^k y^{n-k} = nx(x+y)^{n-1}.$$

**Suma 3.** Niech  $n \in \mathbb{N}$  oraz  $x \in \mathbb{R} \setminus \{1\}$ . Wtedy

$$\sum_{k=0}^n x^k = \frac{1-x^{n+1}}{1-x}.$$

**Suma 4.** Niech  $n \in \mathbb{N}$  oraz  $x \in \mathbb{R} \setminus \{1\}$ . Wtedy

$$\sum_{k=0}^n k \cdot x^k = \frac{x}{(1-x)^2} \cdot (nx^{n+1} - (n+1)x^n + 1).$$

**Suma 5.** Niech  $x \in (-1; 1)$ . Wtedy

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}.$$

**Suma 6.** Niech  $x \in (-1; 1)$ . Wtedy

$$\sum_{k=0}^{\infty} kx^k = \frac{x}{(1-x)^2}.$$

**Suma 7.** Niech  $m \in \mathbb{N}$  oraz  $x \in (-1; 1)$ . Wtedy

$$\sum_{k=m}^{\infty} \binom{k-1}{m-1} x^k = \frac{x^m}{(1-x)^m}.$$

**Suma 8.** Niech  $m \in \mathbb{N}$  oraz  $x \in (-1; 1)$ . Wtedy

$$\sum_{k=m}^{\infty} k \cdot \binom{k-1}{m-1} x^k = \frac{mx^m}{(1-x)^{m+1}}.$$

**Nierówność 1.** Niech  $a, b \in \mathbb{N}$ ,  $a < b$  oraz  $f : [a; b] \rightarrow \mathbb{R}$  będzie funkcją ciągłą i malejącą. Wtedy

$$\int_a^b f(x) \, dx \leq \sum_{k=a}^b f(k) \leq f(a) + \int_a^b f(x) \, dx.$$

**Nierówność 2.** Niech  $n \in \mathbb{N}_+$ . Wtedy

$$H_n \leq 1 + \log(n).$$

**Nierówność 3.** Niech  $x \in (0; 1)$ . Wtedy

$$\frac{1}{\log(\frac{1}{1-x})} \leq \frac{1}{x}.$$

**Nierówność 4** (Nierówność między średnimi). Niech  $x_1, x_2, \dots, x_n \geq 0$ . Wtedy

$$\log(x_1 \cdots x_n) \leq n \cdot \log\left(\frac{x_1 + \cdots + x_n}{n}\right).$$

Równość zachodzi wtedy i tylko wtedy gdy  $x_1 = \cdots = x_n$ .

**Nierówność 5** (Nierówność Markowa). Niech  $X$  będzie zmienną losową taką, że  $X \geq 0$  oraz  $\mathbb{E}[X] < \infty$ . Wtedy dla dowolnego  $t > 0$

$$\mathbb{P}[X \geq t] \leq \frac{\mathbb{E}[X]}{t}.$$

**Nierówność 6** (Nierówność Cauchy'ego-Schwarza). Niech  $X, Y$  będą zmiennymi losowymi takimi, że  $\mathbb{E}[X^2], \mathbb{E}[Y^2] < \infty$ . Wtedy

$$\mathbb{E}[X \cdot Y] \leq \sqrt{\mathbb{E}[X^2]} \cdot \sqrt{\mathbb{E}[Y^2]}.$$

**Nierówność 7** (Nierówność Jensena dla wartości oczekiwanej). Niech  $n \in \mathbb{N}_+$  oraz  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  będzie funkcją wypukłą zaś  $X_1, X_2, \dots, X_n : \Omega \rightarrow \mathbb{N}$  będą zmiennymi losowymi (niekoniecznie niezależnymi). Wtedy

$$g(\mathbb{E}[X_1], \dots, \mathbb{E}[X_n]) \leq \mathbb{E}[g(X_1, \dots, X_n)].$$

Jeśli  $g$  jest wklęsła to nierówność zachodzi w drugą stronę. W szczególności, ponieważ  $\max$  jest funkcją wypukłą, a  $\min$  wklęsłą, mamy:

$$\max(\mathbb{E}[X_1], \dots, \mathbb{E}[X_n]) \leq \mathbb{E}[\max(X_1, \dots, X_n)],$$

$$\min(\mathbb{E}[X_1], \dots, \mathbb{E}[X_n]) \geq \mathbb{E}[\min(X_1, \dots, X_n)].$$

## Rozdział 3

# Modele propagacji losowej

Dany jest graf spójny nieskierowany  $G = (V, E)$ . Propagacja na takim grafie jest procesem stochastycznym. Zakładamy, że czas dla tego procesu jest dyskretny i mierzony w jednostkach naturalnych, zatem za zbiór chwil przyjmujemy  $\mathbb{N}$ . Niech  $\mathcal{Q}$  będzie skończonym zbiorem stanów, jakie mogą przyjmować wierzchołki  $G$ . W każdej chwili  $t \in \mathbb{N}$  każdy wierzchołek  $v \in V$  znajduje się w pewnym stanie  $Q \in \mathcal{Q}$ . Definiujemy zmienną losową  $\mathbf{X} : \mathbb{N} \times V \rightarrow \mathcal{Q}$ , taką, że  $\mathbf{X}_t(v) = Q$  wtedy i tylko wtedy, gdy wierzchołek  $v$  w chwili  $t$  znajduje się w stanie  $Q$ .

### 3.1 Model SI

Model **Susceptible—Infected (SI)** opisuje propagację w sieci, w której każdy wierzchołek znajduje się w jednym z dwóch stanów: podatny ( $S$ ) lub zainfekowany ( $I$ ). Mamy więc  $\mathcal{Q} = \{S, I\}$ . Początkowo ustalony wierzchołek  $s \in V$  znajduje się w stanie  $I$ , natomiast pozostałe wierzchołki są w stanie  $S$ . A więc

$$\mathbf{X}_0(v) = \begin{cases} I, & \text{jeśli } v = s, \\ S, & \text{jeśli } v \neq s. \end{cases}$$

W każdej jednostce czasu dowolny zainfekowany wierzchołek może zarazić każdego swojego sąsiada z prawdopodobieństwem  $p$ , dla ustalonego  $p \in (0; 1)$ . Wierzchołek raz zainfekowany pozostaje w tym stanie na zawsze. W modelu **SI** liczba zainfekowanych wierzchołków jest funkcją niemalejącą w czasie. Dla uproszczenia notacji kładziemy

$$q = 1 - p, \quad \mathcal{S}_t = \{v \in V : \mathbf{X}_t(v) = S\}, \quad \mathcal{I}_t = \{v \in V : \mathbf{X}_t(v) = I\}.$$

Rozkład prawdopodobieństwa w tym modelu jest definiowany przez następujące zależności:

$$\begin{aligned}\mathbb{P}[\mathbf{X}_{t+1}(u) = S \mid \mathbf{X}_t(u) = S] &= \prod_{v \in N(u) \cap \mathcal{I}_t} q, \\ \mathbb{P}[\mathbf{X}_{t+1}(u) = I \mid \mathbf{X}_t(u) = S] &= 1 - \prod_{v \in N(u) \cap \mathcal{I}_t} q, \\ \mathbb{P}[\mathbf{X}_{t+1}(u) = S \mid \mathbf{X}_t(u) = I] &= 0, \\ \mathbb{P}[\mathbf{X}_{t+1}(u) = I \mid \mathbf{X}_t(u) = I] &= 1.\end{aligned}$$

Zdefiniujmy teraz zmienne losowe opisujące istotne dla nas własności. Dla każdego  $v \in V$  definiujemy zmienną  $X_v$  określającą chwilę czasu zarażenia wierzchołka  $v$ . Formalnie

$$X_v = \min\{t \in \mathbb{N} : v \in \mathcal{I}_t\}.$$

Jeśli taka chwila nie istnieje (tzn. w danym przebiegu procesu wierzchołek  $v$  nigdy się nie zarazi), to przyjmujemy  $X_v = \infty$ . Później udowodnimy (Twierdzenie 1), że w modelu **SI** mamy  $\mathbb{P}[X_v = \infty] = 0$ . Wyznaczenie rozkładu  $X_v$  jak i  $\mathbb{E}[X_v]$  da nam sporo informacji o propagacji na grafie w zależności od jego topologii. Następnie dla każdego  $t \in \mathbb{N}$  definiujemy zmienną losową  $Y_t$  oznaczającą liczbę zainfekowanych wierzchołków w chwili  $t$ . Zatem

$$Y_t = |\mathcal{I}_t|.$$

Interesować nas będzie rozkład prawdopodobieństwa  $Y_t$  oraz wartość oczekiwana  $\mathbb{E}[Y_t]$ . Pokażemy, że  $\mathbb{E}[Y_t] \rightarrow |V|$  wraz z  $t \rightarrow \infty$ . Dlatego też nie będziemy badać asymptotyki  $\mathbb{E}[Y_t]$  względem  $t$ . Dodatkowo definiujemy zmienną  $Z$  opisującą czas całkowitego zarażenia grafu:

$$Z = \min\{t \in \mathbb{N} : \mathcal{I}_t = V\}.$$

Jeśli ta sytuacja nigdy nie nastąpi to  $Z = \infty$ . Dla propagacji **SI** zainfekowanie całego grafu jest jednakże zdarzeniem pewnym. Alternatywnie możemy zapisać  $Z = \max_{v \in V} X_v$ . Wyznaczanie rozkładu  $Z$  oraz jej wartości oczekiwanej dla konkretnych rodzin grafów będzie głównym celem w tym modelu.

## 3.2 Model SIR

Model **Susceptible—Infected—Recovered (SIR)** rozszerza model **SI** o dodanie trzeciego stanu. Stanem tym jest  $R$  (Recovered). W tym modelu

mamy  $\mathcal{Q} = \{S, I, R\}$ . Stan  $R$  jest trwał — wierzchołek, który wyzdrowiał, nie może już ani się zarazić, ani nikogo zakazić. Zarażony wierzchołek może przejść z  $I$  do stanu  $R$  z prawdopodobieństwem  $\alpha \in (0; 1)$ . Przyjmujemy, że w każdej rundzie wierzchołki najpierw przekazują infekcję, a potem dopiero mogą wyzdrowieć. Dla uproszczenia notacji kładziemy

$$\beta = 1 - \alpha, \quad \mathcal{R}_t = \{v \in V : \mathbf{X}_t(v) = R\}.$$

Rozkład prawdopodobieństwa w tym modelu jest definiowany przez następujące zależności:

$$\begin{aligned} \mathbb{P}[\mathbf{X}_{t+1}(u) = S \mid \mathbf{X}_t(u) = S] &= \prod_{v \in N(u) \cap \mathcal{I}_t} q, \\ \mathbb{P}[\mathbf{X}_{t+1}(u) = I \mid \mathbf{X}_t(u) = S] &= 1 - \prod_{v \in N(u) \cap \mathcal{I}_t} q, \\ \mathbb{P}[\mathbf{X}_{t+1}(u) = R \mid \mathbf{X}_t(u) = S] &= 0, \\ \mathbb{P}[\mathbf{X}_{t+1}(u) = S \mid \mathbf{X}_t(u) = I] &= 0, \\ \mathbb{P}[\mathbf{X}_{t+1}(u) = I \mid \mathbf{X}_t(u) = I] &= \beta, \\ \mathbb{P}[\mathbf{X}_{t+1}(u) = R \mid \mathbf{X}_t(u) = I] &= \alpha, \\ \mathbb{P}[\mathbf{X}_{t+1}(u) = S \mid \mathbf{X}_t(u) = R] &= 0, \\ \mathbb{P}[\mathbf{X}_{t+1}(u) = I \mid \mathbf{X}_t(u) = R] &= 0, \\ \mathbb{P}[\mathbf{X}_{t+1}(u) = R \mid \mathbf{X}_t(u) = R] &= 1. \end{aligned}$$

Tak jak poprzednio rozważmy zmienne

$$X_v = \min\{t \in \mathbb{N} : v \in \mathcal{I}_t\}.$$

W kontraste dla modelu **SI** zachodzi  $\mathbb{P}[X_v = \infty] > 0$ . (patrz Twierdzenie 4). A więc poza rozkładem  $X_v$  możemy wyznaczyć  $\mathbb{E}[X_v | X_v < \infty]$ . Istnienie stanu  $R$  narzuca pomysł rozważania podobnej zmiennej na pierwszy czas wyzdrowienia wierzchołka  $v$ . Ale transmisja ze stanu  $I$  do  $R$  na pojedynczym wierzchołku jest rozkładem  $\text{Geo}(\alpha)$ , a więc zmienna ta była by po prostu sumą rozkładu geometrycznego z  $X_v$ . Z tego powodu tego nie będziemy jej rozważać. Zamiast rozważać liczbę tylko zainfekowanych lub tylko wyzdrowiałych wierzchołków będziemy rozważać liczbę nie podatnych wierzchołków po  $t$  krokach. Kładziemy więc

$$Y_t = |\mathcal{I}_t \cup \mathcal{R}_t|.$$

Zdarzenie  $\mathbb{P}[\mathcal{I}_t = \emptyset] \rightarrow 1$  wraz z  $t \rightarrow \infty$  a więc możemy zdefiniować zmienną  $Z$  oznaczającą czas wygaśnięcia infekcji. Dla uproszczenia będziemy rozważać moment w którym żaden nowy wierzchołek nie może być zarażony. A więc

$$Z = \min\{t \in \mathbb{N} : \forall v \in \mathcal{I}_t \quad \mathcal{S}_t \cap N(v) = \emptyset\}.$$

Dodatkowo definiujemy zmienną  $W$  będącą liczbą finalnie wyzdrowiałych wierzchołków.

$$W = |\{v \in V : X_v < \infty\}|.$$

Rozkłady  $Z, W$  jak i ich wartości oczekiwane będą naszym głównym zainteresowaniem w tym modelu.

### 3.3 Model SIS

Model **Susceptible—Infected—Susceptible (SIS)** rozszerza model **SI** o powracanie wierzchołków zarażonych do stanu podatnego. Wierzchołek zainfekowany może powrócić do stanu podatnego z prawdopodobieństwem  $\alpha \in (0; 1)$ . Tutaj mamy również  $\mathcal{Q} = \{S, I\}$ . W modelu **SIS** liczba zainfekowanych wierzchołków może oscylować w czasie i nie musi osiągnąć stanu pełnego zakażenia. Dla uproszczenia notacji kładziemy  $\beta = 1 - \alpha$ . Przyjmujemy, że w każdej rundzie wierzchołki najpierw przekazują infekcję, a potem dopiero mogą powrócić w stan podatności. Rozkład prawdopodobieństwa w tym modelu jest definiowany przez następujące zależności:

$$\begin{aligned} \mathbb{P}[\mathbf{X}_{t+1}(u) = S \mid \mathbf{X}_t(u) = S] &= \prod_{v \in N(u) \cap \mathcal{I}_t} q, \\ \mathbb{P}[\mathbf{X}_{t+1}(u) = I \mid \mathbf{X}_t(u) = S] &= 1 - \prod_{v \in N(u) \cap \mathcal{I}_t} q, \\ \mathbb{P}[\mathbf{X}_{t+1}(u) = S \mid \mathbf{X}_t(u) = I] &= \alpha, \\ \mathbb{P}[\mathbf{X}_{t+1}(u) = I \mid \mathbf{X}_t(u) = I] &= \beta. \end{aligned}$$

Zmienne losowe opisujące istotne własności są tutaj podobne jak w modelu **SI**. Dla każdego  $v \in V$  kładziemy

$$X_v = \min\{t \in \mathbb{N} : v \in \mathcal{I}_t\},$$

oraz dla  $t \in \mathbb{N}$  definiujemy

$$Y_t = |\mathcal{I}_t|.$$

Dodatkowo definiujemy zmienną losową opisującą czas wygaśnięcia infekcji:

$$Z = \min\{t \in \mathbb{N} : \mathcal{I}_t = \emptyset\}.$$

Wygaśnięcie infekcji zachodzi z prawdopodobieństwem 1 (patrz Twierdzenie 5). Mamy  $\mathbb{P}[X_v = \infty] > 0$  a więc interesować nas będzie  $\mathbb{E}[X_v | X_v < \infty]$ . Dalej wykażemy, że  $\mathbb{E}[Y_t]$  dąży do 0, a więc także nie będą nas interesować. Skupimy się tylko na rozkładzie  $Y_t$ . Jeśli chodzi o zmienną  $Z$  to przyjrzyjmy się zarówno jej rozkładowi jak i wartości oczekiwanej.

