

# **Probabilistyczne modele propagacji w grafach**

Bartosz Łabuz

Kierunek: Informatyka Algorytmiczna (INA)

Promotor: dr Dominik Bojko

Słowa kluczowe: propagacja losowa, modele epidemiologiczne, teoria grafów

## Streszczenie

Niniejsza praca poświęcona jest analizie matematycznej procesów probabilistycznej propagacji w grafach nieskierowanych. Rozważane są trzy klasyczne modele epidemiologiczne: model SI, w którym wierzchołki po zarażeniu pozostają zainfekowane na zawsze; model SIR, gdzie zainfekowane wierzchołki mogą przejść do stanu wyzdrowiałego; oraz model SIS, w którym wierzchołki mogą ponownie stać się podatne na zarażenie. Głównym celem pracy jest wyznaczenie rozkładów prawdopodobieństwa oraz wartości oczekiwanych kluczowych zmiennych losowych opisujących dynamikę propagacji, takich jak czas zarażenia poszczególnych wierzchołków, liczba zarażonych wierzchołków w danej chwili oraz czas całkowitego zarażenia lub wygaśnięcia infekcji. Praca wykorzystuje narzędzia z teorii prawdopodobieństwa, teorii grafów oraz teorii procesów stochastycznych. Wszystkie rezultaty teoretyczne zostały zweryfikowane poprzez symulacje numeryczne. Praca kończy się omówieniem możliwych zastosowań praktycznych oraz kierunków dalszych badań.

## Abstract

This thesis is devoted to the mathematical analysis of probabilistic propagation processes in undirected graphs. Three classical epidemiological models are considered: the SI model, where vertices remain infected forever once infected; the SIR model, where infected vertices can transition to a recovered state; and the SIS model, where vertices can become susceptible to infection again. The main objective is to determine probability distributions and expected values of key random variables describing propagation dynamics, such as infection time of individual vertices, number of infected vertices at a given time, and time to total infection or extinction. The work employs tools from probability theory, graph theory, and the theory of stochastic processes. All theoretical results were verified through numerical simulations. The thesis concludes with a discussion of potential practical applications and directions for future research.

# Spis treści

<b>1 Wstęp</b>	<b>5</b>
1.1 Motywacja i zastosowania . . . . .	5
1.2 Cel pracy . . . . .	5
1.3 Zakres pracy . . . . .	6
<b>2 Podstawy matematyczne</b>	<b>7</b>
2.1 Notacja . . . . .	7
2.2 Rodziny grafów . . . . .	7
2.3 Rozkłady prawdopodobieństwa . . . . .	8
2.4 Fakty, sumy i nierówności . . . . .	10
<b>3 Modele propagacji losowej</b>	<b>13</b>
3.1 Model SI . . . . .	13
3.2 Model SIR . . . . .	14
3.3 Model SIS . . . . .	15
<b>4 Scenariusz aplikacyjny</b>	<b>17</b>
4.1 Model SI — Sieć bez zabezpieczeń . . . . .	17
4.2 Model SIR — Sieć z automatycznymi łatkami bezpieczeństwa . . . . .	18
4.3 Model SIS — Sieć z tymczasowymi zabezpieczeniami . . . . .	19
<b>5 Analiza modelu SI</b>	<b>21</b>
5.1 Dwa wierzchołki, jedna krawędź . . . . .	21
5.2 Trójkąt . . . . .	21
5.3 Całkowita infekcja pewna . . . . .	22
5.4 Grafy ścieżkowe . . . . .	23
5.5 Grafy gwiazdne . . . . .	24
5.6 Uniwersalne ograniczenia na czas zarażenia . . . . .	25
5.7 Grafy cykliczne . . . . .	27
5.8 Grafy pełne . . . . .	29
5.9 Drzewa . . . . .	31
5.10 Eksperymenty . . . . .	32
<b>6 Rozkłady umierające</b>	<b>37</b>
6.1 Rozkład umierający geometryczny . . . . .	37
6.2 Rozkład umierający ujemny dwumianowy . . . . .	38

<b>7 Analiza modelu SIR</b>	<b>39</b>
7.1 Dwa wierzchołki, jedna krawędź	39
7.2 Pewne wygaśnięcie	40
7.3 Grafy ścieżkowe	40
7.4 Grafy gwiazdne	41
7.5 Eksperymenty	43
<b>8 Analiza modelu SIS</b>	<b>47</b>
8.1 Pewne wygaśnięcie	47
8.2 Dwa wierzchołki, jedna krawędź	47
8.3 Grafy pełne	49
8.4 Eksperymenty	50
<b>9 Podsumowanie</b>	<b>55</b>
9.1 Wnioski	55
9.2 Kierunki dalszych badań	56
<b>Bibliografia</b>	<b>59</b>

