Probabilistyczne modele propagacji w grafach.

Bartosz Łabuz

17 listopada 2025

Spis treści

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 społecznościowych, 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 propagację.

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 $f : \mathbb{R} \to \mathbb{R}$ jest funkcją to przez $f(\pm\infty)$ oznaczamy $\lim_{x\to\pm\infty} f(x)$. Symbol Pochhammer'a oznaczamy $(x)_n = x(x + 1) \dots (x + n - 1)$.

Niech $G = (V, E)$ będzie grafem prostym nieskierowanym. Stopień wierzchołka $v \in V$ oznaczamy $\deg(v)$. Zbiór sąsiadów $v \in V$ oznaczamy $\mathrm{N}(v)$. Odległość między u i v oznaczamy $\mathrm{d}(u, v)$ dla $u, v \in V$. Ekscentryczność $v \in V$ oznaczamy $\epsilon(v) = \max_{u\in V} \mathrm{d}(u, v)$. Przez $\delta(G)$ i $\Delta(G)$ oznaczamy odpowiednio minimalny i maksymalny stopień wierzchołka w grafie G.

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

2.2 Rodziny grafów

Graf ścieżkowy P_n

Dla $n \in \mathbb{N}_+$ 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

Dla $n \in \mathbb{N}_+$ graf gwiazda ma zbiór wierzchołków $V = \{0, 1, \ldots, n\}$ oraz zbiór krawędzi $E = \{\{0, i\} : i \in \{1, 2, \ldots, n\}\}$. Oznaczamy go przez S_n.

Graf cykliczny C_n

Dla $n \in \mathbb{N}_+$ graf cykliczny ma zbiór wierzchołków $V = \{1, 2, \ldots, n\}$ oraz zbiór krawędzi $E = \{\{i, i+1\} : i \in \{1, 2, \ldots, n-1\}\} \cup \{\{n, 1\}\}$. Oznaczamy go przez C_n.

Graf pełny K_n

Dla $n \in \mathbb{N}_+$ graf pełny ma zbiór wierzchołków $V = \{1, 2, \ldots, n\}$ oraz zbiór krawędzi $E = \{\{i, j\} : i, j \in \{1, 2, \ldots, n\} \wedge i \neq j\}$. Oznaczamy go przez K_n.

2.3 Rozkłady prawdopodobieństwa

Rozkład Bernoulliego $\mathrm{Ber}(p)$

Próba Bernoulliego to doświadczenie losowe, którego wynik może być jednym z dwóch:

- sukces z prawdopodobieństwem $p \in (0; 1)$,
- porażka z prawdopodobieństwem $1 - p$.

Zmienna losowa X przyjmująca wartość 1 w przypadku sukcesu i 0 w przypadku porażki ma rozkład Bernoulliego. Oznaczamy $X \sim \mathrm{Ber}(p)$.

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

Rozkład dwumianowy opisuje liczbę sukcesów w n próbach Bernoulliego. Niech X będzie zmienną losową przyjmującą wartości w $\{0, 1, \ldots, n\}$, a każda próba ma prawdopodobieństwo sukcesu $p \in (0; 1)$. Wtedy:

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

Wartość oczekiwana i wariancja mają postać:

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

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

Rozkład geometryczny $\mathrm{Geo}(p)$

Rozkład geometryczny opisuje liczbę prób Bernoulliego potrzebnych do uzyskania pierwszego sukcesu. Niech X będzie zmienną losową przyjmującą wartości w $\mathbb{N}_+$, a 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 mają postać:
$$\mathbb{E}[X] = \frac{1}{p}, \quad \mathrm{Var}[X] = \frac{1-p}{p^2}.$$
Oznaczamy $X \sim \mathrm{Geo}(p)$.

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

Rozkład ujemny dwumianowy opisuje liczbę prób Bernoulliego potrzebnych do uzyskania m sukcesów. Niech X oznacza liczbę prób, przy czym każda próba ma prawdopodobieństwo sukcesu $p \in (0;1)$, a liczba sukcesów $m \in \mathbb{N}_+$ jest ustalona. Wtedy:
$$\mathbb{P}[X = k] = \binom{k-1}{m-1} p^m (1-p)^{k-m}, \quad k \geq m.$$
Wartość oczekiwana i wariancja mają postać:
$$\mathbb{E}[X] = \frac{m}{p}, \quad \mathrm{Var}[X] = \frac{m(1-p)}{p^2}.$$
Oznaczamy $X \sim \mathrm{NegBin}(m,p)$.

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

Zdefiniujmy funkcje
$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}, \quad \mathbf{\Phi}(t) = \int_{-\infty}^{t} \varphi(x)\, \mathrm{d}x.$$
Niech $\mu \in \mathbb{R}$ oraz $\sigma > 0$. Zmienna losowa X ma rozkład normalny, jeśli jej funkcja gęstości wyraża się wzorem:
$$f_X(t) = \frac{1}{\sigma} \cdot \varphi\Big(\frac{t-\mu}{\sigma}\Big), \quad t \in \mathbb{R}.$$

Dystrybuanta jest równa:

$$\mathbb{P}[X \le t] = \mathbf{\Phi}\Big(\frac{t-\mu}{\sigma}\Big), \quad t \in \mathbb{R}.$$

Wartość oczekiwana i wariancja:

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

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

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

Rozkład umierający geometryczny KGeo(p, α)

Rozkład umierający geometryczny jest wariantem rozkładu geometrycznego, w którym proces może zostać przerwany z prawdopodobieństwem α w każdym kroku. Niech X będzie zmienną losową przyjmującą wartości w $\mathbb{N}_+ \cup \{\infty\}$. Wtedy:

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

oraz

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

Dystrybuanta jest równa:

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

Wartość oczekiwana wynosi nieskończoność. Natomiast

$$\mathbb{E}[X | X < \infty] = \frac{1}{1-(1-p)(1-\alpha)}.$$

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

2.4 Fakty, sumy i nierówności

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

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

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

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

Fakt 4. *Niech* $X_1, X_2, \ldots, X_m \sim \mathrm{Geo}(p)$ *będą IID oraz* $Y = X_1 + \cdots + X_m$. *Wtedy*

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

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* $x, y > 0$. *Wtedy*

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

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

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

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*

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

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

$$\int_a^b f(x)\,\mathrm{d}x \le \sum_{k=a}^{b} f(k) \le f(a) + \int_a^b f(x)\,\mathrm{d}x.$$

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

$$H_n \le 1 + \log(n)$$

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

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

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

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

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

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

$$\mathbb{P}[X \ge t] \le \frac{\mathbb{E}[X]}{t}.$$

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 $g : \mathbb{R}^n \to \mathbb{R}$ będzie funkcją wypukłą zaś $X_1, X_2, \ldots, X_n : \Omega \to \mathbb{N}$ będą zmiennymi losowymi (niekoniecznie niezależnymi). Wtedy*

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

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

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

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

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 $\mathbf{X} : \mathbb{N} \times V \to \mathcal{Q}$, 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:

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

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 topologi.