## Rozdział 4

# Analiza modelu SI

### 4.1 Dwa wierzchołki, jedna krawędź

Na samym początku rozważmy najprostszy graf, czyli taki o dwóch wierzchołkach  $u, v$  połączonych krawędzią. Za wierzchołek startowy wybierzmy  $u$ . Istnieją tylko dwa możliwe stany systemu:  $\mathcal{I}_t = \{u\}$  oraz  $\mathcal{I}_t = \{u, v\}$ . Przejście ze stanu  $\mathcal{I}_t = \{u\}$  do  $\mathcal{I}_t = \{u, v\}$  następuje z prawdopodobieństwem  $p$  w każdej jednostce czasu. Zatem czas zarażenia drugiego wierzchołka  $X_v$  ma rozkład geometryczny,  $X_v \sim \text{Geo}(p)$ .

Rozważmy teraz rozkład  $Y_t$ . Mamy  $\mathbb{P}[Y_t = 1] = q^t$ , ponieważ próba zarażenia musiałaby nie udało się  $t$  razy, oraz  $\mathbb{P}[Y_t = 2] = 1 - q^t$ . Stąd  $\mathbb{E}[Y_t] = 1 \cdot q^t + 2 \cdot (1 - q^t) = 2 - q^t$ .

Jeśli chodzi o zmienną  $Z$ , to zachodzi  $Z = \max\{X_u, X_v\} = X_v$ , a więc również  $Z \sim \text{Geo}(p)$  oraz  $\mathbb{E}[Z] = \frac{1}{p}$ .

### 4.2 Trójkąt

Przyjrzymy się teraz nieco większemu grafowi — trójkątowi. Niech jeden z wierzchołków będzie źródłem  $s$ , a pozostałe  $u, v$ . Aby poinformować  $u$ , musimy uzyskać sukces bezpośrednio od  $s$  lub zarazić  $v$ , a następnie  $u$ . Zatem zachodzi  $X_u = \min\{A, B\}$ , gdzie  $A \sim \text{Geo}(p)$  oraz  $B \sim \text{NegBin}(2, p)$ . Wiemy, że  $\mathbb{P}[A \leq t] = 1 - q^t$ . Z kolei

$$\begin{aligned}\mathbb{P}[B \leq t] &= \sum_{k=2}^t (k-1) \cdot p^2 q^{k-2} = \frac{p^2}{q} \cdot \frac{q}{(1-q)^2} \cdot ((t-1)q^t - tq^{t-1} + 1) \\ &= 1 - q^t - tpq^{t-1}.\end{aligned}$$

Dalej, z Faktu 2 mamy

$$\mathbb{P}[X_u \leq t] = 1 - (1 - (1 - q^t)) \cdot (1 - (1 - q^t - tpq^{t-1})) = 1 - q^{2t} - tpq^{2t-1}.$$

Jeśli chodzi o liczbę zainfekowanych po  $t$  krokach, to skoro mamy trzy wartości do policzenia. Oczywiście  $\mathbb{P}[Y_t = 1] = q^{2t}$ . Aby po  $t$  chwilach tylko dwa węzły były zainfekowane, musimy zarazić któryś z wierzchołków po  $1 \leq k \leq t$  rundach z prawdopodobieństwem  $2pq \cdot q^{2 \cdot (k-1)}$ , a następnie uzyskać  $t - k$  porażek. Na każdą z nich mamy szansę równą  $q^2$ . A zatem

$$\mathbb{P}[Y_t = 2] = \sum_{k=1}^t 2pq \cdot q^{2 \cdot (k-1)} \cdot q^{2 \cdot (t-k)} = 2tpq^{2t-1}.$$

Na koniec mamy  $\mathbb{P}[Y_t = 3] = 1 - q^{2t} - 2tpq^{2t-1}$ . Ponadto:

$$\mathbb{E}[Y_t] = 1 \cdot q^{2t} + 2 \cdot 2tpq^{2t-1} + 3 \cdot (1 - q^{2t} - 2tpq^{2t-1}) = 3 - 2q^{2t} - 2tpq^{2t-1}.$$

Propagacja może się zakończyć na dwa sposoby. Pierwszy z nich to sytuacja, w której przez  $t - 1$  jednostek czasu żadne zakażenie nie zaszło, a w chwili  $t$  zarażają się oba wierzchołki. Prawdopodobieństwo tego przypadku wynosi  $p^2 q^{2 \cdot (t-1)}$ . Druga możliwość to taka, w której w  $k$ -tym kroku (dla  $1 \leq k \leq t - 1$ ) zaraził się jeden z wierzchołków — z prawdopodobieństwem  $2pq \cdot q^{2 \cdot (k-1)}$ , a potem przez kolejne  $t - 1 - k$  kroków trzeci wierzchołek nie został zainfekowany, na co mamy prawdopodobieństwo  $(q^2)^{t-1-k}$ , aż do chwili  $t$ . To ostatnie przejście ma prawdopodobieństwo  $1 - q^2$ . Podsumowując:

$$\begin{aligned} \mathbb{P}[Z = t] &= p^2 q^{2t-2} + \sum_{k=1}^{t-1} (2pq q^{2k-2}) \cdot (q^{2t-2k-2}) \cdot (1 - q^2) \\ &= p^2 q^{2t-2} + 2pq^{2t-3} \cdot (t-1)(1-q^2). \end{aligned}$$

Mamy też  $\mathbb{P}[Z > t] = \mathbb{P}[Y_t \neq 3] = q^{2t} + 2tpq^{2t-1}$ . Wartość oczekiwana wynosi więc:

$$\begin{aligned} \mathbb{E}[Z] &= \sum_{t=0}^{\infty} \mathbb{P}[Z > t] = \sum_{t=0}^{\infty} q^{2t} + 2tpq^{2t-1} \\ &= \frac{1}{1 - q^2} + \frac{2p}{q} \cdot \frac{q^2}{(1 - q^2)^2} = \frac{-3q^2 + 2q + 1}{(1 - q^2)^2} = \frac{4 - 3p}{p(2 - p)^2}. \end{aligned}$$

Wykonaliśmy dość sporo obliczeń jak na tak mały graf. Możemy więc zauważyć, że istnienie cykli w grafie znacząco komplikuje sytuację, jeśli chodzi o model **SI**.

### 4.3 Całkowita infekcja pewna

Dość intuicyjny jest fakt, że każdy wierzchołek zostanie kiedyś zarażony z prawdopodobieństwem 1. Co za tym idzie cały graf znajdzie się w stanie  $I$  oraz  $Z < \infty$ . Postaramy się teraz formalnie tego dowieść.

**Twierdzenie 1.** *Niech  $G = (V, E)$  będzie grafem spójnym,  $s \in V$  źródłem infekcji oraz  $v \in V \setminus \{s\}$ . Wtedy zachodzą następujące tożsamości:*

$$\mathbb{P}[X_v = \infty] = 0, \quad \lim_{t \rightarrow \infty} \mathbb{E}[Y_t] = |V|, \quad \mathbb{P}[Z < \infty] = 1.$$

*Dowód.* Rozważmy propagację jako łańcuch Markova. W danej chwili stanem jest  $\mathcal{I}_t$ . Stan absorbujący wynosi  $V$ . Jeśli

$$\ell = \max_{u \in \mathcal{I}_t, v \in \mathcal{S}_t} d(u, v)$$

to po  $\ell$  rundach możemy przejść do stanu całkowitej infekcji grafu z dodatnim prawdopodobieństwem. A zatem z Faktu 9 mamy  $\mathbb{P}[Z < \infty] = 1$ . Ponadto  $\mathbb{P}[\max_{v \in V} X_v = \infty] = \mathbb{P}[Z = \infty] = 0$  a więc dla dowolnego  $v \in V$  mamy  $\mathbb{P}[X_v = \infty] = 0$ . Dalej zauważmy, że  $0 \leq Y_t \leq |V|$ . Z Twierdzenia Lebesgue o zbieżności ograniczonej mamy

$$\lim_{t \rightarrow \infty} \mathbb{E}[Y_t] = \mathbb{E}[\lim_{t \rightarrow \infty} Y_t] = \mathbb{E}\left[\lim_{t \rightarrow \infty} \sum_{v \in V} \mathbf{1}_{v \in \mathcal{I}_t}\right] = \mathbb{E}\left[\sum_{v \in V} 1\right] = \mathbb{E}[|V|] = |V|.$$

□

### 4.4 Grafy ścieżkowe

Jako pierwszą rodzinę grafów rozważmy grafy ścieżkowe  $P_n$ . Założymy, że proces zaczyna się w wierzchołku  $s = 1$ . Zatem infekcja rozchodzi się po grafie „od lewej do prawej”. Zauważmy, że czasy zarażenia kolejnych wierzchołków tworzą ciąg zmiennych losowych:

$$X_1 = 0, \quad X_v = X_{v-1} + U_v, \quad v \in \{2, 3, \dots, n\},$$

gdzie  $U_2, U_3, \dots, U_n \sim \text{Geo}(p)$  oraz  $U_2, U_3, \dots, U_n$  są niezależne. Widzimy zatem, że  $X_v = U_2 + U_3 + \dots + U_v$ , a więc z Faktu 4  $X_v$  ma rozkład ujemny dwumianowy:

$$X_v \sim \text{NegBin}(v - 1, p),$$

oraz

$$\mathbb{E}[X_v] = \frac{v-1}{p}.$$

Ustalmy  $t \in \mathbb{N}$  i przejdźmy do obliczania rozkładu  $Y_t$ . Zauważmy, że liczba dodatkowych zakażeń poza startowym wierzchołkiem do czasu  $t$  to po prostu liczba sukcesów w  $t$  niezależnych prób Bernoulliego. Musimy jednak pamiętać, że  $Y_t$  nie może przekroczyć  $n$ . Zatem mamy dokładnie:

$$Y_t = \min\{n, 1 + B_t\}, \quad B_t \sim \text{Bin}(t, p).$$

Pozwala to na wyznaczenie PMF dla  $Y_t$ . Dla  $1 \leq k \leq n - 1$  mamy:

$$\mathbb{P}[Y_t = k] = \mathbb{P}[B_t = k - 1] = \binom{t}{k-1} p^{k-1} q^{t-k+1},$$

oraz dla  $k = n$ :

$$\mathbb{P}[Y_t = n] = \mathbb{P}[B_t \geq n - 1] = \sum_{j=n-1}^t \binom{t}{j} p^j q^{t-j}.$$

Przejdźmy teraz do obliczania wartości oczekiwanej  $Y_t$ :

$$\begin{aligned} \mathbb{E}[Y_t] &= \sum_{k=1}^{n-1} k \cdot \mathbb{P}[Y_t = k] + n \cdot \mathbb{P}[Y_t = n] \\ &= \sum_{k=1}^{n-1} k \cdot \binom{t}{k-1} p^{k-1} q^{t-k+1} + n \cdot \sum_{j=n-1}^t \binom{t}{j} p^j q^{t-j} \\ &= \sum_{j=0}^t \min\{n, 1 + j\} \cdot \binom{t}{j} p^j q^{t-j}. \end{aligned}$$

Policzmy teraz asymptotykę dla  $n \rightarrow \infty$ . Wtedy  $n > 1 + j$  dla wszystkich  $0 \leq j \leq t$ , a więc:

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E}[Y_t] &= \sum_{j=0}^t (1 + j) \binom{t}{j} p^j q^{t-j} = \sum_{j=0}^t \binom{t}{j} p^j q^{t-j} + \sum_{j=0}^t j \binom{t}{j} p^j q^{t-j} \\ &= (p + q)^t + tp(p + q)^{t-1} = 1 + tp. \end{aligned}$$

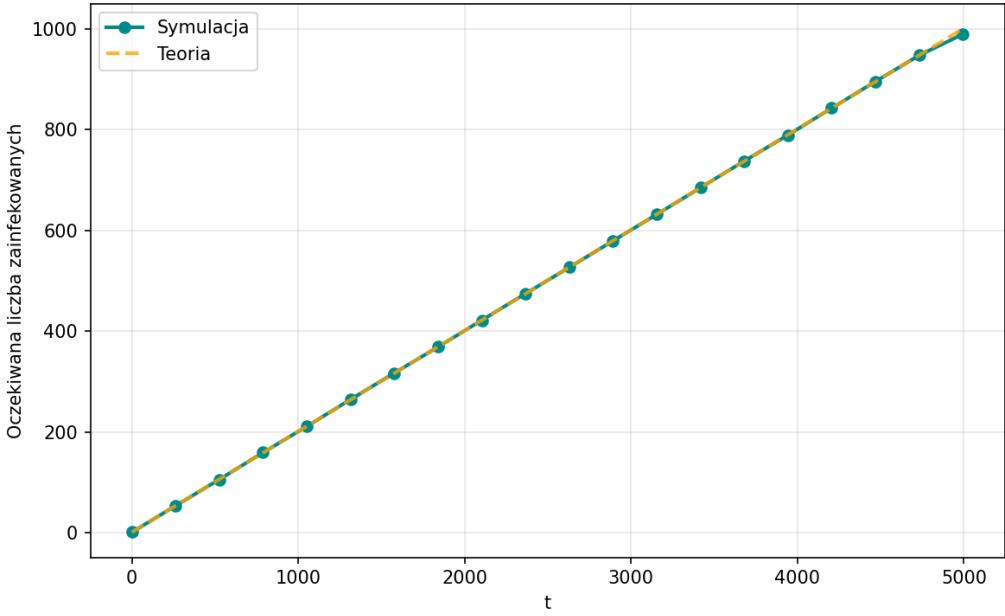
Czas całkowitego zainfekowania grafu to  $Z = \max\{X_1, X_2, \dots, X_n\} = X_n$ . Zatem rozkład zmiennej  $Z$  jest już nam znany,

$$Z \sim \text{NegBin}(n - 1, p),$$

a wartość oczekiwana wynosi

$$\mathbb{E}[Z] = \frac{n - 1}{p}.$$

Sprawdźmy, czy nasze obliczenia teoretyczne zgadzają się z empirycznie wyznaczonymi wartościami. Ustalmy  $p = 0.2$ . W celu estymacji  $\mathbb{E}[Y_t]$  przeprowadźmy 2000 symulacji propagacji na grafie  $P_n$ . Następnie dla  $n \in \{1, 2, \dots, 1000\}$  tą samą liczbą symulacji oszacujmy  $\mathbb{E}[Z]$ . Wyniki eksperymentu niemal idealnie pokrywają się z przewidywaniami, to jest  $1 + tp$  dla  $\mathbb{E}[Y_t]$  (zob. Wykres 1) oraz  $\frac{n-1}{p}$  dla  $\mathbb{E}[Z]$  (zob. Wykres 2).