# Rozdział 1

## Wstęp

### 1.1 Motywacja i zastosowania

Propagację zjawisk w sieciach ludzkość obserwuje od starożytności — od epidemii chorób zakaźnych, przez rozprzestrzenianie się plotek i pogłosek, aż po współczesne wirusy komputerowe i kampanie marketingowe w mediach społecznościowych. Zrozumienie mechanizmów tej propagacji ma kluczowe znaczenie praktyczne: pozwala przewidywać rozwój epidemii, projektować skuteczne kampanie informacyjne, zabezpieczać infrastrukturę cyfrową przed atakami czy optymalizować algorytmy w sieciach rozproszonych. Najbardziej naturalną reprezentacją matematyczną sieci relacji są grafy: wierzchołki reprezentują jednostki (ludzi, komputery, węzły sieci), a krawędzie określają połączenia między nimi (kontakty społeczne, łącza sieciowe, kanały komunikacji). Połączenie teorii grafów z rachunkiem prawdopodobieństwa pozwala stworzyć precyzyjne modele stochastyczne, które uwzględniają losowy charakter propagacji — nie każdy kontakt prowadzi do transmisji, a różne jednostki mogą różnić się podatnością na zarażenie. Modele propagacji znajdą zastosowanie w epidemiologii (modelowanie rozprzestrzeniania się chorób zakaźnych), cyberbezpieczeństwie (analiza rozprzestrzenniania się wirusów komputerowych), marketingu (viralowe kampanie reklamowe), teorii algorytmów (propagacja wartości ekstremalnych w sieciach rozproszonych) oraz naukach społecznych (dyfuzja innowacji, rozprzestrzenianie się opinii).

### 1.2 Cel pracy

Celem niniejszej pracy jest kompleksowa analiza probabilistycznych modeli propagacji w grafach, ze szczególnym uwzględnieniem wpływu topologii sieci na dynamikę procesów stochastycznych. Konkretnie, praca ma na celu:

- zamodelowanie rozprzestrzeniania się informacji w sieciach przy użyciu procesów grafowych,
- teoretyczna analiza procesów losowej propagacji w grafach,
- teoretyczne wyznaczenie rozkładów prawdopodobieństwa i wartości oczekiwanych kluczowych zmiennych losowych dla wybranych rodzin grafów,
- przeprowadzenie symulacji komputerowych w celu weryfikacji wyników teoretycznych,

- zbadanie, jak topologia grafu wpływa na tempo i zasięg propagacji,
- przedstawienie praktycznego scenariusza aplikacyjnego ilustrującego użyteczność modeli.

## 1.3 Zakres pracy

Praca składa się z następujących części:

**Rozdział 2: Podstawy matematyczne.** Zawiera notację używaną w całej pracy, definicje badanych rodzin grafów, wykorzystywane rozkłady prawdopodobieństwa wraz z ich własnościami, oraz zbiór faktów, sum i nierówności stosowane w dowodach i obliczeniach.

**Rozdział 3: Modele propagacji losowej.** Wprowadza formalne definicje trzech badanych modeli: SI, SIR i SIS. Dla każdego modelu definiujemy zmienne losowe opisujące kluczowe aspekty propagacji oraz diagramy przejść między stanami.

**Rozdział 4: Scenariusz aplikacyjny.** Ilustruje praktyczne zastosowanie modeli na przykładzie rozprzestrzeniania się wirusa komputerowego w różnych topologiiach sieciowych.

**Rozdział 5: Analiza modelu SI.** Przeprowadzamy szczegółową analizę teoretyczną modelu SI dla rodzin grafów ścieżkowych, gwiazdkowych, cyklicznych, pełnych oraz drzew. Wyznaczamy rozkłady prawdopodobieństwa i wartości oczekiwane zmiennych zdefiniowanych dla tego modelu. Wyprowadzamy ogólne ograniczenia na czas zarażenia. Prezentujemy wyniki symulacji potwierdzające teoretyczne przewidywania.