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 $\mathbb{E}[Y_t] \to |V|$ wraz z $t \to \infty$. 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 sytuacją nigdy by nie nastąpiła to mielibyśmy $Z = \infty$. Jednakże dla propagacji **SI** zachodzi zainfekowanie całego grafu jest zdarzeniem pewnym, $\mathbb{P}[Z < \infty] = 1$. Alternatywnie możemy zapisać $Z = \max_{v \in V} X_v$. Wyznacznie rozkładu Z oraz wartości oczekiwanej dla konkretnych rodzin grafów będzie głównym celem w tym modelu. Dodatkowo interesować nas będzie asymptotyczne oszacowanie $\mathbb{E}[Z]$ względem liczby wierzchołków.

3.2 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:

$$\mathbb{P}[\mathbf{X}_{t+1}(u) = S \mid \mathbf{X}_t(u) = S] = \prod_{v \in \mathrm{N}(u) \cap \mathcal{I}_t} q,$$

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

$$\mathbb{P}[\mathbf{X}_{t+1}(u) = S \mid \mathbf{X}_t(u) = I] = \alpha,$$

$$\mathbb{P}[\mathbf{X}_{t+1}(u) = I \mid \mathbf{X}_t(u) = I] = \beta.$$

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:

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

Wygaśnięcie infekcji jest zachodzi z prawdopodobieństwem 1 (patrz Twierdzenie 4). W kontraście dla modelu **SI** mamy $\mathbb{P}[X_v = \infty] > 0$ a więc $\mathbb{E}[X_v] = \infty$. Przekierujemy naszą uwagę zatem na $\mathbb{E}[X_v | X_v < \infty]$. Dalej wykażemy, że $\mathbb{E}[Y_t]$ dążą do 0, a więc także nie będą nas interesować. Skupimy się tylko na rozkładzie Y_t. Dla zmiennej Z przyjrzyjmy się zarówno rozkładowi jak i wartości oczekiwanej.

3.3 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 trwały — wierzchołek, który wyzdrowiał, nie może już ani się zarazić, ani nikogo zakazić. Zarażony wierzchołek może przejść z I do stanu R z prawdopodobieństwem $\alpha \in (0;1)$. 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:

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

Tak jak poprzednio rozważmy zmienne

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

W modelu **SIR** również zachodzi $\mathbb{P}[X_v = \infty] > 0$. A więc poza rozkładem X_v możemy wyznaczyć $\mathbb{E}[X_v | X_v < \infty]$. Istnieje 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 $\mathrm{Geo}(\alpha)$, a więc zmienna ta była by po prostu sumą rozkładu geometrycznego z X_v. Z tego powodu tego nie będziemy jej rozważać. Zamiast rozważać liczbe tylko zainfekowanych lub tylko wyzdrowiałych wierzchołków będziemy rozważać liczbe nie podatnych wierzchołków po t krokach. Kładziemy więc

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

Podobnie jak w **SIS** zdarzenie $\mathbb{P}[\mathcal{I}_t = \varnothing] \to 1$ wraz z $t \to \infty$ a więc możemy zdefiniować zmienną Z oznaczającą czas wygaśnięcia infekcji. Dla uproszczenia będziemy rozważać moment w którym żaden nowy wierzchołek nie może byc zarażony. A więc

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

Dodatkowo kładziemy zmienną W będącą liczba finalnie wyzdrowiałych wierzchołków.

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

Rozkłady Z, W jak i ich wartości oczekiwane będą głownym zainteresowaniem w tym modelu.

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, $X_v \sim \mathrm{Geo}(p)$.

Rozważmy teraz rozkład Y_t. Mamy $\mathbb{P}[Y_t = 1] = q^t$, ponieważ próba zarażenia musiałaby nie udać 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 $Z \sim \mathrm{Geo}(p)$ oraz $\mathbb{E}[Z] = \frac{1}{p}$.

4.2 Trójkąt

Przyjrzyjmy 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. Możemy więc zapisać $X_u = \min\{A, B\}$, gdzie $A \sim \mathrm{Geo}(p)$ oraz $B \sim \mathrm{NegBin}(2, p)$. Wiemy, że $\mathbb{P}[A \leq t] = 1 - q^t$. Z kolei

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

gdzie skorzystaliśmy z Sumy 4. Dalej, z Faktu 2 mamy

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

Jeśli chodzi o liczbę zainfekowanych po t krokach, to skoro mamy trzy wierzchołki, mamy też 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 $1 \le k \le t$ 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. Podsumowując:

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

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

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

Widzimy, że $\mathbb{E}[Y_t] \to 3$ przy $t \to \infty$, co jest zgodne z intuicją.

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 $1 \le k \le t-1$) 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 $1-q^2$ szans. Łącznie dostajemy:

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

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

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

(patrz Suma 5 oraz Suma 6). 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:*

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

Dowód. Weźmy $v \in V \setminus \{s\}$. Wybierzemy ustaloną ścieżkę od s do v, $s = u_0u_1u_2 \dots u_\ell = v$ gdzie ℓ to długość tej ścieżki ($\ell \geq 1$). Aby v został zainfekowane, wystarczy, że po kolei na tej ścieżce zajdą sukcesy transmisji. Rozważmy następujące zdarzenia zainfekowania tych wierzchołków w bloku czasu indeksowanym przez $k \in \mathbb{N}$:

$$A_k = \{\forall j \in \{1, \dots, \ell\} \quad X_{u_j} = k\ell + j\}.$$

Próby transmisji w różnych rundach są niezależne a więc $\mathbb{P}[A_k] = p^\ell > 0$. Jeśli chociaż jedno ze zdarzeń A_k zajdzie to wtedy $v \in \mathcal{I}_t$ dla pewnego t. Mamy zatem

$$\begin{aligned}\mathbb{P}[X_v = \infty] &\leq \mathbb{P}\Big[\bigcap_{k=0}^{\infty} A_k^c\Big] = \lim_{m\to\infty} \mathbb{P}\Big[\bigcup_{k=0}^{m} A_k^c\Big] \\ &= \lim_{m\to\infty} \prod_{k=0}^{m} 1 - p^\ell = \lim_{m\to\infty} (1 - p^\ell)^{m+1} = 0.\end{aligned}$$