Rysunek 1:  $\mathbb{E}[Y_t]$  dla  $P_n$  w funkcji  $t$ .

## 4.5 Grafy gwiazdne

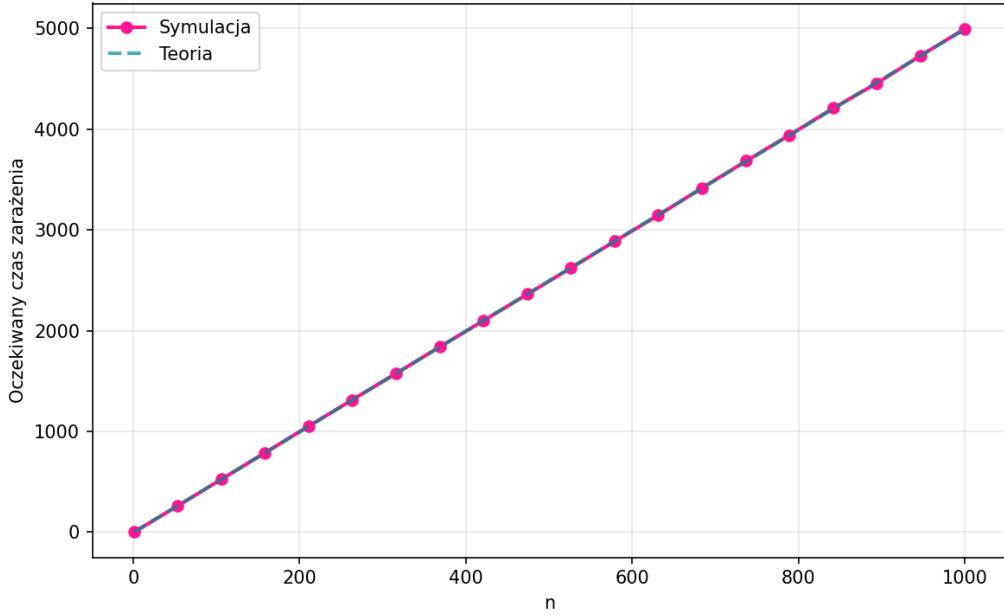
Następnie rozpatrzmy rodzinę grafów gwiazd  $S_n$ . Niech źródłem będzie centralny wierzchołek grafu, czyli  $s = 0$ . Propagacja rozchodzi się tutaj po każdym ramieniu gwiazdy niezależnie a więc zmienne  $X_1, X_2, \dots, X_n$  są niezależne. Stąd dla każdego  $v \in \{1, 2, \dots, n\}$  mamy

$$X_v \sim \text{Geo}(p),$$

oraz

$$\mathbb{E}[X_v] = \frac{1}{p}.$$

Kwestia zmiennej  $Y_t$  jest również prosta. Ponieważ propagacja działa niezależnie na każdym wierzchołku,  $Y_t$  odpowiada liczbie sukcesów w  $n$  próbach Bernoulliego. Sukcesem pojedynczej próby jest zdarzenie, że zmienna



Rysunek 2:  $\mathbb{E}[Z]$  dla  $P_n$  w funkcji  $n$ .

$X_v$  o rozkładzie geometrycznym osiągnie sukces w czasie co najwyżej  $t$ . Skoro  $\mathbb{P}[X_v \leq t] = 1 - q^t$  to mamy

$$Y_t = 1 + B_t, \quad B_t \sim \text{Bin}(n, 1 - q^t).$$

W konsekwencji:

$$\mathbb{E}[Y_t] = 1 + n \cdot (1 - q^t).$$

Przejdźmy teraz do zmiennej  $Z$ . Mamy  $Z = \max\{X_1, X_2, \dots, X_n\}$ . Ponieważ zmienne te są IID, z Faktu 1 otrzymujemy:

$$\mathbb{P}[Z \leq t] = (1 - q^t)^n.$$

Policzmy wartość oczekiwana całkowitego zainfekowania grafu:

$$\begin{aligned}
\mathbb{E}[Z] &= \sum_{k=1}^{\infty} \mathbb{P}[Z \geq k] = \sum_{k=1}^{\infty} 1 - \mathbb{P}[Z \leq k-1] = \sum_{k=1}^{\infty} 1 - (1-q^{k-1})^n \\
&= \sum_{k=0}^{\infty} 1 - (1-q^k)^n = \sum_{k=0}^{\infty} \left(1 - \sum_{j=0}^n \binom{n}{j} (-1)^j q^{kj}\right) \\
&= \sum_{k=0}^{\infty} \sum_{j=1}^n \binom{n}{j} (-1)^{j+1} q^{kj} = \sum_{j=1}^n \sum_{k=0}^{\infty} \binom{n}{j} (-1)^{j+1} (q^j)^k \\
&= \sum_{j=1}^n \binom{n}{j} \frac{(-1)^{j+1}}{1-q^j}.
\end{aligned}$$

Nie jest to jednak szczególnie elegancka forma, więc spróbujmy wyznaczyć asymptotykę  $\mathbb{E}[Z]$ . Zauważmy, że

$$\mathbb{E}[Z] = \sum_{k=0}^{\infty} 1 - (1-q^k)^n.$$

Z Nierówności 1 możemy przybliżyć tę sumę całką. Niech  $f(x) = 1 - (1-e^{-\lambda x})^n$ , gdzie  $\lambda = -\log(q)$ . Oczywiście  $f(0) = 1$ ,  $f(\infty) = 0$ , a funkcja  $f$  jest malejąca, więc:

$$\int_0^{\infty} f(x) dx \leq \mathbb{E}[Z] \leq 1 + \int_0^{\infty} f(x) dx.$$

Podstawiamy  $u = 1 - e^{-\lambda x}$ . Wtedy  $du = \lambda e^{-\lambda x} dx$ , a więc  $dx = \frac{1}{\lambda} \cdot \frac{1}{1-u} du$ . Ponadto  $u(0) = 0$ ,  $u(\infty) = 1$  (bo  $\lambda > 0$ ). Zatem całka ma postać:

$$\frac{1}{\lambda} \int_0^1 \frac{1-u^n}{1-u} du = \frac{1}{\lambda} \sum_{j=0}^{n-1} \frac{1}{j+1} = \frac{H_n}{\lambda}.$$

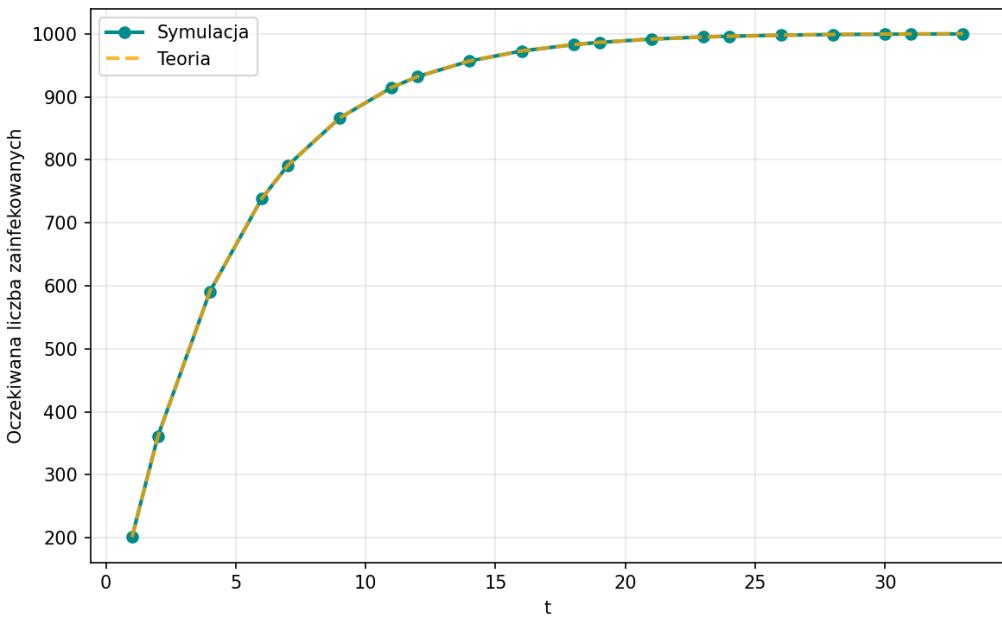
Zauważmy, że  $-\log(q) = \log\left(\frac{1}{1-p}\right)$ , zatem:

$$\frac{H_n}{\log\left(\frac{1}{1-p}\right)} \leq \mathbb{E}[Z] \leq \frac{H_n}{\log\left(\frac{1}{1-p}\right)} + 1.$$

Stąd asymptotyczny czas pełnego zarażenia grafu  $S_n$  wynosi:

$$\mathbb{E}[Z] \sim \frac{H_n}{\log\left(\frac{1}{1-p}\right)}.$$

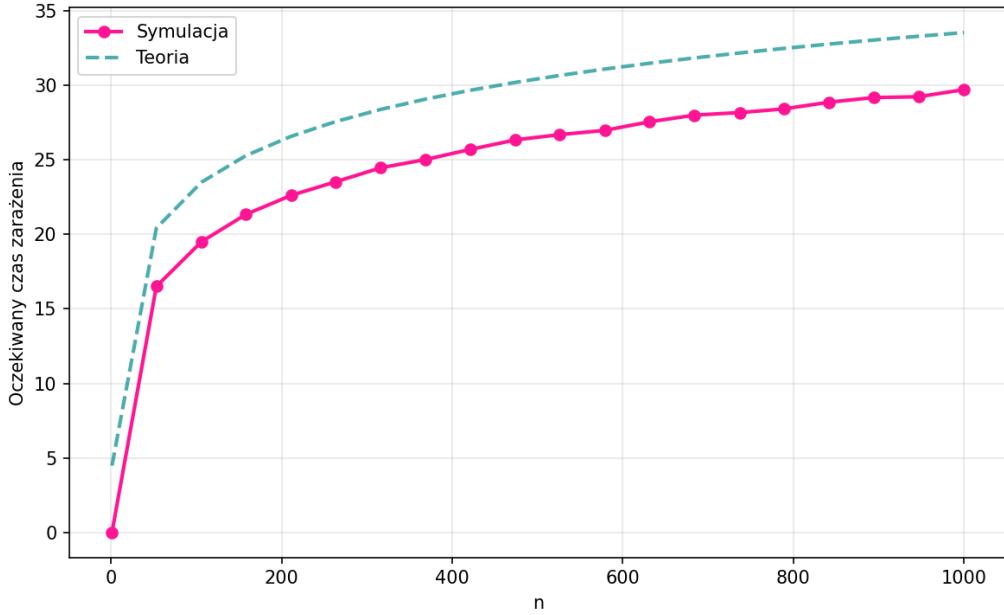
Przeprowadźmy teraz symulacje. Ustalmy  $p = 0.2$  oraz  $n = 1000$ . Dla każdego  $t \in \{1, 2, \dots, \log(n)\}$  wykonajmy 2000 powtórzeń propagacji na  $S_n$ , a następnie dla  $n \in \{1, 2, \dots, 1000\}$  oszacujmy  $\mathbb{E}[Z]$ . Wyniki przedstawiono na Wykresie 3 oraz Wykresie 4. Dla  $\mathbb{E}[Y_t]$  obserwujemy niemal idealne dopasowanie do przewidywanego kształtu. Natomiast dla  $\mathbb{E}[Z]$  wartości empiryczne są około o 1 mniejsze od oczekiwanych, co stanowi bardzo dobre przybliżenie.



Rysunek 3:  $\mathbb{E}[Y_t]$  dla  $S_n$  w funkcji  $t$ .

## 4.6 Ograniczenia na czas zarażenia

Po rozważeniu dwóch rodzin grafów dostrzegamy znaczną różnicę w wartościach oczekiwanych zmiennych  $Y_t$  oraz  $Z$ . Dla grafów ścieżkowych minimalna liczba rund potrzebnych do zainfekowania całego grafu wynosi  $t = n - 1$ , natomiast dla gwiazd jest to zaledwie  $t = 1$ . Widzimy więc, że w pewnym sensie najlepszy przypadek sprzyjający szybkiemu rozprzestrzenianiu się infekcji zachodzi wtedy, gdy źródło  $s$  jest połączone ze wszystkimi pozostałymi wierzchołkami grafu. Z drugiej strony, najgorsza sytuacja ma miejsce, gdy istnieje odległy węzeł z niewielką liczbą ścieżek prowadzących do niego — tak jak w przypadku grafów ścieżkowych. Teraz postaramy się uogólnić tę obserwację.



Rysunek 4:  $\mathbb{E}[Z]$  dla  $S_n$  w funkcji  $n$ .

**Twierdzenie 2.** Niech  $G = (V, E)$  będzie grafem spójnym, a  $G' = (V, E')$  jego spójnym podgrafem. Założmy, że  $\mathbf{X}$  opisuje proces stochastyczny w modelu **SI** prowadzony równocześnie na  $G$  oraz  $G'$  z tym samym źródłem  $s \in V$ . Jeśli przez  $X'_v$ ,  $Y'_t$  oraz  $Z'$  oznaczamy odpowiednie zmienne losowe dla  $G'$ , to zachodzą nierówności:

$$X_v \leq X'_v, \quad Y_t \geq Y'_t, \quad Z \leq Z'.$$

*Dowód.* Oznaczmy przez  $\mathcal{I}_t$  zbiór zainfekowanych wierzchołków w grafie  $G$ , a przez  $\mathcal{I}'_t$  — w grafie  $G'$ . Wtedy  $\mathcal{I}'_t \subseteq \mathcal{I}_t$  dla każdego  $t \in \mathbb{N}$ . Ustalmy  $v \in V$  i niech  $X'_v = a$ . Wtedy  $v \in \mathcal{I}'_a$ , a więc także  $v \in \mathcal{I}_a$ , co implikuje  $X_v \leq a = X'_v$ . Analogicznie, dla ustalonego  $t \in \mathbb{N}$  z faktu, że  $\mathcal{I}'_t \subseteq \mathcal{I}_t$ , mamy  $|\mathcal{I}'_t| \leq |\mathcal{I}_t|$ , a zatem  $Y'_t \leq Y_t$ . Na koniec, jeśli  $Z' = b$ , to  $\mathcal{I}'_b = V$ , a więc  $V \subseteq \mathcal{I}_b$ , co prowadzi do  $Z \leq b = Z'$ .  $\square$

Intuicyjnie, wynik ten jest oczywisty — mając mniej krawędzi w grafie, potrzebujemy więcej czasu, aby informacja (lub infekcja) rozprzestrzeniła się po całym grafie. W praktyce oznacza to, że jeśli znamy średni czas pełnego zainfekowania dowolnego podgrafa  $G$ , to otrzymujemy górne ograniczenie dla całego grafu. Spróbujmy teraz oszacować z góry wartość  $\mathbb{E}[Z]$  dla dowolnego grafu.