**Rozdział 6: Rozkłady umierające.** Wprowadzamy rozkłady umierający geometryczny i umierający ujemny dwumianowy, które modelują propagację z możliwością przerwania procesu. Rozkłady te są potrzebne do analizy modeli SIR i SIS, gdzie infekcja może wygasnąć.

**Rozdział 7: Analiza modelu SIR.** Koncentruje się na modelu SIR dla grafów ścieżkowych i gwiazdowych. Wyznaczamy rozkłady i wartości oczekiwane odpowiednich zmiennych losowych. Wyniki weryfikujemy symulacjami.

**Rozdział 8: Analiza modelu SIS.** W modelu SIS ograniczamy się do grafu pełnego. Wykorzystujemy formalizm łańcuchów Markova do wyznaczenia interesujących nas rozkładów. Przeprowadzamy symulacje dla różnych wartości parametrów pokazujące oscylacyjny charakter propagacji.

# 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  $\lim_{x \rightarrow \pm\infty} f(x)$ .

Niech  $G = (V, E)$  będzie grafem prostym nieskierowanym. Przez  $N(v) = \{u \in V : \{u, v\} \in E\}$  oznaczamy zbiór sąsiadów wierzchołka  $v$ . 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 (w skrócie PMF) oznaczamy  $\mathbb{P}[X = t]$  a dystrybuante (w skrócie CDF) oznaczamy  $F_X(t) = \mathbb{P}[X \leq t]$  dla  $t \in \mathbb{R}$ . Mówimy, że zmienne losowe  $X_1, \dots, X_n$  są IDD (independent and identically distributed) jeżeli są one niezależne oraz  $F_{X_1} = \dots = F_{X_n}$ .

### 2.2 Rodziny grafów

W tej pracy ograniczymy rozważanie propagacji do konkretnych rodzin grafów. Są to grafy ścieżkowe, gwiazdne, cykliczne oraz pełne. Poniżej znajdują się szczegółowe definicje zbioru wierzchołków  $V$  oraz zbioru krawędzi  $E$  tych rodzin dla  $n \in \mathbb{N}_+$ . Dodatkowo opisy te zawierają wizualizacje reprezentanta rodziny dla  $n = 5$ .

#### Graf ścieżkowy $P_n$

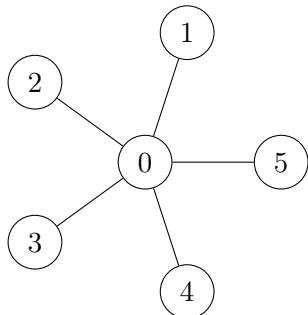
$$V = \{1, 2, \dots, n\}, \quad E = \{\{i, i + 1\} : i \in \{1, 2, \dots, n - 1\}\}.$$



Rysunek 2.1: Graf  $P_5$

#### Graf gwiazda $S_n$

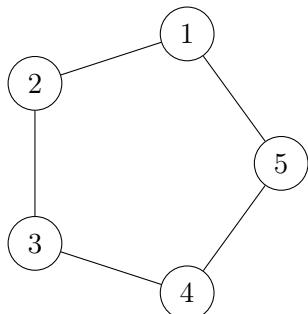
$$V = \{0, 1, \dots, n\}, \quad E = \{\{0, i\} : i \in \{1, 2, \dots, n\}\}.$$



**Rysunek 2.2:** Graf  $S_5$

**Graf cykliczny  $C_n$**

$$V = \{1, 2, \dots, n\}, \quad E = \{\{i, i + 1\} : i \in \{1, 2, \dots, n - 1\}\} \cup \{\{n, 1\}\}.$$



**Rysunek 2.3:** Graf  $C_5$

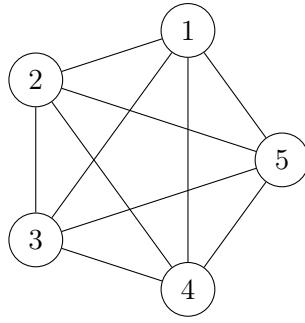
**Graf pełny  $K_n$**

$$V = \{1, 2, \dots, n\}, \quad E = \{\{i, j\} : i, j \in \{1, 2, \dots, n\} \wedge i \neq j\}.$$

## 2.3 Rozkłady prawdopodobieństwa

Przedstawiamy specyfikacje klasycznych rozkładów prawdopodobieństwa obejmującą ich funkcje masy, wartości oczekiwane i wariancje.