Stąd oczywiscie $\mathbb{P}[X_v = \infty] = 0$. Dalej zauważmy, że $0 \leq Y_t \leq |V|$. Z Twierdzenia Lebesgue o zbieżności ograniczonej mamy

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

Wreszcie

$$\mathbb{P}[Z = \infty] = \mathbb{P}[\max_{v\in V} X_v = \infty] = 0.$$

Stąd $\mathbb{P}[Z < \infty] = 1$. □

4.4 Grafy ścieżkowe

Jako pierwszą rodzinę grafów rozważmy grafy ścieżkowe P_n. Załóżmy, że proces zaczyna się w wierzchołku $s = 1$. Zatem infekcja rozchodzi się po grafie

„od lewej do prawej". Dla tej rodziny grafów uda nam się wyznaczyć dokładny rozkład prawdopodobieństwa. Zauważmy, że czasy zarażenia kolejnych wierzchołków tworzą ciąg zmiennych losowych:

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

gdzie $U_2, U_3, \dots, U_n \sim \mathrm{Geo}(p)$ oraz $U_2, U_3, \dots, U_n$ są niezależne. Widzimy zatem, że $X_v = U_1 + U_2 + \cdots + U_{v-1}$, a więc z Faktu 4 X_v ma rozkład ujemny dwumianowy:

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

Ponadto mamy:

$$\mathbb{E}[X_v] = \frac{v-1}{p}, \quad \mathrm{Var}[X_v] = \frac{(v-1)(1-p)}{p^2}.$$

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:

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

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

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

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 $n \to \infty$. Wtedy $n > 1+j$ dla wszystkich $0 \leq j \leq t$, a więc:

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

gdzie sumy obliczamy korzystając z Sumy 1 oraz Sumy 2. Stąd:

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

Czas całkowitego zainfekowania grafu P_n to $Z = \max\{X_1, X_2, \ldots, X_n\} = X_n$. Zatem rozkład zmiennej Z jest już nam znany: $Z \sim \mathrm{NegBin}(n-1, p)$, 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 $n \in \{1, 2, \ldots, 1000\}$ tą samą liczbą symulacji oszacujmy $\mathbb{E}[Z]$.

Wyniki eksperymentu niemal idealnie pokrywają się z przewidywanymi kształtami, to jest $1 + tp$ dla $\mathbb{E}[Y_t]$ (zob. Wykres 1) oraz $\frac{n-1}{p}$ dla $\mathbb{E}[Z]$ (zob. Wykres 2).

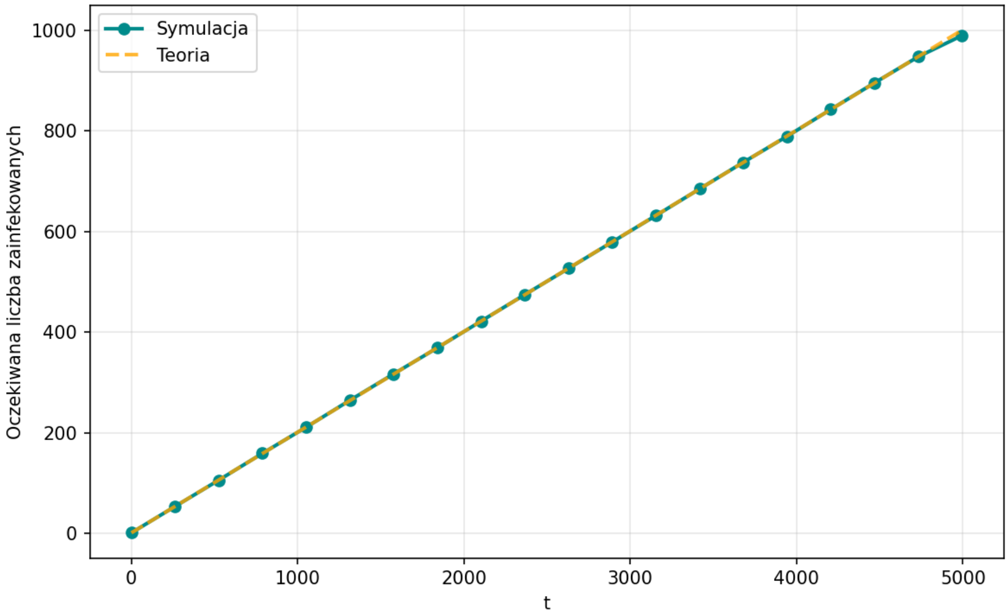
4.5 Grafy gwiezdne

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. Stąd mamy $X_v \sim \mathrm{Geo}(p)$ dla każdego $v \in \{1, 2, \ldots, n\}$, a zmienne $X_1, X_2, \ldots, X_n$ są od siebie niezależne. Otrzymujemy zatem:

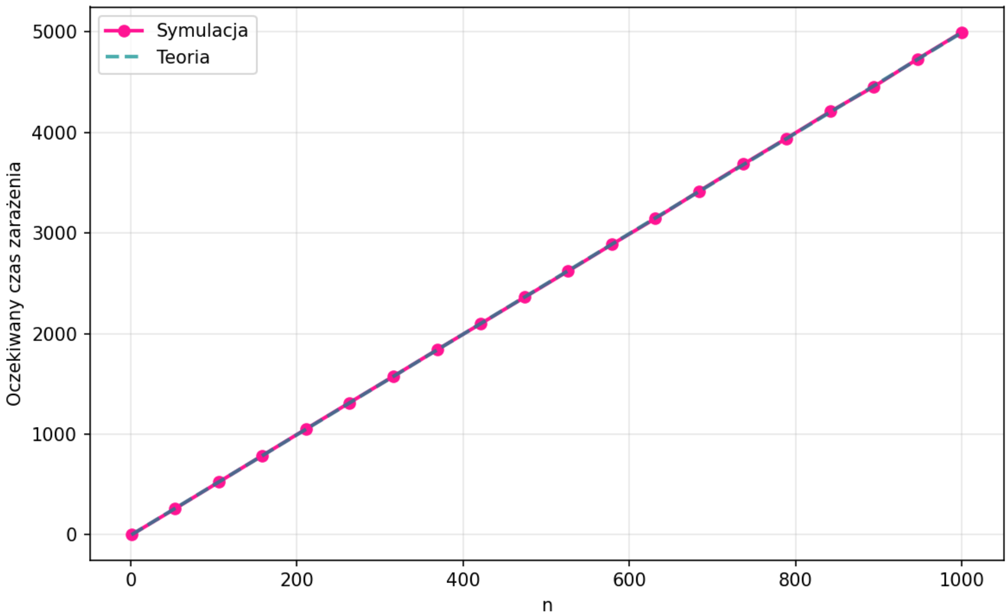
$$\mathbb{E}[X_v] = \frac{1}{p}, \quad \mathrm{Var}[X_v] = \frac{1-p}{p^2}.$$