**Twierdzenie 3.** Niech  $G = (V, E)$  będzie grafem o  $n$  wierzchołkach oraz  $s \in V$  — ustalonym źródłem. Oznaczmy  $\lambda = \log(\frac{1}{1-p})$  oraz  $h = \epsilon(s)$ . Wtedy zachodzi:

$$\mathbb{E}[Z] \leq h + \frac{h}{\lambda} \left( \log\left(\frac{n-1}{h}\right) + 1 \right).$$

*Dowód.* Dla  $0 \leq j \leq h$  zdefiniujmy zbiory  $A_j = \{v \in V : d(s, v) = j\}$  oraz położmy  $a_j = |A_j|$ . Mamy oczywiście  $a_0 = 1$ , a więc  $a_1 + \dots + a_h = n - 1$ . Dalej zdefiniujmy zmienne losowe:

$$T_j = \min\{t \in \mathbb{N} : A_j \subseteq \mathcal{I}_t\}.$$

Zmienna  $T_j$  określa czas potrzebny na zainfekowanie wszystkich wierzchołków w odległości  $j$  od źródła. Udowodnijmy teraz pomocniczy lemat.

**Lemat 1.** Niech  $U_j = T_j - T_{j-1}$  dla  $1 \leq j \leq h$ . Wtedy:

$$\mathbb{E}[U_j] \leq \frac{H_{a_j}}{\lambda} + 1.$$

*Dowód.* Zmienna  $U_j$  opisuje czas potrzebny na zainfekowanie wierzchołków z  $A_j$ , zakładając, że wszystkie wierzchołki z  $A_{j-1}$  są już zainfekowane. Skonstruujmy podgraf  $G'$  w taki sposób, by każdy wierzchołek z  $A_j$  był połączony dokładnie jedną krawędzią z pewnym wierzchołkiem ze zbioru  $A_{j-1}$ . Wtedy proces propagacji na  $G'$  jest izomorficzny z tym na grafie gwiazdy  $S_{a_j}$ , gdzie  $a_j = |A_j|$ . Z Twierdzenia 2 wynika, że zmienna  $U_j$  jest ograniczona przez całkowity czas infekcji w  $S_{a_j}$ . Zgodnie z wcześniejszymi wynikami, jego wartość oczekiwana wynosi co najwyżej  $\frac{H_{a_j}}{\lambda} + 1$ , co kończy dowód.  $\square$

Wróćmy do dowodu twierdzenia. Mamy  $T_h = \sum_{j=1}^h U_j$  a więc

$$\begin{aligned} \mathbb{E}[Z] &= \mathbb{E}[T_h] = \mathbb{E}\left[\sum_{j=1}^h U_j\right] = \sum_{j=1}^h \mathbb{E}[U_j] \leq \sum_{j=1}^h \frac{H_{a_j}}{\lambda} + 1 \\ &= h + \frac{1}{\lambda} \sum_{j=1}^h H_{a_j} \leq h + \frac{1}{\lambda} \sum_{j=1}^h 1 + \log(a_j) \\ &= h + \frac{h}{\lambda} \cdot \left(1 + \sum_{j=1}^h \log(a_j)\right) = h + \frac{h}{\lambda} \cdot \left(1 + \log\left(\prod_{j=1}^h a_j\right)\right) \\ &\leq h + \frac{h}{\lambda} \cdot \left(1 + \log\left(\frac{1}{h} \sum_{j=1}^h a_j\right)\right) = h + \frac{h}{\lambda} \cdot \left(1 + \log\left(\frac{n-1}{h}\right)\right), \end{aligned}$$

gdzie w linijce pierwszej wykorzystujemy Lemat 1, w drugiej Nierówność 2 a w piątej nierówność między średnimi (4).  $\square$

Porównajmy teraz powyższy wynik z wcześniejszymi obserwacjami. Dla rodziny grafów  $P_n$  mamy  $h = n - 1$ . Korzystając z Nierówności 3, otrzymujemy:

$$\mathbb{E}[Z] \leq (n - 1) \left(1 + \frac{1}{p}\right).$$

Faktyczna wartość oczekiwana wynosi  $\frac{n-1}{p}$ , więc oszacowanie jest dość dokładne. Z kolei dla rodziny grafów  $S_n$  mamy  $h = 1$  oraz  $n + 1$  wierzchołków, stąd:

$$\mathbb{E}[Z] \leq 1 + \frac{\log(n) + 1}{\log\left(\frac{1}{1-p}\right)}.$$

Ponownie otrzymujemy zaskakująco dobre przybliżenie. Należy zwrócić uwagę, że zarówno  $P_n$  jak i  $S_n$  są grafami rzadkimi. Analizując grafy pełnej przekonamy się, że dla grafów gęstych oszacowanie to nie będzie skuteczne.

## 4.7 Grafy cykliczne

Przejdźmy teraz do grafów cyklicznych. W rozważaniach dla trójkąta, to jest  $C_3$ , mogliśmy zauważyć, że cykl w tym grafie sprawiał trudności. W ogólnym przypadku nie jest lepiej. Rozważmy graf  $C_n$ . Niech źródłem będzie wierzchołek  $n$ . Ustalmy wierzchołek  $v \in \{1, 2, \dots, n-1\}$ . Niech  $a = \min\{v, n-v\}$  oraz  $b = \max\{v, n-v\}$ . Oczywiście  $a \leq b$ . Od źródła do tego wierzchołka są dwie ścieżki: jedna o długości  $a$ , druga o długości  $b$ . Propagacja rozchodzi się po nich równolegle i niezależnie. Dla  $j \in \{1, 2, \dots, n-1\}$  położymy  $N_j \sim \text{NegBin}(j, p)$ . Zmienne te są niezależne. Mamy wtedy

$$X_v \sim \min\{N_a, N_b\}.$$

Niech  $F_j(t)$  będzie dystrybuantą zmiennej  $N_j$ . Z Faktu 2 mamy

$$\mathbb{P}[X_v \leq t] = 1 - (1 - F_a(t)) \cdot (1 - F_b(t)).$$

Nie ma co liczyć na wyznaczenie eleganckiej postaci na PMF czy CDF dla  $X_v$ . Postaramy się więc przybliżyć wartość oczekiwana dla dużych  $n$ . Kładziemy  $A \sim \mathcal{N}(\mu_a, \sigma_a^2)$ ,  $B \sim \mathcal{N}(\mu_b, \sigma_b^2)$  gdzie

$$\mu_a = \frac{a}{p}, \quad \sigma_a^2 = \frac{aq}{p^2}, \quad \mu_b = \frac{b}{p}, \quad \sigma_b^2 = \frac{bq}{p^2}.$$

Z centralnego twierdzenia granicznego możemy przybliżyć  $N_a \approx A$ ,  $N_b \approx B$ . Zatem mamy  $\mathbb{E}[X_v] \approx \mathbb{E}[\min\{A, B\}]$  oraz

$$\mathbb{E}[\min\{A, B\}] = \mathbb{E}\left[\frac{A + B - |A - B|}{2}\right] = \frac{\mathbb{E}[A] + \mathbb{E}[B] - \mathbb{E}[|A - B|]}{2}.$$

Położmy  $C = A - B$ . Korzystając z Faktu 5 mamy  $C \sim \mathcal{N}(\mu_a - \mu_b, \sigma_a^2 + \sigma_b^2)$ . Oznaczmy  $\eta = \mu_a - \mu_b$  oraz  $\xi = \sqrt{\sigma_a^2 + \sigma_b^2}$ . Potrzebujemy teraz następującego lematu:

**Lemat 2.** *Niech  $X \sim \mathcal{N}(\mu, \sigma^2)$ . Wtedy*

$$\mathbb{E}[|X|] = 2\sigma \cdot \varphi\left(\frac{\mu}{\sigma}\right) + \mu \cdot (2\Phi\left(\frac{\mu}{\sigma}\right) - 1).$$

*Dowód.*

$$\mathbb{E}[|X|] = \int_{-\infty}^{\infty} \frac{|x|}{\sigma} \varphi\left(\frac{x - \mu}{\sigma}\right) dx = \int_0^{\infty} \frac{x}{\sigma} \varphi\left(\frac{x - \mu}{\sigma}\right) dx - \int_{-\infty}^0 \frac{x}{\sigma} \varphi\left(\frac{x - \mu}{\sigma}\right) dx.$$

Oznaczmy  $c = \frac{\mu}{\sigma}$  oraz podstawmy  $z = \frac{x - \mu}{\sigma}$ . Zatem  $x = \mu + \sigma z$ ,  $dx = \sigma dz$ . Dla  $x > 0$  mamy  $z > -c$  zaś dla  $x < 0$  mamy  $z < -c$ . Otrzymujemy więc

$$\begin{aligned} & \int_{-c}^{\infty} (\mu + \sigma z) \varphi(z) dz - \int_{-\infty}^{-c} (\mu + \sigma z) \varphi(z) dz \\ &= \mu \int_{-c}^{\infty} \varphi(z) dz + \sigma \int_{-c}^{\infty} z \varphi(z) dz - \mu \int_{-\infty}^{-c} \varphi(z) dz - \sigma \int_{-\infty}^{-c} z \varphi(z) dz \\ &= \mu \left( \int_{-c}^{\infty} \varphi(z) dz - \int_{-\infty}^{-c} \varphi(z) dz \right) + \sigma \left( \int_{-c}^{\infty} z \varphi(z) dz - \int_{-\infty}^{-c} z \varphi(z) dz \right) \\ &= \mu \left( \int_{-\infty}^{\infty} \varphi(z) dz - 2 \int_{-\infty}^{-c} \varphi(z) dz \right) + \sigma \left( -\varphi(z)|_{-c}^{\infty} + \varphi(z)|_{-\infty}^{-c} \right) \\ &= \mu \cdot (1 - 2\Phi(-c)) + \sigma \cdot (-\varphi(\infty) + \varphi(-c) + \varphi(-c) - \varphi(-\infty)) \\ &= \mu \cdot (2\Phi(c) - 1) + 2\sigma \cdot \varphi(c). \end{aligned}$$

gdzie skorzystaliśmy z tożsamości  $\Phi(-x) = 1 - \Phi(x)$ ,  $\varphi(-x) = \varphi(x)$ ,  $\varphi(\pm\infty) = 0$ ,  $\int_{-\infty}^{\infty} \varphi(x) dx = 1$  oraz  $\int x \varphi(x) dx = -\varphi(x)$ .  $\square$

Z Lematu 2 dostajemy  $\mathbb{E}[C] = 2\xi \cdot \varphi\left(\frac{\eta}{\xi}\right) + \eta \cdot (2\Phi\left(\frac{\eta}{\xi}\right) - 1)$ . Ostatecznie

$$\begin{aligned} \mathbb{E}[X_v] &\approx \frac{1}{2} \left( \mu_a + \mu_b - 2\xi \varphi\left(\frac{\eta}{\xi}\right) - \eta \left( 2\Phi\left(\frac{\eta}{\xi}\right) - 1 \right) \right) \\ &= \frac{\mu_a + \mu_b}{2} - \xi \varphi\left(\frac{\eta}{\xi}\right) - (\mu_a - \mu_b) \left( \Phi\left(\frac{\eta}{\xi}\right) - \frac{1}{2} \right) \\ &= \mu_a \left( 1 - \Phi\left(\frac{\eta}{\xi}\right) \right) + \mu_b \Phi\left(\frac{\eta}{\xi}\right) - \eta \varphi\left(\frac{\eta}{\xi}\right). \end{aligned}$$

Przenalizujmy teraz zachowanie asymptotyczne otrzymanego wyrażenia. Skoro  $a + b = n$  to niech  $a = rn$ ,  $b = (1 - r)n$  dla pewnego  $r \in (0; 1)$ . Dalej

$$\frac{\eta}{\xi} = \frac{\mu_a - \mu_b}{\sqrt{\sigma_a^2 + \sigma_b^2}} = \frac{\frac{a}{p} - \frac{b}{p}}{\sqrt{\frac{aq}{p^2} + \frac{bq}{p^2}}} = \frac{(2r - 1)\sqrt{n}}{\sqrt{q}}.$$

Musimy rozważyć dwa przypadki. Jeśli  $a < b$ , co za tym idzie  $r < \frac{1}{2}$  to  $\frac{\eta}{\xi} \rightarrow -\infty$  wraz z  $n \rightarrow \infty$ . Wtedy też  $\varphi(\frac{\eta}{\xi}) \rightarrow 0$  oraz  $\Phi(\frac{\eta}{\xi}) \rightarrow 0$  a więc  $\mathbb{E}[X_v] \rightarrow \frac{a}{p}$ . Zaś gdy  $a = b$  to  $r = \frac{1}{2}$  jak i  $\frac{\eta}{\xi} = 0$ . Wiemy, że  $\varphi(0) = \frac{1}{\sqrt{2\pi}}$  oraz  $\Phi(0) = \frac{1}{2}$ . Wstawiając otrzymamy  $\mathbb{E}[X_v] \rightarrow \frac{n}{2p} - \frac{\sqrt{np}}{p\sqrt{2\pi}}$ . Podsumowując mamy następujący wynik:

$$\mathbb{E}[X_v] \sim \frac{\min\{v, n - v\}}{p}.$$

Jest to całkowicie zgodne z intuicją. Wierzchołki w grafie  $C_n$  zachowują się podobnie jak w grafach  $P_n$ .