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

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



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



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

W konsekwencji:

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

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

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

Policzmy wartość oczekiwaną 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:

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

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

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

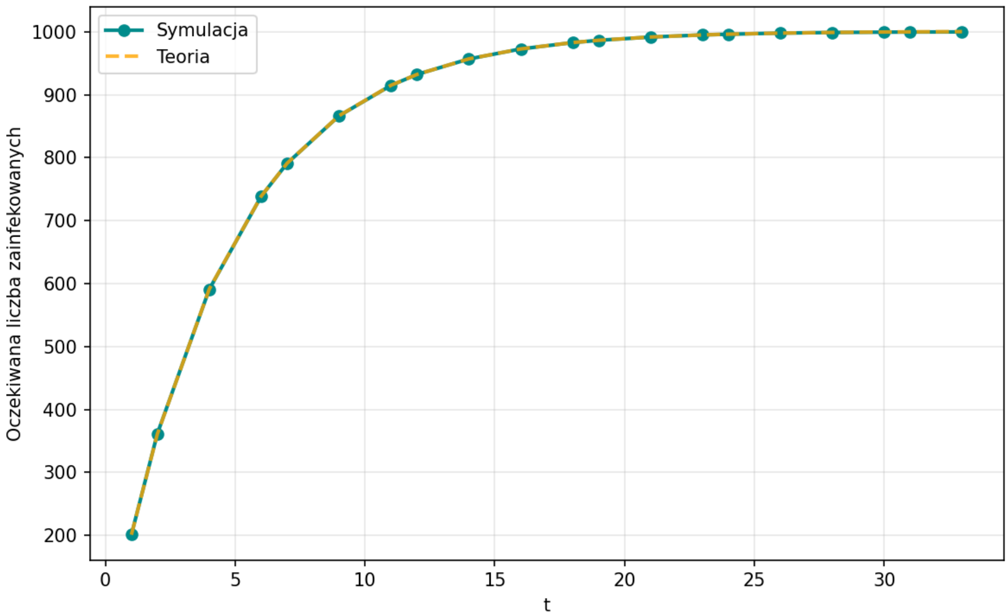
Zauważmy, że $-\log(q) = \log\left(\frac{1}{1-p}\right)$, zatem:

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

Stąd asymptotyczny czas pełnego zarażenia grafu S_n ma postać:

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

Przeprowadźmy teraz symulacje. Ustalmy $p = 0.2$ oraz $n = 1000$. Dla każdego $t \in \{1, 2, \ldots, \log(n)\}$ wykonajmy 2000 powtórzeń propagacji na S_n, a następnie dla $n \in \{1, 2, \ldots, 1000\}$ oszacujmy $\mathbb{E}[Z]$. Wyniki przedstawiono na Wykresie 3 oraz Wykresie 4.

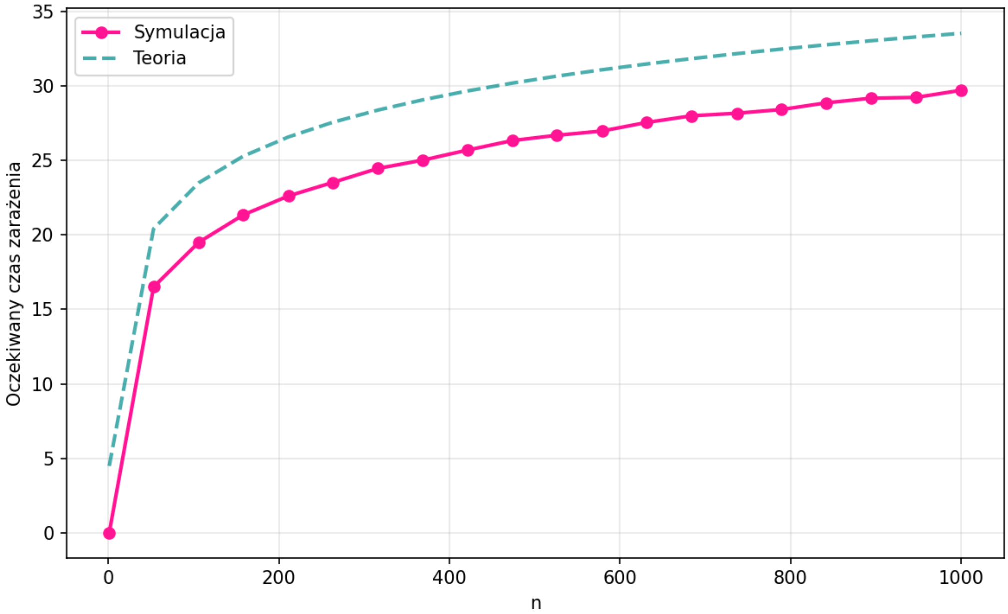


Rysunek 3: $\mathbb{E}[Y_t]$ dla $\mathrm{S_n}$ w funkcji t.

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.

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



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

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.

Twierdzenie 2. *Niech $G = (V, E)$ będzie grafem spójnym, a $G' = (V, E')$ jego spójnym podgrafem. Załóżmy, 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' oznaczymy 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 podgrafu 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, a $s \in V$ — ustalonym źródłem. Oznaczmy $\lambda = \log\left(\frac{1}{1-p}\right)$ 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 $A_j = \{v \in V : \mathrm{d}(s, v) = j\}$ oraz liczności $a_j = |A_j|$. Mamy oczywiście $a_0 = 1$, a więc $a_1 + \cdots + a_h = n - 1$. Zdefiniujmy dalej 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. $\square$

Przejdźmy do dowodu głównego twierdzenia. Mamy $T_h = \sum_{j=1}^{h} U_j$, zatem:

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

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

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 — szczególnie dla grafów rzadkich.

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 $v \in \{1, 2, \ldots, n-1\}$. 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 $j \in \{1, 2, \ldots, n-1\}$ połóżmy $N_j \sim \mathrm{NegBin}(j, p)$. 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

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

Nie ma co liczyć na wyznaczenie eleganckiej postaci na PMF czy CDF dla X_v. Postaramy się więc przybliżyć wartość oczekiwaną dla dużych n. Z centralnego twierdzenia granicznego możemy przybliżyć $N_a \approx A,\ N_b \approx B$ dla $A \sim \mathcal{N}(\mu_a, \sigma_a^2),\ B \sim \mathcal{N}(\mu_b, \sigma_b^2)$ gdzie

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

Zatem mamy

$$\mathbb{E}[X_v] \approx \mathbb{E}[\min\{A,B\}] = \mathbb{E}\Big[\frac{A+B-|A-B|}{2}\Big] = \frac{\mathbb{E}[A]+\mathbb{E}[B]-\mathbb{E}[|A-B|]}{2}.$$

Połóżmy $C = A - B$. 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*

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

Dowód.

$$\mathbb{E}[|X|] = \int_{-\infty}^{\infty} \frac{|x|}{\sigma}\varphi\Big(\frac{x-\mu}{\sigma}\Big)\,\mathrm{d}x = \int_0^{\infty} \frac{x}{\sigma}\varphi\Big(\frac{x-\mu}{\sigma}\Big)\,\mathrm{d}x - \int_{-\infty}^{0} \frac{x}{\sigma}\varphi\Big(\frac{x-\mu}{\sigma}\Big)\,\mathrm{d}x.$$

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

$$\begin{aligned}
&\int_{-c}^{\infty} (\mu + \sigma z)\varphi(z)\,\mathrm{d}z - \int_{-\infty}^{-c} (\mu + \sigma z)\varphi(z)\,\mathrm{d}z \\
&= \mu\int_{-c}^{\infty} \varphi(z)\,\mathrm{d}z + \sigma\int_{-c}^{\infty} z\varphi(z)\,\mathrm{d}z - \mu\int_{-\infty}^{-c} \varphi(z)\,\mathrm{d}z - \sigma\int_{-\infty}^{-c} z\varphi(z)\,\mathrm{d}z \\
&= \mu\Big(\int_{-c}^{\infty} \varphi(z)\,\mathrm{d}z - \int_{-\infty}^{-c} \varphi(z)\,\mathrm{d}z\Big) + \sigma\Big(\int_{-c}^{\infty} z\varphi(z)\,\mathrm{d}z - \int_{-\infty}^{-c} z\varphi(z)\,\mathrm{d}z\Big) \\
&= \mu\Big(\int_{-\infty}^{\infty} \varphi(z)\,\mathrm{d}z - 2\int_{-\infty}^{-c} \varphi(z)\,\mathrm{d}z\Big) + \sigma\Big(- \varphi(z)\big|_{-c}^{\infty} + \varphi(z)\big|_{-\infty}^{-c}\Big) \\
&= \mu \cdot \Big(1 - 2\mathbf{\Phi}(-c)\Big) + \sigma \cdot \Big(- \varphi(\infty) + \varphi(-c) + \varphi(-c) - \varphi(-\infty)\Big) \\
&= \mu \cdot \Big(2\mathbf{\Phi}(c) - 1\Big) + 2\sigma \cdot\ \varphi(c).
\end{aligned}$$

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

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

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

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 $\frac{\eta}{\xi} \to -\infty$ wraz z $n \to \infty$. Wtedy też $\varphi\left(\frac{\eta}{\xi}\right) \to 0$ oraz $\mathbf{\Phi}\left(\frac{\eta}{\xi}\right) \to 0$ a więc $\mathbb{E}[X_v] \to \frac{a}{p}$.

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 $\mathbf{\Phi}(0) = \frac{1}{2}$. Wstawiając otrzymamy $\mathbb{E}[X_v] \to \frac{n}{2p} - \frac{\sqrt{np}}{p\sqrt{2\pi}}$. 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 dobiega końca. Każda z tych dróg jak w przypadku grafu ścieżkowego ma rozkład dwumianowy. Możemy zapisać zatem

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

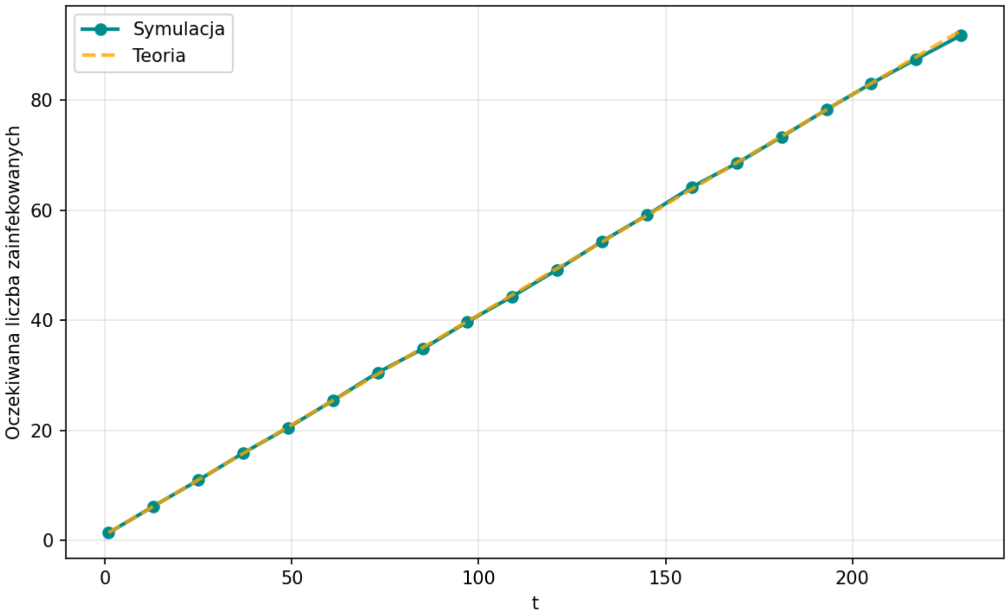
Z Faktu 3 mamy $L_t + R_t \sim \mathrm{Bin}(2t, p)$. Widzimy zatem, że rozkład Y_t dla grafu C_n pokrywa się ze zmienną Y_{2t} dla grafów typu P_n. Z wcześniejszego wyniku dla grafów ścieżek dostajemy