W celu wyznaczenie rozkładu  $Y_t$  dokonajmy obserwacji, że gdy dwie drogi zarażania spotkają się to propagacja dobiera końca. Każda z tych dróg jak w przypadku grafu ścieżkowego ma rozkład dwumianowy. Możemy zapisać zatem

$$Y_t \sim \min\{n, 1 + L_t + R_t\}, \quad L_t, R_t \sim \text{Bin}(t, p).$$

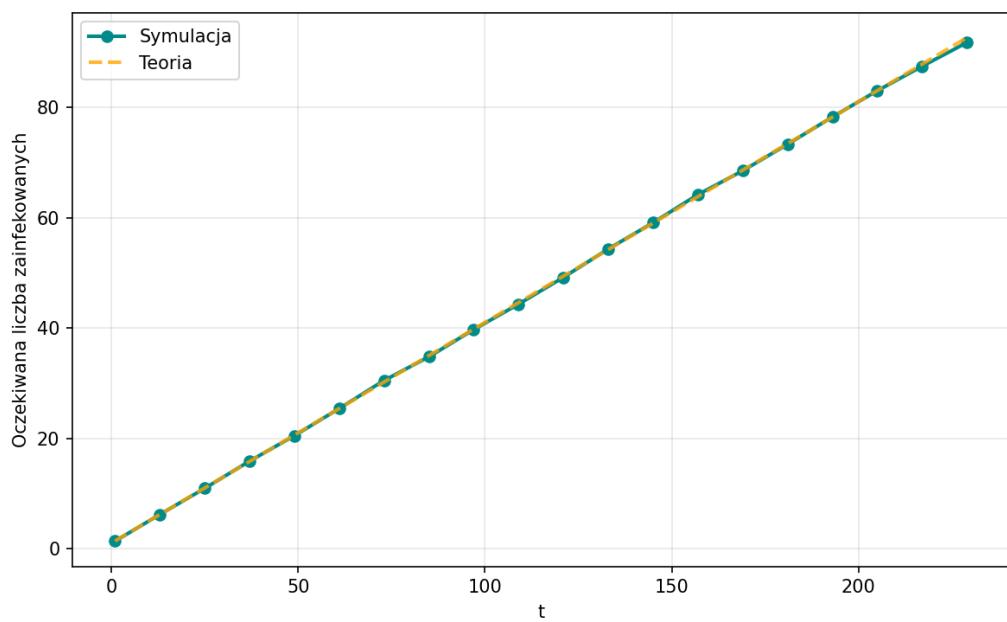
Z Faktu 3 mamy  $L_t + R_t \sim \text{Bin}(2t, p)$ . Widzimy zatem, że rozkład  $Y_t$  dla grafu  $C_n$  pokrywa się ze zmienną  $Y_{2t}$  dla grafów typu  $P_n$ . Z wcześniejszego wyniku dla grafów ścieżek dostajemy

$$\lim_{n \rightarrow \infty} \mathbb{E}[Y_t] = 1 + 2tp.$$

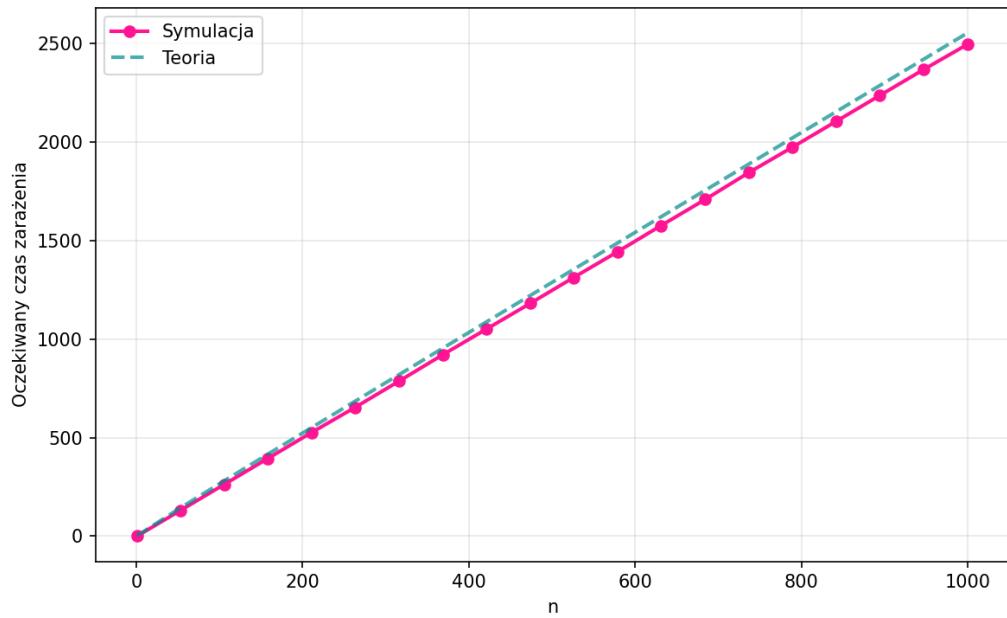
Teraz możemy wyznaczyć rozkład i wartość oczekiwana zmiennej  $Z$ . Dla  $n$  parzystego najdalej oddalony wierzchołek od źródła to  $\frac{n}{2}$  a dla  $n$  nieparzystego to  $\lfloor \frac{n+1}{2} \rfloor$ . Asymptotycznie nie ma to znaczenia, możemy przyjąć  $v = \frac{n}{2}$ . Mamy więc  $Z = \max_{v \in V} X_v = X_{\frac{n}{2}}$ . Stąd

$$\mathbb{E}[Z] \approx \mathbb{E}[X_{\frac{n}{2}}] \approx \frac{n}{2p} - \frac{\sqrt{np}}{p\sqrt{2\pi}}.$$

Również intuicyjny wynik. Aby się upewnić czy nie przesadziliśmy z szacowaniem zwróciły się ku symulacji. Dla  $p = 0.2$ ,  $n \in \{3, 4, \dots, 1000\}$  policzmy wartość oczekowaną całkowitego zarażenia po razy 2000 razy. Z wykresów (zob. Wykresie 5, Wykresie 6) widzimy, że empiryczny wynik pokrywa się asymptotycznie z teoretycznym.



Rysunek 5:  $\mathbb{E}[Y_t]$  dla  $C_n$  w funkcji  $t$



Rysunek 6:  $\mathbb{E}[Z]$  dla  $C_n$  w funkcji  $n$

## 4.8 Grafy pełne

Graf pełny  $K_n$  intuicyjnie powinien mieć najszybszą propagację bo przecież ma maksymalną liczbę krawędzi. Za źródło możemy przyjąć dowolny wierzchołek  $s \in V$  ze względu na symetrię. Początkowo rozkład  $X_v$  pokrywa się z rozkładem gwiazdy natomiast w każdej kolejnej rundzie mocno się komplikuje. Nie mamy co liczyć na jakiekolwiek sensowne wyznaczenie rozkładu  $X_v$ . Podejdźmy do problemu na razie heurystycznie. Zauważmy, że jeśli  $Y_1 = a$  to rozkład zmiennej  $Y_2$  wynosi  $Y_2 = a + B$  dla  $B \sim \text{Bin}(n - a, 1 - q^n)$ . Zatem

$$\mathbb{E}[Y_2 | Y_1 = a] = n \cdot (1 - q^a) + aq^a.$$

Mamy  $Y_1 - 1 \sim \text{Bin}(n - 1, p)$  oraz  $\mathbb{E}[Y_1] = 1 + (n - 1)p$ . Możemy również założyć, że również  $a \approx \mathbb{E}[Y_1]$  a co za tym idzie

$$\mathbb{E}[Y_2 | Y_1 = a] \approx n(1 - q^{1+(n-1)p}) + (1 + (n - 1)p)q^{1+(n-1)p}.$$

Jeśli  $n \rightarrow \infty$  to wyrażenie to jest bliskie  $n$ . Spodziewamy się zatem, że zaledwie po dwóch rundach cały graf  $K_n$  będzie zainfekowany. Zweryfikujmy teraz ten heurystyczny argument symulacją w Pythonie. Ustalmy  $p = 0.2$  i dla  $n \in \{2, 3, \dots, 1000\}$  odpalmy propagację. Widzimy, (zob. Wykresie 7, Wykresie 8) że dla  $n > 200$  mamy  $\mathbb{E}[Z] \approx 2$ . Możemy więc wysunąć hipotezę: Dla grafu  $K_n$  mamy:

$$\lim_{n \rightarrow \infty} \mathbb{E}[Z] = 2.$$

Postarajmy się ją teraz udowodnić. Żeby to zrobić najpierw wyznaczmy asymptotykę  $\mathbb{E}[Y_2]$ . Oznaczamy  $U = Y_1 = 1 + W$  gdzie  $W \sim \text{Bin}(n - 1, p)$ . Z prawa całkowitej wartości oczekiwanej mamy

$$\mathbb{E}[Y_2] = n \cdot (1 - \mathbb{E}[q^U]) + \mathbb{E}[Uq^U].$$

Musimy wyznaczyć  $\mathbb{E}[q^U]$  jak i  $\mathbb{E}[Uq^U]$ . Korzystając z Faktu 6 dostajemy

$$\mathbb{E}[q^U] = \mathbb{E}[q^{1+W}] = q \cdot \mathbb{E}[q^W] = q(q + pq)^{n-1} = q^n(1 + p)^{n-1}.$$

Dla drugiej wartości mamy zaś

$$\begin{aligned} \mathbb{E}[Uq^U] &= \mathbb{E}[(1 + W)q^{1+W}] = q(\mathbb{E}[q^W] + \mathbb{E}[Wq^W]) \\ &= q(q^{n-1}(1 + p)^{n-1} + (n - 1)pq^{n-1}(1 + p)^{n-2}) \\ &= q^n(1 + p)^{n-2}(1 + p + (n - 1)p) = q^n(1 + p)^{n-2}(1 + np). \end{aligned}$$

Podstawiając przed chwilą wyrażenia wzory do wzoru na  $\mathbb{E}[Y_2]$  dostaniemy

$$\begin{aligned} \mathbb{E}[Y_2] &= n - nq^n(1 + p)^{n-1} + q^n(1 + p)^{n-2}(1 + np) = \\ &= n - (n - 1)q^n(1 + p)^{n-2} = n - (n - 1)(1 + p)^{-2}(1 - p^2)^n. \end{aligned}$$

Położmy  $\varepsilon_n = (n-1)(1+p)^{-2}(1-p^2)^n$ . Wtedy  $\mathbb{E}[Y_2] = n - \varepsilon_n$ . Z nierówności Markowa (5) otrzymujemy  $\mathbb{P}[Z \geq 3] = \mathbb{P}[n - Y_2 \geq 1] \leq \mathbb{E}[n - Y_2] = \varepsilon_n$ . Dalej zauważmy, że  $\mathbb{P}[Z = 1] = p^{n-1}$  bo wszystkie próby zarażenia w rundzie pierwszej musiałby się powieść. Ograniczmy teraz z dwóch stron  $\mathbb{E}[Z]$ . Z dołu mamy

$$\mathbb{E}[Z] = \sum_{k=1}^{\infty} \mathbb{P}[Z \geq k] \geq \mathbb{P}[Z \geq 1] + \mathbb{P}[Z \geq 2] = 1 + 1 - p^{n-1} = 2 - p^{n-1}.$$

Zajmijmy się teraz oszacowaniem górnym. Zauważmy, że graf  $K_n$  zawiera  $P_n$  jako podgraf. Ustalmy jeden z tych podgrafów. Niech  $Z'$  będzie zmienną losową czasu całkowitego zarażenia dla tego podgrafa. Z Twierdzenia 2 mamy  $Z \leq Z'$  a co za tym idzie  $\mathbb{E}[Z^2] \leq \mathbb{E}[(Z')^2]$ . Przypomnijmy, że  $Z' \sim \text{NegBin}(n-1, p)$  a więc  $\mathbb{E}[(Z')^2] = \frac{(n-1)^2 + (n-1)q}{p^2}$ . Wnioskujemy, że

$$\begin{aligned} \mathbb{E}[Z] &= \sum_{k=1}^{\infty} \mathbb{P}[Z \geq k] = \mathbb{P}[Z \geq 1] + \mathbb{P}[Z \geq 2] + \sum_{k=3}^{\infty} \mathbb{P}[Z \geq k] \\ &= 1 + 1 - p^{n-1} + \mathbb{E}[Z \cdot \mathbf{1}_{Z \geq 3}] \leq 2 - p^{n-1} + \sqrt{\mathbb{E}[Z^2]} \sqrt{\mathbb{E}[\mathbf{1}_{Z \geq 3}]} \\ &\leq 2 - p^{n-1} + \sqrt{\mathbb{E}[(Z')^2]} \sqrt{\mathbb{P}[Z \geq 3]} \\ &\leq 2 - p^{n-1} + \sqrt{\frac{(n-1)^2 + (n-1)q}{p^2}} \sqrt{\varepsilon_n}. \end{aligned}$$

gdzie wykorzystaliśmy nierówność Cauchy'ego-Schwarza (6). Ostatecznie dostajemy

$$2 - p^{n-1} \leq \mathbb{E}[Z] \leq 2 - p^{n-1} + \sqrt{\frac{(n-1)^2 + (n-1)q}{p^2}} \sqrt{\varepsilon_n},$$

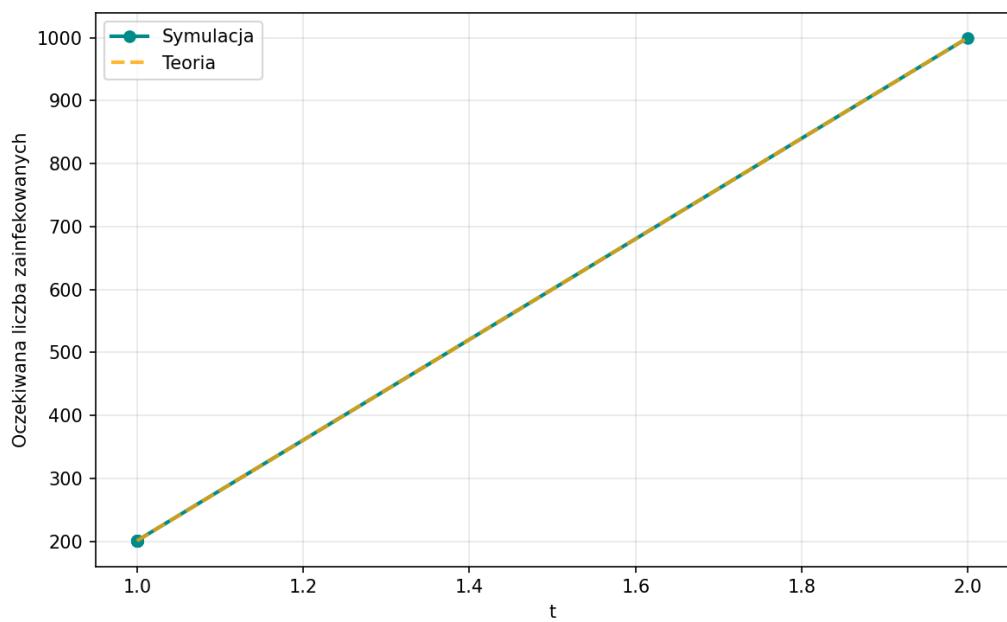
a zatem

$$\lim_{n \rightarrow \infty} \mathbb{E}[Z] = 2.$$

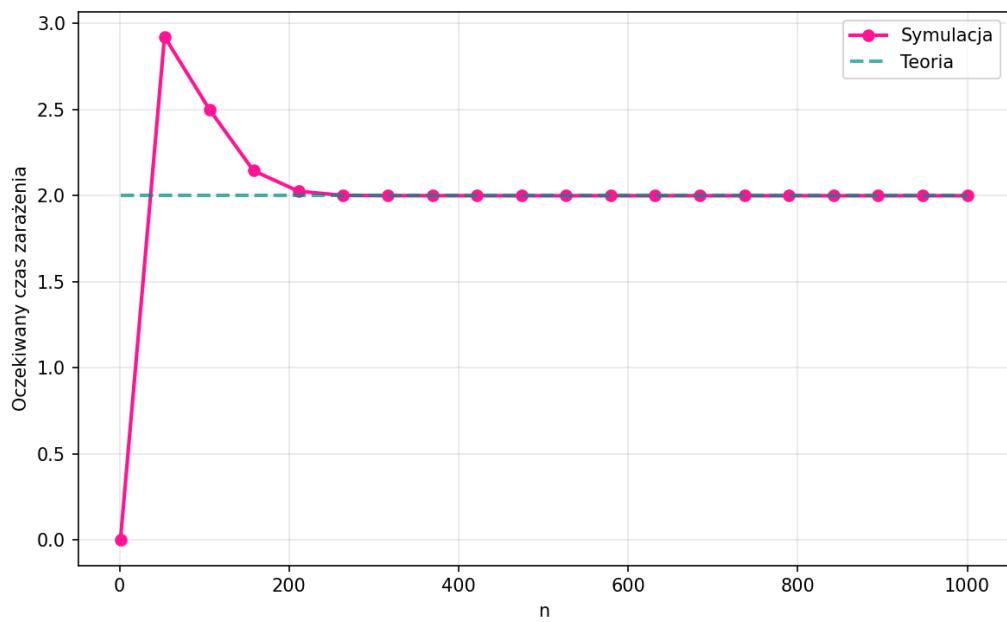
Jeżeli zaledwie po dwóch rundach cały graf jest poinformowany to rozkłady  $X_v$  czy  $Y_t$  nie są dla nas istotne. Spójrzmy jeszcze na oszacowanie, które otrzymamy stosując Twierdzenie 3 dla grafu pełnego. Wynosi ono

$$1 + \frac{\log(n-1) + 1}{\log(\frac{1}{1-p})}.$$

Ograniczenie to zdaje się nie być za dobre czego przyczyną jest zapewne fakt, że  $K_n$  jest grafem gęstym.