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

Teraz możemy wyznaczyć $\mathbb{E}[Z]$. Dla n parzystego najdalej oddalony wierzchołek od źródła to $\frac{n}{2}$ a dla n nieparzystego to $\lfloor\frac{n\pm1}{2}\rfloor$. Asymptotycznie nie ma to znaczenia, możemy przyjąć $v = \frac{n}{2}$. Stąd

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

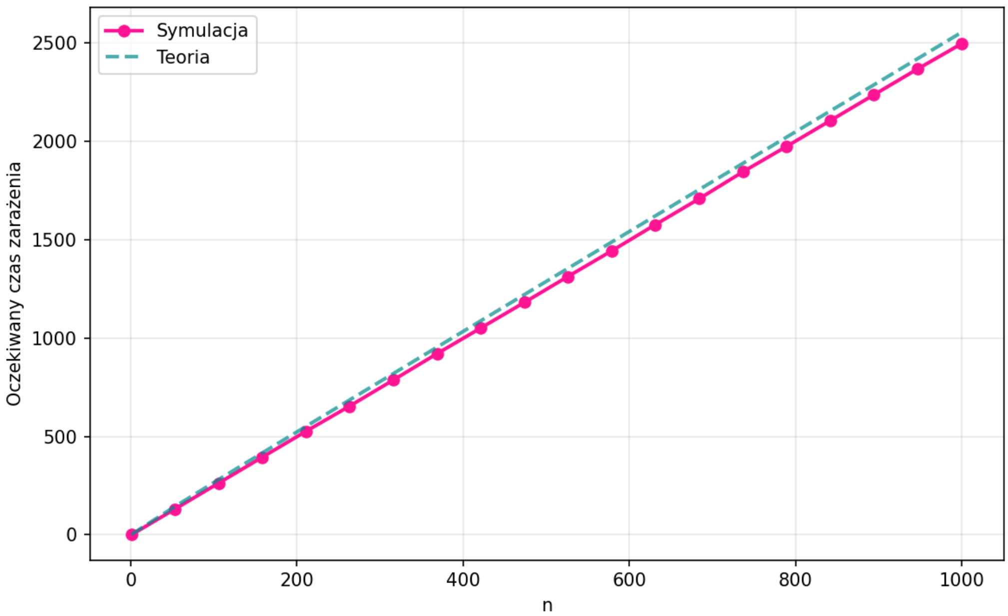
Również intuicyjny wynik. Aby się upewnić czy nie przesadziliśmy z szacowanie zwróćmy się ku symulacji. Dla $p = 0.2$, $n \in \{3, 4, \ldots, 1000\}$ policzmy wartość oczekiwaną całkowitego zarażenia po razy 2000 razy. Z wykresów (zob. Wykresie 5, Wykresie 6) widzimy, że empiryczny wynik pokrywa się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



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

4.8 Grafy pełne

Graf pełny K_n intuicyjnie powinien mieć najszybszą propagację ze względu na maksymalną liczbę krawędzi. Za źródło możemy przyjąć dowolny wierzchołek $s \in V$ ze względu na symetrie. Początkowo rozkład X_v pokrywa się z rozkładem gwiazdy natomiast w każdej kolejnej rundzie mocno się komplikuje. Zachodzi bowiem $Y_t = a$ to $\mathbb{P}[X_v = t + 1 | Y_t = a] = 1 - q^a$. Nie mamy



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

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

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

Mamy $Y_1 \sim \text{Bin}(n-1, p)$ 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

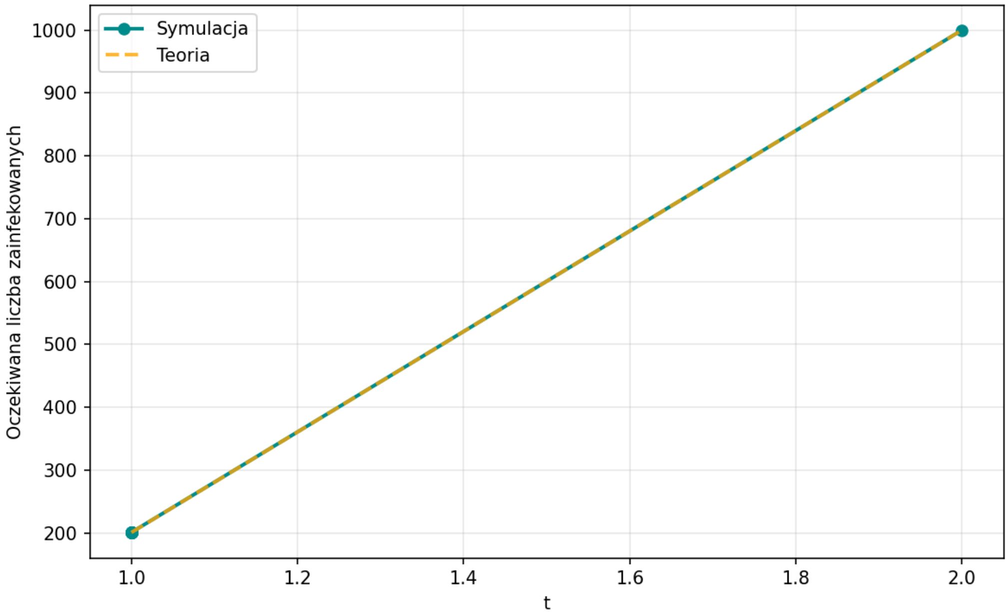
$$\mathbb{E}[Y_2 \mid Y_1 = a] \approx n(1 - q^{1+(n-1)p}) + (1 + (n-1)p)q^{1+(n-1)p}.$$

Jeśli $n \to \infty$ 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 $n \in \{2, 3, \ldots, 1000\}$ 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:

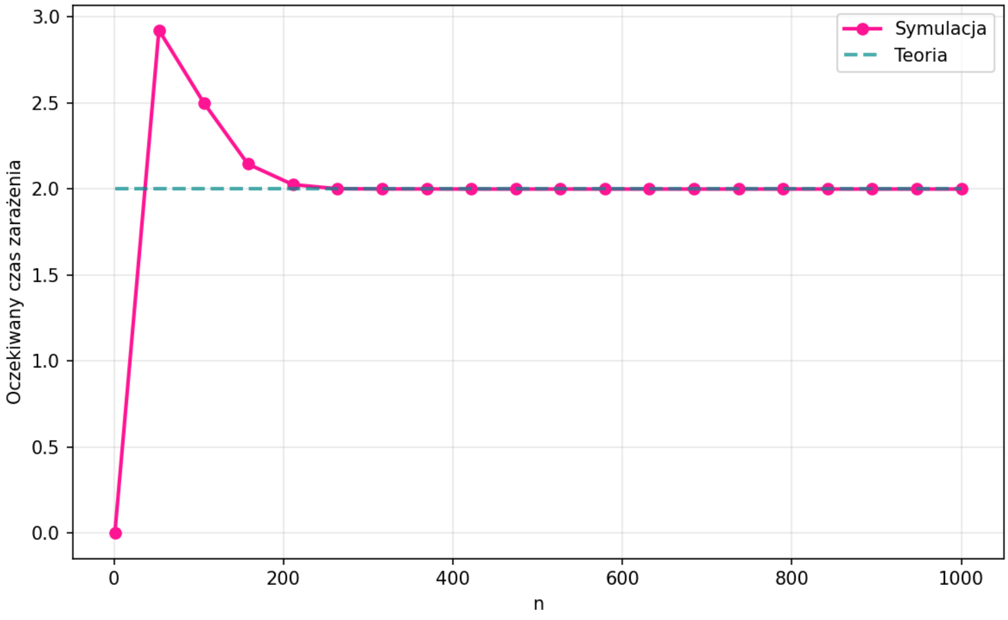
$$\lim_{n\to\infty} \mathbb{E}[Z] = 2.$$

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

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



Rysunek 7: $\mathbb{E}[Y_t]$ dla $\mathrm{K_n}$ w funkcji t



Rysunek 8: $\mathbb{E}[Z]$ dla $\mathrm{K_n}$ w funkcji n

Musimy wyznaczyć $\mathbb{E}[q^U]$ jak i $\mathbb{E}[Uq^U]$.

Lemat 3. *Niech* $X \sim \mathrm{Bin}(m,p)$. *Wtedy*

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

Dowód. Do obliczenia tych wartości posłuży nam Suma 1 jak i Suma 2.

$$\mathbb{E}[z^X] = \sum_{k=0}^{m} z^k \cdot \binom{m}{k} p^k q^{m-k} = \sum_{k=0}^{m} \binom{m}{k} (pz)^k q^{m-k} = (q+pz)^m.$$

$$\mathbb{E}[Xz^X] = \sum_{k=0}^{m} kz^k \binom{m}{k} p^k q^{m-k} = \sum_{k=0}^{m} k \binom{m}{k} (pz)^k q^{m-k} = mpz(q+pz)^{m-1}.$$

□

W naszym przypadku 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

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

Połóżmy $\varepsilon_n = (n-1)(1+p)^{-2}(1-p^2)^n$. Wtedy $\mathbb{E}[Y_2] = n - \varepsilon_n$. Z nierówności Markova (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 by 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 podgrafu. Z Twierdzenia 2 mamy

$Z \leq Z'$ a co za tym idzie $\mathbb{E}[Z^2] \leq \mathbb{E}[(Z')^2]$. Przypomnijmy, że $Z' \sim \mathrm{NegBin}(n-1, p)$ a więc $\mathbb{E}[(Z')^2] = \frac{(n-1)^2+(n-1)q}{p^2}$.

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

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

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

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 fakt, że K_n jest grafem gęstym.

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 $d_v = \mathrm{d}(s, v)$. Ustalmy $v \in V$. Skoro G jest drzewem to istnieje dokładnie jedna ścieżka od s do v, powiedzmy $s, v_1, \ldots, v_k, v$. Ponieważ infekcja rozprzestrzenia się od korzenia s wzdłuż 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 $\{s, v_1, \ldots, v_k, v\}$. Stąd

$$X_v \sim \mathrm{NegBin}(\mathrm{d}_v, p),$$

oraz

$$\mathbb{E}[X_v] = \frac{\mathrm{d}_v}{p}, \quad \mathrm{Var}[X_v] = \frac{\mathrm{d}_v \cdot (1-p)}{p^2}.$$

Lemat 4. *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:

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

□

Przejdźmy teraz to obliczania średniej liczby zainfekowanych wierzchołków w czasie t. Oznaczmy przez $F(t; m, p)$ dystrybuante zmiennej o rozkładzie $\mathrm{NegBin}(m, p)$. Z Lematu 4 otrzymujemy

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

Połóżmy $a_j = |\{v \in V : d_v = j\}|$ 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 $L = \{u_1, \ldots, u_m\}$ będzie zbiorem liści w G. Wtedy mamy $Z = \max_{u \in L} X_u$.

Zauważmy, że $\epsilon(s) = \max_{u \in L} \mathrm{d}_u$ i jest to wysokość drzewa. Oznaczmy ją przez h. Z nierówności Jensena (7) otrzymujemy

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

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 $h \in \{1, 2, \ldots, n-1\}$ i poszukajmy drzew 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 oczekiwaną 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 $a_1 = \cdots = a_h$. Czyli innymi słowy, nasze drzewo ma tyle samo węzłów na każdej głębokości. Połóżmy $a_1 = b$. 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 analize 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

Analiza modelu SIS

5.1 Dwa wierzchołki, jedna krawędź

W celu oswojenia się z bardziej skomplikowanym modelem, jakim jest **SIS**, przeanalizujmy graf K_2. Niech $V = \{u, v\}$ oraz niech u będzie wierzchołkiem startowym. Istnieją cztery możliwe stany systemu: $\mathcal{I}_t = \varnothing$, $\mathcal{I}_t = \{u\}$, $\mathcal{I}_t = \{v\}$, $\mathcal{I}_t = \{u, v\}$. Stan, w którym żaden wierzchołek nie jest zainfekowany, jest stanem absorbującym. Ponadto, z każdego pozostałego stanu możemy przejść do dowolnego innego. 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,

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

A zatem

$$\mathbb{P}[X_v = t] = p(q\beta)^{t-1}, \quad t \geq 1.$$

oraz

$$\mathbb{P}[X_v < \infty] = \frac{p}{1 - q\beta}, \quad \mathbb{P}[X_v = \infty] = \frac{q\alpha}{1 - q\beta}.$$

Z założeń modelu $q, \alpha \in (0; 1)$, a zatem $\mathbb{P}[X_v = \infty] \neq 0$. Mamy $\mathbb{E}[X_v] = \infty$. Powinniśmy się więc spodziewać, że niezależnie od wartości parametrów p oraz α infekcja się nie rozprzestrzeni. Nie jest to jednak pożądany rezultat. Wartość oczekiwaną warunkowa wynosi:

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

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:

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

Połóżmy

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

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 oczekiwaną, to mamy $\mathbb{E}[Y_t] = 0 \cdot a_t + 1 \cdot b_t + 2 \cdot c_t = b_t + 2c_t$. Z Twierdzenia 4 mamy

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

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:

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

5.2 Pewne wygaśnięcie

W powyższych obliczeniach mogliśmy zauważyć istotną różnicę pomiędzy modelem **SI** a **SIS**. Mianowicie jest niezerowe prawdopodobieństwo, że wierzchołki w grafie nigdy nie zostaną zainfekowane, a oczekiwana liczba zarażonych dąży wraz z upływem czasu do zera. Dla tak małego grafu jakim jest K_2 jest to dość niespodziewany rezultat. Postaramy się teraz udowodnić tą obserwację dla dowolnego grafu.