Rysunek 7:  $E[Y_t]$  dla  $K_n$  w funkcji  $t$



Rysunek 8:  $E[Z]$  dla  $K_n$  w funkcji  $n$

## 4.9 Drzewa

Rozważmy drzewo  $G = (V, E)$  oraz ustalony wierzchołek początkowy  $s \in V$ , który traktujemy jako korzeń drzewa. Dla  $v \in V$  oznaczmy  $d_v = d(s, v)$ . Ustalmy  $v \in V$ . Skoro  $G$  jest drzewem to istnieje dokładnie jedna ścieżka od  $s$  do  $v$ , powiedzmy  $s, v_1, \dots, v_k, v$ . Ponieważ infekcja rozprzestrzenia się od korzenia  $s$  wzduż krawędzi drzewa, każde zakażenie wymaga sukcesu w niezależnym doświadczeniu Bernoulliego o prawdopodobieństwie  $p$ . W konsekwencji, aby infekcja dotarła z  $s$  do  $v$ , musi wystąpić  $d_v$  kolejnych sukcesów. Zatem rozkład  $X_v$  pokrywa się z rozkładem tej zmiennej dla grafu  $P_{d_v+1}$  na wierzchołkach  $\{s, v_1, \dots, v_k, v\}$ . Stąd

$$X_v \sim \text{NegBin}(d_v, p),$$

oraz

$$\mathbb{E}[X_v] = \frac{d_v}{p}.$$

**Lemat 3.** *Dla dowolnego  $t \in \mathbb{N}$  wartość oczekiwana zmiennej  $Y_t$  wyraża się wzorem*

$$\mathbb{E}[Y_t] = \sum_{v \in V} \mathbb{P}[X_v \leq t].$$

*Dowód.* Mamy  $Y_t = |\{v \in V : X_v \leq t\}|$  zatem  $Y_t = \sum_{v \in V} \mathbf{1}_{\{X_v \leq t\}}$ . Nakładając na tą równość operator  $\mathbb{E}$  otrzymujemy:

$$\mathbb{E}[Y_t] = \mathbb{E}\left[\sum_{v \in V} \mathbf{1}_{\{X_v \leq t\}}\right] = \sum_{v \in V} \mathbb{E}[\mathbf{1}_{\{X_v \leq t\}}] = \sum_{v \in V} \mathbb{P}[X_v \leq t].$$

□

Przejdźmy teraz do obliczania średniej liczby zainfekowanych wierzchołków w czasie  $t$ . Oznaczmy przez  $F(t; m, p)$  dystrybuantę zmiennej o rozkładzie NegBin( $m, p$ ). Z Lematu 3 otrzymujemy

$$\mathbb{E}[Y_t] = \sum_{v \in V} F(t; d_v, p).$$

Położymy  $a_j = |\{v \in V : d_v = j\}|$  dla  $0 \leq j \leq h$ . Wtedy

$$\mathbb{E}[Y_t] = \sum_{j=0}^h a_j \cdot F(t; j, p).$$

Ponadto gdy  $t < j \leq h$  to  $F(t; j, p)$ , bo żaden wierzchołek w odległości od korzenia większej niż liczba rund nie może zostać zarażony. Możemy więc zmniejszyć granice sumowania

$$\mathbb{E}[Y_t] = \sum_{j=0}^{\min\{h,t\}} a_j \cdot F(t; j, p).$$

Oszacujmy teraz średni czas całkowity czas propagacji drzewa. Niech  $L = \{u_1, \dots, u_m\}$  będzie zbiorem liści w  $G$ . Wtedy mamy  $Z = \max_{u \in L} X_u$ . Zauważmy, że  $\epsilon(s) = \max_{u \in L} d_u$  i jest to wysokość drzewa. Oznaczmy ją przez  $h$ . Z nierówności Jense'a (7) otrzymujemy

$$\mathbb{E}[Z] = \mathbb{E}[\max_{u \in L} X_u] \geq \max_{u \in L} \mathbb{E}[X_u] = \max_{u \in L} \frac{d_u}{p} = \frac{h}{p}.$$

Aby ograniczyć  $\mathbb{E}[Z]$  z góry skorzystamy z Twierdzenia 3:

$$\mathbb{E}[Z] \leq h + h \cdot \frac{\log(\frac{n-1}{h}) + 1}{\log(\frac{1}{1-p})}.$$

Ograniczenia te są różnych rzędów wielkości. Jednakże nie da się ich poprawić dla ogólnego drzewa znając tylko liczbę jego wierzchołków i wysokość. Ustalmy  $n \in \mathbb{N}_+$  oraz  $h \in \{1, 2, \dots, n-1\}$  i poszukajmy drzewa o  $n$  wierzchołkach i wysokości  $h$  osiągających zarówno dolne jak i górne ograniczenie na  $\mathbb{E}[Z]$ . Dla dolnej nierówności możemy wziąć drzewo składające się ze ścieżki długości  $h$  oraz  $n-1-h$  liści bezpośrednio przy korzeniu. Wtedy  $\mathbb{E}[Z] \approx \frac{h}{p}$ . Aby znaleźć drzewo osiągające górne ograniczenie musimy wrócić do dowodu Twierdzenia 3. Udowadniając granicę na wartość oczekiwana korzystamy z trzech nierówności. Pierwsza z nich to Nierówność 2. Jest ona bardzo ciasna a ponadto nie zależy od grafu. Druga z nich to nierówność między średnią arytmetyczną a geometryczną (4). Aby uzyskać równość potrzebujemy mieć  $a_1 = \dots = a_h$ . Czyli innymi słowy, nasze drzewo ma tyle samo węzłów na każdej głębokości. Położymy  $a_1 = b$ . Wtedy  $hb = n-1$  a więc  $b = \lfloor \frac{n-1}{h} \rfloor$ . Na koniec zostaje nierówność wynikająca z Lematu 1. Sam lemat daje nierówność, której nie da się poprawić, co wiemy poprzez analizę dla grafów gwiazd. Lecz dla drzewa będzie ona najmniej luźna, jeżeli każdy wierzchołek w warstwie  $A_j$  będzie miał dokładnie jedną krawędź łączącą go z wierzchołkiem w warstwie  $A_{j+1}$ , gdzie  $0 \leq j \leq h-1$ . Zatem drzewo składa się z korzenia oraz  $b$  rozłącznych ścieżek, każda o długości  $h$ . I taki graf osiąga ograniczenie górne na  $\mathbb{E}[Z]$ . Widzimy zatem, że nasze ograniczenia nie są do poprawienia bez dodatkowych parametrów grafu. Podsumowując możemy następująco szacować

przewidywany czas całkowitego poinformowania drzewa o wysokości  $h$ :

$$\frac{h}{p} \leq \mathbb{E}[Z] \leq h + h \cdot \frac{\log(\frac{n-1}{h}) + 1}{\log(\frac{1}{1-p})}.$$

## Rozdział 5

# Rozkłady wymierające

W celu ułatwienia analizy modeli **SIR** oraz **SIS** wprowadzimy nowe rozkłady Prawdopodobieństwa uwzględniające możliwość przerwania propagacji. Te zmienne losowe będą przyjmować wartość nieskończoność w sytuacji, w której nastąpi przerwanie przed pożądanym rezultatem. Jako że następuje to z dodatnim prawdopodobieństwem,  $\mathbb{P}[X = \infty] > 0$ , to ich wartości oczekiwane wynosić będą nieskończoność,  $\mathbb{E}[X] = \infty$ . Interesować nas więc będzie wartość oczekiwana warunkowa przy warunku, że wartość zmiennej losowej jest skończona a więc  $\mathbb{E}[X|X < \infty]$ .

### 5.1 Rozkład umierający geometryczny

Rozkład umierający geometryczny jest wariantem rozkładu geometrycznego, w którym proces może zostać przerwany z prawdopodobieństwem  $\alpha \in (0; 1)$  po każdej próbie. Zmienna  $X$  ma rozkład umierający geometryczny, jeżeli opisuje liczbę prób Bernoulliego potrzebnych do uzyskania pierwszego sukcesu w przypadku, w którym sukces nastąpi przed przetrwaniem eksperymentu. Jeśli natomiast proces zostanie zabity szybciej niż pierwsza porażka to wtedy  $X = \infty$ . Dla wygody oznaczmy  $q = 1 - p$ ,  $\beta = 1 - \alpha$ . Aby sukces nastąpił po  $k$  rundach to potrzebujemy  $k - 1$  niepowodzeń próby jak i jej przerwania a następnie sukcesu. A zatem

$$\mathbb{P}[X = k] = p(q\beta)^{k-1}, \quad k \in \mathbb{N}_+.$$

Dystrybuanta jest więc równa:

$$\mathbb{P}[X \leq t] = \sum_{k=0}^t \mathbb{P}[X = k] = \sum_{k=0}^t p(q\beta)^{k-1} = p \frac{1 - (q\beta)^t}{1 - q\beta}.$$

Ponadto

$$\mathbb{P}[X < \infty] = \sum_{k=0}^{\infty} \mathbb{P}[X = k] = \sum_{k=0}^{\infty} p(q\beta)^{k-1} = \frac{p}{1 - q\beta}.$$

Suma wszystkich prawdopodobieństw wynosi 1. A więc

$$\mathbb{P}[X = \infty] = 1 - \frac{p}{1 - q\beta} = \frac{q\alpha}{1 - q\beta}.$$

Wartość oczekiwana wynosi nieskończoność. Natomiast

$$\begin{aligned}\mathbb{E}[X|X < \infty] &= \frac{1}{\mathbb{P}[X < \infty]} \sum_{t=1}^{\infty} t \cdot \mathbb{P}[X_v = t] = \frac{1 - q\beta}{p} \sum_{t=1}^{\infty} t(q\beta)^{t-1} p \\ &= (1 - q\beta) \cdot \frac{1}{(1 - q\beta)^2} = \frac{1}{1 - q\beta}.\end{aligned}$$

Oznaczamy  $X \sim \text{KGeo}(p, \alpha)$ .

## 5.2 Rozkład umierający dwumianowy

Rozkład umierający dwumianowy jest wariantem rozkładu dwumianowego, w którym doświadczenia mogą zostać przerwane z prawdopodobieństwem  $\alpha \in (0; 1)$  po każdej próbie. Zmienna  $X$  ma rozkład umierający dwumianowy, jeżeli opisuje liczbę sukcesów podczas  $n$  prób Bernoulliego, gdzie po każdej próbie proces doświadczeń może zostać zabity. Dla wygody oznaczmy  $q = 1 - p$ ,  $\beta = 1 - \alpha$ . Niech  $F$  będzie zmienną określającą czas porażki. Dla  $1 \leq j \leq n - 1$  mamy  $\mathbb{P}[F = j] = \alpha\beta^{j-1}$  oraz  $\mathbb{P}[F = n] = \beta^{n-1}$ . Jeśli proces zostanie zabity po  $j$  rundach to wtedy  $X$  ma rozkład  $\text{Bin}(j, p)$ . A zatem korzystając z prawdopodobieństwa całkowitego dla  $k \in \{0, 1, \dots, n\}$  mamy

$$\begin{aligned}\mathbb{P}[X = k] &= \sum_{j=k}^n \mathbb{P}[X = k|F = j] \cdot \mathbb{P}[F = j] = \\ &= \binom{n}{k} p^k q^{n-k} \beta^{n-1} + \alpha \sum_{j=k}^{n-1} \binom{j}{k} p^k q^{j-k} \beta^{j-1}.\end{aligned}$$

Dalej mamy  $\mathbb{E}[X|F = j] = jp$  oraz

$$\begin{aligned}\mathbb{E}[F] &= \sum_{j=1}^n j \cdot \mathbb{P}[F = j] = n\beta^{n-1} + \sum_{j=1}^{n-1} j\alpha\beta^{j-1} = \\ &= n\beta^{n-1} + \alpha \frac{1 - n\beta^{n-1} + (n-1)\beta^n}{(1-\beta)^2} = \\ &= \frac{1}{\alpha}(n\alpha\beta^{n-1} + 1 - n\beta^{n-1} + n\beta^n - \beta^n) = \\ &= \frac{1}{\alpha}(1 - \beta^n + n\beta^{n-1}(\alpha + \beta - 1)) = \frac{1 - \beta^n}{\alpha}.\end{aligned}$$

Z prawa całkowitej wartości oczekiwanej mamy

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|F]] = p\mathbb{E}[F] = \frac{p}{\alpha}(1 - \beta^n).$$

Oznaczamy  $X \sim \text{KBin}(n, p, \alpha)$ .

### 5.3 Rozkład umierający ujemny dwumianowy

Rozkład umierający ujemny dwumianowy jest wariantem rozkładu ujemnego dwumianowego, w którym doświadczenia mogą zostać przerwane z prawdopodobieństwem  $\alpha \in (0; 1)$  po każdej próbie. Zmienna  $X$  ma rozkład umierający ujemny dwumianowy, jeżeli opisuje liczbę prób Bernoulliego potrzebnych do uzyskania  $m$  sukcesów próbach Bernoulliego, gdzie po każdej próbie proces może się zakończyć. Jeśli proces zostanie zabity szybciej niż zajdzie  $m$  sukcesów to przyjmujemy  $X = \infty$ . Dla wygody oznaczmy  $q = 1 - p$ ,  $\beta = 1 - \alpha$ . Alternatywnie istnieją niezależne zmienne  $Y_1, Y_2, \dots, Y_m \sim \text{KGeo}(p, \alpha)$  takie, że  $X = Y_1 + Y_2 + \dots + Y_m$ . Aby wyznaczyć rozkład  $X$  ustalmy  $k \geq m$  i niech  $K = \{\mathbf{y} \in \mathbb{N}_+^m : y_1 + \dots + y_m = k\}$ . Wtedy

$$\begin{aligned}\mathbb{P}[X = k] &= \sum_{\mathbf{y} \in K} \mathbb{P}[Y_1 = y_1, \dots, Y_m = y_m] = \sum_{\mathbf{y} \in K} \prod_{j=1}^m \mathbb{P}[Y_j = y_j] \\ &= \sum_{\mathbf{y} \in K} \prod_{j=1}^m p(q\beta)^{y_j-1} = \sum_{\mathbf{y} \in K} p^m (q\beta)^{y_1+\dots+y_m-m} \\ &= \sum_{\mathbf{y} \in K} p^m (q\beta)^{k-m} = |K| \cdot p^m (q\beta)^{k-m}.\end{aligned}$$

Jeśli  $\mathbf{y} \in K$  to podstawiamy  $\mathbf{y} = \mathbf{z} + 1$  dla  $\mathbf{z} \in \mathbb{N}^m$ . Wtedy  $z_1 + \dots + z_m = k - m$  a więc  $|K| = \binom{k-1}{m-1}$ . Zatem

$$\mathbb{P}[X = k] = \binom{k-1}{m-1} p^m (q\beta)^{k-m}, \quad k \geq m.$$

Dalej mamy

$$\begin{aligned}
\mathbb{P}[X < \infty] &= \sum_{k=m}^{\infty} \mathbb{P}[X = k] = \sum_{k=m}^{\infty} \binom{k-1}{m-1} p^m (q\beta)^{k-m} = \\
&= p^m (q\beta)^{-m} \sum_{k=m}^{\infty} \binom{k-1}{m-1} (q\beta)^k = p^m (q\beta)^{-m} \frac{(q\beta)^m}{(1-q\beta)^m} \\
&= \frac{p^m}{(1-q\beta)^m}.
\end{aligned}$$

co daje nam

$$\mathbb{P}[X = \infty] = 1 - \frac{p^m}{(1-q\beta)^m}.$$

Jeśli chodzi o wartość oczekiwana to

$$\begin{aligned}
\mathbb{E}[X|X < \infty] &= \frac{1}{\mathbb{P}[X < \infty]} \sum_{k=m}^{\infty} k \cdot \mathbb{P}[X = k] \\
&= \frac{(1-q\beta)^m}{p^m} \sum_{k=m}^{\infty} k \binom{k-1}{m-1} p^m (q\beta)^{k-m} = \\
&= \frac{(1-q\beta)^m}{(q\beta)^m} \sum_{k=m}^{\infty} k \binom{k-1}{m-1} (q\beta)^k = \\
&= \frac{(1-q\beta)^m}{(q\beta)^m} \cdot \frac{m(q\beta)^m}{(1-q\beta)^{m+1}} = \frac{m}{1-q\beta}.
\end{aligned}$$

Oznaczamy  $X \sim \text{KNegBin}(m, p, \alpha)$ .