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

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

Dowód. Weźmy $v \in V \setminus \{s\}$. Jeśli $\{s, v\} \notin E$ to jedną z możliwości, która sprawi, że $X_v = \infty$, to wygaśnięcie infekcji po pierwszej jednostce czasu. Zatem $\mathbb{P}[X_v = \infty] \geq \alpha$. Gdy zaś $\{s, v\} \in E$ to $\mathbb{P}[X_v = \infty] \geq q\alpha$, bo musimy jeszcze zagwarantować, że v nie zostanie zainfekowany zanim infekcja w s wygaśnie. Skoro $q, \alpha > 0$ to otrzymujemy $\mathbb{P}[X_v = \infty] > 0$.

Rozważmy propagację jako łańcuch Markova, gdzie stanem w czasie t jest $\mathcal{I}_t$. Prawdopodobieństwo przejścia do stanu bez infekcji wynosi

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

Istotne jest to, że jest to dodatnia liczba. Kładziemy

$$r^* = \min\{r(\mathcal{I}_t) : \mathcal{I}_t \subseteq V\}.$$

Prawdopodobieństwo przetrwania infekcji po k rundach wynosi $(1 - r^*)^k$. Ale skoro $r^* > 0$ to wartość ta dąży do 0 wraz z $k \to \infty$. Mamy więc $\mathbb{P}[Z = \infty] = 0$ co jest równoważne z $\mathbb{P}[Z < \infty] = 1$.

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

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

□

5.3 Grafy pełne

Po analizie dla K_2 postarajmy się ją uogólnić dla K_n. Ze względu na symetrie dowolny stan z taką samą liczbą zainfekowanych wierzchołków jest izomorficzny. Z Twierdzenia 4 wiemy, że $\mathbb{E}[Y_t] \to 0$. Ustalmy $t \in \mathbb{N}$. Niech A_t będzie zmienną losową oznaczająca 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 $A_t \sim \mathrm{Bin}(i, \beta)$. Dalej każdy podatny wierzchołek może zostać zarażony niezależnie przez $n - i$ sąsiadów. Szansa, że któremuś z nich się uda wynosi $1 - q^i$. Stąd $B_t \sim \mathrm{Bin}(n - i, 1 - q^i)$. 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 $P_{i\to j} = \mathbb{P}[Y_{t+1} = j | Y_t = i]$ dla $j \in \{0, 1, \dots, n\}$. Wtedy korzystając z Faktu 7 mamy

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

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

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

Macierz przejść dla naszego łańcucha Markova dana jest

$$\mathbf{P} = \left[P_{i\to j}\right]_{0\le i,j\le n}.$$

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

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

Licząc wartość oczekiwaną dostajemy

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

Rozdział 6

Analiza modelu SIR

6.1 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. Rozkład X_v będzie tutaj taki sam jak dla modelu **SIS** a więc

$$\mathbb{P}[X_v = t] = p(q\beta)^{t-1}, \quad t \geq 1$$

oraz

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

Dalej 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 $t \to \infty$ 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 $Z \sim \mathrm{Geo}(1-q\beta)$. Mamy

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

6.2 Pewne wygaśnięcie

Podobnie jak w **SIS** w modelu **SIR** wygaśnięcie infekcji jest zdarzeniem pewnym. Różnica polega na tym, że wszystkie kiedykolwiek zainfekowane węzły przejdą w stan wyzdrowiały. Sformalizujmy ten fakt prostym twierdzeniem.

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

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

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

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

Dowód. Fakt $\mathbb{P}[X_v = \infty] > 0$ możemy udowodnić analogicznie jak w dowodzie Twierdzenia 4. Podobnie $\mathbb{P}[Z < \infty] = 1$ bo wymarcie infekcji jest stanem absorbującym. Zauważmy, że dla $v \in V$ mamy

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

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

6.3 Grafy gwiezdne

Rozpatrzmy propagacje na rodzinie S_n, gdzie źródłem jest wierzchołek 0. Dla $v \in \{1, 2, \ldots, n\}$ mamy $X_v \sim \mathrm{KGeo}(p, \alpha)$. Zauważmy natomiast, że zmienne $X_1, X_2, \ldots, X_n$ sa zależne, bo gdy centralny wierzchołek wyzdrowieje żaden z liści nie może już zostać zainfekowany.

Przyjrzyjmy się teraz zmiennej Y_t. Połóżmy $C = \min\{\tau \in \mathbb{N} : 0 \in \mathcal{R}_\tau\}$. Jeśli centrum zarazi się po j rundach to liście gwiazdy mogą byc zarażane przez $\min\{j, t\}$ rund. Y_t zachowuje się tak samo jak w modelu **SI** dla grafów gwiezdnych. Przypomnijmy, że rozkład ten wynosi $1 + B_{j,t}$ gdzie $B_j \sim \mathrm{Bin}(n, 1 - q^{\min\{j,t\}})$. 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 $C \sim \mathrm{Geo}(\alpha)$ 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 4. 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 $t \to \infty$ 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 zwięzłej 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

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

Dalej

$$\int_0^\infty e^{-bx}(1-(1-e^{-ax})^n)\,\mathrm{d}x = \int_0^\infty e^{-bx}\mathrm{d}x - \int_0^\infty e^{-bx}(1-e^{-ax})^n \mathrm{d}x.$$

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 $\mathrm{d}x = \frac{1}{a}e^{ax}\mathrm{d}u$. Mamy zatem $e^{-bx} = (1-u)^{\frac{b}{a}},\ e^{ax} = \frac{1}{1-u}$. A więc całka wynosi

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

gdzie skorzystaliśmy z Faktu 6. Ostatecznie otrzymujemy

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