## Rozdział 6

# Analiza modelu SIR

### 6.1 Dwa wierzchołki, jedna krawędź

W celu oswojenia się z bardziej skomplikowanym modelem, jakim jest **SIR**, przeanalizujmy graf o jednej krawędzi. Niech  $V = \{u, v\}$  oraz niech  $u$  będzie wierzchołkiem startowym. Oczywiście  $X_u = 0$ . Pierwsza runda jest identyczna jak w modelu **SI**, a więc  $\mathbb{P}[X_v = 1] = p$ . Jeżeli węzeł  $v$  nie zostanie poinformowany w rundzie pierwszej, a propagacja nie wygaśnie, to sytuacja się powtórzy. Zachodzi to z prawdopodobieństwem  $q\beta$ . Aby  $X_v = t$ , potrzebujemy, by ten cykl nastąpił  $t - 1$  razy. Widzimy zatem, że  $X_v$  ma rozkład umierający geometryczny,

$$X_v \sim \text{KGeo}(p, \alpha),$$

a co za tym idzie

$$\mathbb{E}[X_v | X_v < \infty] = \frac{1}{1 - q\beta}.$$

Z założeń modelu  $p, \alpha \in (0; 1)$ , a zatem  $\mathbb{P}[X_v = \infty] \neq 0$ . Jest cecha kontrastująca **SIR** względem **SI**.

Przyjrzymy się teraz zmiennej  $Y_t$ . Zauważmy, że  $\mathbb{P}[Y_t = 2] = \mathbb{P}[X_v \leq t]$  a więc

$$\mathbb{P}[Y_t = 2] = \sum_{k=1}^t p(q\beta)^{k-1} = p \frac{1 - q^t \beta^t}{1 - q\beta}, \quad \mathbb{P}[Y_t = 1] = 1 - p \frac{1 - q^t \beta^t}{1 - q\beta}.$$

Stąd

$$\mathbb{E}[Y_t] = 1 + p \frac{1 - q^t \beta^t}{1 - q\beta}.$$

Przechodząc w granicę  $t \rightarrow \infty$  dostajemy

$$\mathbb{P}[W = 1] = 1 - \frac{p}{1 - q\beta}, \quad \mathbb{P}[W = 2] = \frac{p}{1 - q\beta}.$$

Co oczywiście daje nam natychmiast

$$\mathbb{E}[W] = 1 + \frac{p}{1 - q\beta}.$$

Jeśli chodzi o zmienną  $Z$  to propagacja zakończy się gdy wierzchołek  $v$  wyzdrowieje lub  $u$  zostanie poinformowany. Dzieje się to z prawdopodobieństwem  $1 - q\beta$  a więc

$$Z \sim \text{Geo}(1 - q\beta).$$

Mamy

$$\mathbb{E}[Z] = \frac{1}{1 - q\beta}.$$

## 6.2 Pewne wygaśnięcie

W modelu **SIR** wygaśnięcie infekcji jest zdarzeniem pewnym. Sformalizujmy ten fakt prostym twierdzeniem.

**Twierdzenie 4.** *Niech  $G = (V, E)$  będzie grafem spójnym,  $s \in V$  źródłem infekcji oraz  $v \in V \setminus \{s\}$ . Wtedy zachodzą następujące tożsamości:*

$$\mathbb{P}[X_v = \infty] > 0, \quad \mathbb{P}[Z < \infty] = 1.$$

Ponadto dla  $k \in \mathbb{N}$  zachodzi

$$\mathbb{P}[W = k] = \lim_{t \rightarrow \infty} \mathbb{P}[Y_t = k], \quad \mathbb{E}[W] = \lim_{t \rightarrow \infty} \mathbb{E}[Y_t].$$

*Dowód.* Ustalmy  $v \in V \setminus \{s\}$ . Jeśli po pierwszej rundzie źródło wyzdrowieje i nie przekaże dalej infekcji to  $X_v = \infty$ . Zatem  $\mathbb{P}[X_v = \infty] \geq \alpha q > 0$ . Jeśli rozważamy propagacje jako łańcuch Markowa to prawdopodobieństwo przejścia do stanu absorbującego, czyli  $\mathcal{I}_t = \emptyset$ , w czasie  $t$  wynosi

$$\alpha^{|\mathcal{I}_t|} \prod_{u \in \mathcal{I}_t} q^{|\mathbb{N}(u) \cap \mathcal{S}_t|}.$$

Istotne jest to, że jest to dodatnia liczba. A więc z Faktu 9 mamy  $\mathbb{P}[Z < \infty] = 1$ . Zauważmy, że

$$\lim_{t \rightarrow \infty} \mathbf{1}_{v \in \mathcal{I}_t \cup \mathcal{R}_t} \xrightarrow{\text{a.s.}} \mathbf{1}_{X_v < \infty}.$$

Ponadto  $Y_t \leq Y_{t+1}$  oraz  $Y_t \leq |V|$ . A więc z Twierdzenia Lebesgue dostajemy wzory na rozkład i wartość oczekiwana zmiennej  $W$ .  $\square$

### 6.3 Grafy ścieżkowe

Przyjrzyjmy się teraz grafom  $P_n$ . Oczywiście za źródło propagacji wybieramy wierzchołek  $s = 1$ . Podobnie jak w modelu **SI** propagacja rozchodzi się między kolejnymi wierzchołkami niezależnie a więc

$$X_1 = 0, \quad X_v = X_{v-1} + U_v, \quad v \in \{2, 3, \dots, n\},$$

gdzie jednak  $U_2, U_3, \dots, U_n \sim \text{KGeo}(p, \alpha)$  oraz są niezależne. Stąd  $X_v = U_2 + U_3 + \dots + U_v$  a zatem  $X_v$  ma rozkład umierający ujemny dwumianowy:

$$X_v \sim \text{KNegBin}(v - 1, p, \alpha),$$

oraz

$$\mathbb{E}[X_v | X_v < \infty] = \frac{v - 1}{1 - q\beta}.$$

Jeśli chodzi o zmienną  $Y_t$  to zauważmy, że

$$Y_t = \max\{j \in \{1, 2, \dots, n\} : X_j \leq t\}.$$

A więc dla  $k > 0$  mamy

$$\mathbb{P}[Y_t \geq k + 1] = \mathbb{P}[X_k \leq t] = \sum_{j=k}^t \binom{j-1}{k-1} p^k (q\beta)^{j-k}.$$

Dalej mamy

$$\mathbb{E}[Y_t] = \sum_{k=0}^{n-1} \mathbb{P}[Y_t \geq k + 1] = 1 + \sum_{k=1}^{n-1} \sum_{j=k}^t \binom{j-1}{k-1} p^k (q\beta)^{j-k}.$$

Nie ma zbytniej nadziei na zwartą formę tej sumy. Przejdźmy teraz do zmiennej  $W$ .

$$\mathbb{P}[W \geq k + 1] = \lim_{t \rightarrow \infty} \sum_{j=k}^t \binom{j-1}{k-1} p^k (q\beta)^{j-k} = \left(\frac{p}{1 - q\beta}\right)^k.$$

Oznaczmy  $\tau = \frac{p}{1 - q\beta}$ . Wtedy

$$\mathbb{P}[W = k] = (1 - \tau)\tau^{k-1}, \quad k \in \{1, 2, \dots, n-1\}, \quad \mathbb{P}[W = n] = \tau^{n-1}.$$

Ponadto

$$\mathbb{E}[W] = \sum_{k=0}^{n-1} \mathbb{P}[W \geq k + 1] = \sum_{k=0}^{n-1} \tau^k = \frac{1 - \tau^n}{1 - \tau}.$$

Skoncentrujmy naszą uwagę na zmiennej  $Z$ . Niech  $T_j$  oznacza czas przejścia ze stanu w którym wierzchołek  $j$  został dopiero co zainfekowany do kolejnego stanu, gdzie  $j \in V$ . Formalnie  $T_j = \min\{t \in \mathbb{N} : i = X_j \wedge \neg(j \in \mathcal{I}_{t+i} \wedge j+1 \in \mathcal{S}_{t+i})\}$ . Wtedy  $T_j \sim \text{Geo}(1 - q\beta)$ , a ponadto  $T_1, T_2, \dots$  są niezależne. Niech  $G_j(z) = \mathbb{E}[z^{T_j}]$ . Mamy

$$G_j(z) = \sum_{k=1}^{\infty} (1 - q\beta)(q\beta)^{k-1} z^k = \frac{(1 - q\beta)z}{1 - q\beta z}.$$

Zauważmy, że  $Z = T_1 + T_2 + \dots + T_W$ . A zatem

$$\mathbb{E}[z^Z] = \sum_{k=1}^{n-1} \mathbb{P}[W = k] \cdot G_k(z).$$

Funkcja generująca prawdopodobieństwo determinuje rozkład  $Z$ . Natomiast wartość oczekiwana wynosi

$$\mathbb{E}[Z] = \mathbb{E}[W] \cdot \mathbb{E}[T_1] = \frac{1 - \tau^n}{1 - \tau} \cdot \frac{1}{1 - q\beta} = \frac{1 - \tau^n}{\frac{q\alpha}{1 - q\beta}} \cdot \frac{1}{1 - q\beta} = \frac{1 - \tau^n}{q\alpha}.$$

## 6.4 Grafy gwiazdne

Rozpatrzmy propagacje na rodzinie  $S_n$ , gdzie źródłem jest wierzchołek 0. Dla  $v \in \{1, 2, \dots, n\}$  mamy

$$X_v \sim \text{KGeo}(p, \alpha),$$

a co za tym idzie

$$\mathbb{E}[X_v | X_v < \infty] = \frac{1}{1 - q\beta}.$$

Zauważmy natomiast, że zmienne  $X_1, X_2, \dots, X_n$  są zależne, bo gdy centralny wierzchołek wyzdrowieje żaden z liści nie może już zostać zainfekowany.

Przyjrzyjmy się teraz zmiennej  $Y_t$ . Położymy  $C = \min\{\tau \in \mathbb{N} : 0 \in \mathcal{R}_\tau\}$ . Jeśli centrum zarazi się po  $j$  rundach to liście gwiazdy mogą być zarażane przez  $\min\{j, t\}$  rund.  $Y_t$  zachowuje się tak samo jak w modelu **SI** dla grafów gwiazdnych. Przypomnijmy, że rozkład ten wynosi  $1 + B_{j,t}$  gdzie  $B_{j,t} \sim \text{Bin}(n, 1 - q^{\min\{j,t\}})$ . A więc

$$\mathbb{P}[Y_t = k + 1 | C = j] = \binom{n}{k} (1 - q^{\min\{j,t\}}) (q^{\min\{j,t\}})^{n-k}.$$

Ponadto  $C \sim \text{Geo}(\alpha)$  a więc  $\mathbb{P}[C = j] = \alpha\beta^{j-1}$ . Z prawdopodobieństwa całkowitego dostajemy

$$\begin{aligned}\mathbb{P}[Y_t = k+1] &= \sum_{j=1}^{\infty} \mathbb{P}[Y_t = k+1 | C = j] \cdot \mathbb{P}[C = j] = \\ &= \sum_{j=1}^{\infty} \binom{n}{k} (1 - q^{\min\{j,t\}})^k (q^{\min\{j,t\}})^{n-k} \alpha \beta^{j-1} = \\ &= \frac{\alpha}{\beta} \binom{n}{k} \sum_{j=1}^{\infty} (1 - q^{\min\{j,t\}})^k (q^{\min\{j,t\}})^{n-k} \beta^j.\end{aligned}$$

Aby obliczyć wartość oczekiwana skorzystamy z Lematu 3. Mamy  $\mathbb{P}[X_0 \leq t] = 1$  a dla  $v \neq 0$  mamy  $\mathbb{P}[X_v \leq t] = p \frac{1-(q\beta)^t}{1-q\beta}$ . A więc

$$\mathbb{E}[Y_t] = \sum_{v \in V} \mathbb{P}[X_v \leq t] = 1 + np \frac{1 - (q\beta)^t}{1 - q\beta}.$$

Licząc granice tych wyrażeń dla  $t \rightarrow \infty$  dostajemy

$$\mathbb{P}[W = k+1] = \frac{\alpha}{\beta} \binom{n}{k} \sum_{j=1}^{\infty} (1 - q^j)^k (q^j)^{n-k} \beta^j.$$

oraz

$$\mathbb{E}[W] = 1 + \frac{np}{1 - q\beta}.$$

Propagacja się zakończy gdy centrum wyzdrowieje lub wszystkie liście zostaną zarażone. Stąd mamy

$$\begin{aligned}\mathbb{P}[Z > t] &= \mathbb{P}[C > t] \cdot \mathbb{P}[Y_t < n+1 | C > t] = \\ &= \beta^t (1 - \mathbb{P}[Y_t = n+1 | C > t]) = \beta^t (1 - (1 - q^t)^n).\end{aligned}$$

Bezpośrednie liczenie wartości oczekiwanej zmiennej  $Z$  nie da nam zwiększej postaci. Postarajmy się ją za to oszacować korzystając z Nierówności 1. Oznaczmy  $a = -\log(q)$ ,  $b = -\log(\beta)$  oraz  $g(x) = e^{-bx}(1 - (1 - e^{-ax})^n)$ . Wtedy

$$\mathbb{E}[Z] \approx \int_0^\infty g(x) dx.$$

Dalej

$$\int_0^\infty e^{-bx}(1 - (1 - e^{-ax})^n) dx = \int_0^\infty e^{-bx} dx - \int_0^\infty e^{-bx}(1 - e^{-ax})^n dx.$$

Wartość pierwszej całki wynosi  $\frac{1}{b}$ . Aby policzyć drugą podstawmy  $u = 1 - e^{-ax}$ . Oczywiście  $u(0) = 0$ ,  $u(\infty) = 1$ . Wtedy też  $x = -\frac{1}{a} \log(1 - u)$  oraz  $dx = \frac{1}{a}e^{ax}du$ . Mamy zatem  $e^{-bx} = (1 - u)^{\frac{b}{a}}$ ,  $e^{ax} = \frac{1}{1-u}$ . A więc całka wynosi

$$\begin{aligned} \frac{1}{a} \int_0^1 u^n (1 - u)^{\frac{b}{a}-1} du &= \frac{1}{a} B\left(n+1, \frac{b}{a}\right) = \frac{1}{a} \frac{\Gamma(n+1)\Gamma(\frac{b}{a})}{\Gamma(n+1+\frac{b}{a})} \\ &= \frac{n!}{a} \cdot \frac{\Gamma(\frac{b}{a})}{(n+\frac{b}{a})(n-1+\frac{b}{a}) \dots (\frac{b}{a})\Gamma(\frac{b}{a})} = \frac{n!}{b} \cdot \frac{1}{(\frac{b}{a}+1)_n}. \end{aligned}$$

gdzie skorzystaliśmy z Faktu 8. Ostatecznie otrzymujemy

$$\mathbb{E}[Z] \approx \frac{1}{b} \cdot \left(1 - \frac{n!}{(\frac{b}{a}+1)_n}\right) = \frac{1}{\log(\beta)} \cdot \frac{n!}{\left(\frac{\log(\beta)}{\log(q)} + 1\right)_n} - \frac{1}{\log(\beta)}.$$

# Rozdział 7

## Analiza modelu SIS

### 7.1 Pewne wygaśnięcie

Analizę modelu **SIS** zacznijmy od formalnego udowodnienia pewności wygasnięcia infekcji.

**Twierdzenie 5.** *Niech  $G = (V, E)$  będzie grafem spójnym,  $s \in V$  źródłem infekcji oraz  $v \in V \setminus \{s\}$ . Wtedy zachodzą następujące tożsamości:*

$$\mathbb{P}[X_v = \infty] > 0, \quad \lim_{t \rightarrow \infty} \mathbb{E}[Y_t] = 0, \quad \mathbb{P}[Z < \infty] = 1.$$

*Dowód.* Dowód  $\mathbb{P}[X_v = \infty] > 0$  oraz  $\mathbb{P}[Z < \infty] = 1$  wynika z Twierdzenia 4. Skoro infekcja prawie na pewno wymrze to  $\mathbf{1}_{Y_t} \xrightarrow{\text{a.s.}} 0$  a więc i  $Y_t \xrightarrow{\text{a.s.}} 0$ . Zauważmy, że  $0 \leq Y_t \leq |V|$  dla każdego  $t$ . Zatem z Twierdzenia Lebesgue'a o zbieżności ograniczonej otrzymujemy

$$\lim_{t \rightarrow \infty} \mathbb{E}[Y_t] = \mathbb{E}\left[\lim_{t \rightarrow \infty} Y_t\right] = \mathbb{E}[0] = 0.$$

□

### 7.2 Dwa wierzchołki, jedna krawędź

Standardowo na pierwszy przykład grafu rozważamy graf o wierzchołkach  $V = \{u, v\}$  oraz krawędziach  $E = \{\{u, v\}\}$  oraz wierzchołku początkowym  $u$ . Istnieją cztery możliwe stany systemu:  $\mathcal{I}_t = \emptyset$ ,  $\mathcal{I}_t = \{u\}$ ,  $\mathcal{I}_t = \{v\}$ ,  $\mathcal{I}_t = \{u, v\}$ . Stan, w którym żaden wierzchołek nie jest zainfekowany, jest stanem absorbującym. Ponadto, z każdego pozostałego stanu możemy przejść do dowolnego innego. Rozkład  $X_v$  będzie tutaj taki sam jak dla modelu **SIR** a więc

$$X_v \sim \text{KGeo}(p, \alpha),$$

oraz

$$\mathbb{E}[X_v | X_v < \infty] = \frac{1}{1 - q\beta}.$$

Następnie spróbujmy wyznaczyć rozkład  $Y_t$ . Zauważmy, że  $Y_t \in \{0, 1, 2\}$ . Prawdopodobieństwa przejść są następujące:

$$\begin{aligned}\mathbb{P}[Y_{t+1} = 0 | Y_t = 0] &= 1, \\ \mathbb{P}[Y_{t+1} = 0 | Y_t = 1] &= q\alpha, \\ \mathbb{P}[Y_{t+1} = 0 | Y_t = 2] &= \alpha^2, \\ \mathbb{P}[Y_{t+1} = 1 | Y_t = 0] &= 0, \\ \mathbb{P}[Y_{t+1} = 1 | Y_t = 1] &= p\alpha + q\beta, \\ \mathbb{P}[Y_{t+1} = 1 | Y_t = 2] &= 2\alpha\beta, \\ \mathbb{P}[Y_{t+1} = 2 | Y_t = 0] &= 0, \\ \mathbb{P}[Y_{t+1} = 2 | Y_t = 1] &= p\beta, \\ \mathbb{P}[Y_{t+1} = 2 | Y_t = 2] &= \beta^2.\end{aligned}$$

Z wzoru na prawdopodobieństwo całkowite dla  $k \in \{0, 1, 2\}$  mamy:

$$\mathbb{P}[Y_{t+1} = k] = \sum_{j=0}^2 \mathbb{P}[Y_{t+1} = k | Y_t = j] \cdot \mathbb{P}[Y_t = j].$$

Oznaczmy  $a_t = \mathbb{P}[Y_t = 0]$ ,  $b_t = \mathbb{P}[Y_t = 1]$ ,  $c_t = \mathbb{P}[Y_t = 2]$ . Oczywiście  $a_0 = 0$ ,  $b_0 = 1$ ,  $c_0 = 0$ . Stąd otrzymujemy układ równań rekurencyjnych:

$$\begin{cases} a_{t+1} = a_t + q\alpha \cdot b_t + \alpha^2 \cdot c_t \\ b_{t+1} = (p\alpha + q\beta) \cdot b_t + 2\alpha\beta \cdot c_t \\ c_{t+1} = p\beta \cdot b_t + \beta^2 \cdot c_t. \end{cases}$$

Położymy

$$\mathbf{P} = \begin{bmatrix} 1 & q\alpha & \alpha^2 \\ 0 & p\alpha + q\beta & 2\alpha\beta \\ 0 & p\beta & \beta^2 \end{bmatrix}, \quad \mathbf{y}_t = \begin{bmatrix} a_t \\ b_t \\ c_t \end{bmatrix}.$$

Wtedy  $\mathbf{y}_{t+1} = \mathbf{P}\mathbf{y}_t$ , a więc  $\mathbf{y}_t = \mathbf{P}^t\mathbf{y}_0$ . Jeśli chodzi o wartość oczekiwana, to mamy  $\mathbb{E}[Y_t] = 0 \cdot a_t + 1 \cdot b_t + 2 \cdot c_t = b_t + 2c_t$ . Z Twierdzenia 5 mamy

$$\lim_{t \rightarrow \infty} \mathbb{E}[Y_t] = 0.$$

Przyjrzyjmy się teraz zmiennej  $Z$ . Mamy  $\mathbb{P}[Z > t] = \mathbb{P}[Y_t \neq 0] = b_t + c_t$ . Oznaczmy

$$\mathbf{Q} = \begin{bmatrix} p\alpha + q\beta & 2\alpha\beta \\ p\beta & \beta^2 \end{bmatrix}, \quad \mathbf{z}_t = \begin{bmatrix} b_t \\ c_t \end{bmatrix}, \quad \mathbf{1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Wtedy  $\mathbb{P}[Z > t] = \mathbf{1}^\top \mathbf{z}_t = \mathbf{1}^\top \mathbf{Q}^t \mathbf{z}_0$ . Mamy więc

$$\mathbb{E}[Z] = \sum_{t=0}^{\infty} \mathbb{P}[Z > t] = \sum_{t=0}^{\infty} \mathbf{1}^\top \mathbf{Q}^t \mathbf{z}_0 = \mathbf{1}^\top (\mathbf{I} - \mathbf{Q})^{-1} \mathbf{z}_0.$$

Dalej:

$$(\mathbf{I} - \mathbf{Q})^{-1} = \begin{bmatrix} 1 - p\alpha - q\beta & -2\alpha\beta \\ -p\beta & 1 - \beta^2 \end{bmatrix}^{-1} = \frac{1}{\Delta} \begin{bmatrix} 1 - \beta^2 & 2\alpha\beta \\ p\beta & 1 - p\alpha - q\beta \end{bmatrix},$$

oraz

$$\mathbf{1}^\top (\mathbf{I} - \mathbf{Q})^{-1} \mathbf{z}_0 = \mathbf{1}^\top \frac{1}{\Delta} \begin{bmatrix} 1 - \beta^2 \\ p\beta \end{bmatrix} = \frac{1 - \beta^2 + p\beta}{\Delta},$$

gdzie  $\Delta = \det(\mathbf{I} - \mathbf{Q})$ . Obliczmy teraz ten wyznacznik:

$$\begin{aligned} \Delta &= (1 - p\alpha - q\beta)(1 - \beta^2) - (-2\alpha\beta)(-p\beta) \\ &= 1 - p\alpha - q\beta - \beta^2 + p\alpha\beta^2 + q\beta^3 - 2p\alpha\beta \\ &= (1 - \beta^2)(1 - q\beta) - p\alpha(1 + \beta^2) \\ &= (1 - (1 - \alpha)^2)(1 - (1 - p)(1 - \alpha)) - p\alpha(1 + (1 - \alpha)^2) \\ &= \alpha(2 - \alpha)(p + \alpha - p\alpha) - p\alpha(2 - 2\alpha + \alpha^2) \\ &= \alpha(2p + 2\alpha - 2p\alpha - \alpha p - \alpha^2 + p\alpha^2 - 2p + 2\alpha p - p\alpha^2) \\ &= \alpha^2(2 - p - \alpha) = \alpha^2(q + \beta). \end{aligned}$$

Ostatecznie otrzymujemy:

$$\mathbb{E}[Z] = \frac{1 - \beta^2 + p\beta}{\alpha^2(q + \beta)}.$$

### 7.3 Grafy pełne

Po analizie dla  $K_2$  postarajmy się ją uogólnić dla  $K_n$ . Ze względu na symetrię dowolny stan z taką samą liczbą zainfekowanych wierzchołków jest izomorficzny. Z Twierdzenia 5 wiemy, że  $\mathbb{E}[Y_t] \rightarrow 0$ . Ustalmy  $t \in \mathbb{N}$ . Niech  $A_t$  będzie zmienną losową oznaczającą liczbę wierzchołków, które przetrwają rundę  $t$  oraz  $B_t$  liczbę nowo zainfekowanych wierzchołków w tej rundzie. Formalnie

$$A_t = |\{v \in V : v \in \mathcal{S}_t \cap \mathcal{S}_{t+1}\}|, \quad B_t = |\{v \in V : v \in \mathcal{S}_t \cap \mathcal{I}_{t+1}\}|.$$

Oznaczmy  $i = Y_t$  oraz  $j = Y_{t+1}$ . Każdy zarażony wierzchołek może przetrwać niezależnie od siebie w formacie próby Bernoulliego. A więc  $A_t \sim \text{Bin}(i, \beta)$ . Dalej każdy podatny wierzchołek może zostać zarażony niezależnie przez  $n -$

$i$  sąsiadów. Szansa, że któremuś z nich się uda wynosi  $1 - q^i$ . Stąd  $B_t \sim \text{Bin}(n - i, 1 - q^i)$ . Ponadto  $Y_{t+1} = A_t + B_t$ . Możemy policzyć warunkową liczbę oczekiwanych zarażeń w  $t + 1$  kroku:

$$\mathbb{E}[Y_{t+1}|Y_t = i] = \mathbb{E}[A_t + B_t] = i\beta + (n - i)(1 - q^i).$$

Oznaczmy  $P_{i \rightarrow j} = \mathbb{P}[Y_{t+1} = j|Y_t = i]$  dla  $j \in \{0, 1, \dots, n\}$ . Wtedy korzystając z Faktu 7 mamy

$$P_{i \rightarrow j} = \sum_{\ell=\max\{0, j-(n-i)\}}^{\min\{i, j\}} \binom{i}{\ell} \binom{n-i}{j-\ell} \beta^\ell \alpha^{i-\ell} (1-q^i)^{j-\ell} (q^i)^{n-i-(j-\ell)}.$$

Jest to uogólnienie prawdopodobieństw przejść, które wyznaczyliśmy wcześniej dla  $n = 2$ . Ponownie z prawdopodobieństwa całkowitego dla  $j \in \{0, 1, \dots, n\}$  mamy

$$\mathbb{P}[Y_{t+1} = j] = \sum_{i=0}^n \mathbb{P}[Y_{t+1} = j|Y_t = i] \cdot \mathbb{P}[Y_t = i].$$

Zdefiniujmy wektor

$$\mathbf{y}_t = \begin{bmatrix} \mathbb{P}[Y_t = 0] \\ \mathbb{P}[Y_t = 1] \\ \vdots \\ \mathbb{P}[Y_t = n] \end{bmatrix}$$

Wtedy  $\mathbf{y}_t^{(k)} = \mathbb{P}[Y_t = k]$  dla  $k \in \{0, 1, \dots, n\}$ . W uproszczonej notacji możemy zapisać

$$\mathbf{y}_{t+1}^{(j)} = \sum_{i=0}^n P_{i \rightarrow j} \cdot \mathbf{y}_t^{(i)}.$$

Macierz przejść dla naszego łańcucha Markova dana jest

$$\mathbf{P} = [P_{i \rightarrow j}]_{0 \leq i, j \leq n}.$$

Mamy  $\mathbf{y}_{t+1} = \mathbf{P}\mathbf{y}_t$  jak i  $\mathbf{y}_t = \mathbf{P}^t\mathbf{y}_0$ . Zbudowanie tej macierzy jest możliwe w czasie  $\mathcal{O}(n^3)$ .

Jeśli chodzi o zmienną  $Z$  to mamy  $\mathbb{P}[Z > t] = \mathbb{P}[Y_t \neq 0]$ . Kładziemy

$$\mathbf{Q} = [P_{i \rightarrow j}]_{1 \leq i, j \leq n}, \quad \mathbf{z}_t = [\mathbb{P}[Y_t] = k]_{1 \leq k \leq n}, \quad \mathbf{1} = [1]_{1 \leq k \leq n}.$$

Licząc wartość oczekiwana dostajemy

$$\mathbb{E}[Z] = \sum_{t=0}^{\infty} \mathbb{P}[Z > t] = \sum_{t=0}^{\infty} \mathbf{1}^\top \mathbf{z}_t = \sum_{t=0}^{\infty} \mathbf{1}^\top \mathbf{Q}^t \mathbf{z}_0 = \mathbf{1}^\top (\mathbf{I} - \mathbf{Q})^{-1} \mathbf{z}_0.$$