**Rozkład Bernoulliego**  $Ber(p)$



**Rysunek 2.4:** Graf  $K_5$

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. Wtedy:

$$\mathbb{P}[X = 0] = 1 - p, \quad \mathbb{P}[X = 1] = p.$$

Wartość oczekiwana i wariancja wynoszą:

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

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

Zmienna losowa  $X$  ma rozkład dwumianowy, jeżeli opisuję 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:

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

### 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:

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

### 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:

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

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

Dla  $t \in \mathbb{R}$  definiujemy 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:

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

## 2.4 Fakty, sumy i nierówności

W tej sekcji gromadzimy używane w pracy fakty pomocnicze, wzory sumacyjne oraz nierówności dotyczące zmiennych losowych i przekształceń analitycznych.

**Fakt 1.** Niech  $X_1, \dots, X_n$  będą IID zmiennymi losowymi o CDF równej  $F_X$ . Położmy  $Y = \max\{X_1, \dots, X_n\}$ . Wtedy

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

**Fakt 2.** Niech  $X_1, \dots, X_n$  będą IID zmiennymi losowymi o CDF równej  $F_X$ . Położmy  $Y = \min\{X_1, \dots, X_n\}$ . 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  $X \sim \text{Bin}(n, p)$  oraz  $Y \sim \text{Bin}(m, r)$  będą niezależnymi zmiennymi losowymi. Położmy  $Z = X + Y$ . Wtedy dla  $k \in \{0, 1, \dots, n+m\}$  zachodzi

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

**Fakt 5.** 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 6.** Niech  $X_1, \dots, X_m \sim \text{Geo}(p)$  będą IID zmiennymi losowymi. Wtedy

$$X_1 + \dots + X_m \sim \text{NegBin}(m, p).$$

**Fakt 7.** 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

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

**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

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

**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  $x_1, \dots, x_n \geq 0$ . 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 Markova). Niech  $X$  będzie zmienną losową taką, że  $X \geq 0$  oraz  $\mathbb{E}[X] < \infty$ . Wtedy dla dowolnego  $t > 0$  zachodzi

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

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

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

W szczególności, ponieważ  $\max$  jest funkcją wypukłą, mamy:

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

# 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  $Q \in \mathcal{Q}$ . Nie będzie nas interesować przestrzeni zdarzeń elementarnych tego procesu. Definiujemy realizacje zmiennej losowej  $\mathbf{X} : \mathbb{N} \times V \rightarrow \mathcal{Q}$  określającą stany wierzchołków grafu w czasie. Mamy  $\mathbf{X}_t(v) = Q$  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 ustalone źródło  $s \in V$  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

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

Rozkład prawdopodobieństwa w tym modelu jest definiowany przez następujące zależności:

$$\begin{aligned} \mathbb{P}[\mathbf{X}_{t+1}(v) = S \mid \mathbf{X}_t(v) = S] &= q^{i_v}, \\ \mathbb{P}[\mathbf{X}_{t+1}(v) = I \mid \mathbf{X}_t(v) = S] &= 1 - q^{i_v}, \\ \mathbb{P}[\mathbf{X}_{t+1}(v) = S \mid \mathbf{X}_t(v) = I] &= 0, \\ \mathbb{P}[\mathbf{X}_{t+1}(v) = I \mid \mathbf{X}_t(v) = I] &= 1. \end{aligned}$$

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

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

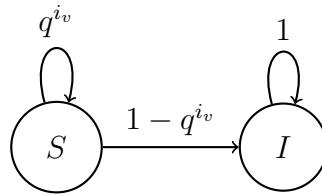
Jeśli taka chwila nie istnieje (tzn. w danym przebiegu procesu wierzchołek  $v$  nigdy się nie zarazi), to przyjmujemy  $X_v = \infty$ . Później udowodnimy ([Twierdzenie 1](#)), że w modelu SI mamy  $\mathbb{P}[X_v = \infty] = 0$ . Wyznaczenie rozkładu  $X_v$ , jak i  $\mathbb{E}[X_v]$  da nam sporo 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  $t \rightarrow \infty$  ([Twierdzenie 1](#)). 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. Wyznacznie rozkładu  $Z$  oraz jej wartości oczekiwanej dla konkretnych rodzin grafów będzie głównym celem w tym modelu.



**Rysunek 3.1:** Diagram przejść dla modelu SI.

## 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 pochłaniający — wierzchołek, który wyzdrowiał, nie może już ani się zarażić, ani nikogo zakazić. Zarażony wierzchołek może przejść z  $I$  do stanu  $R$  z prawdopodobieństwem  $\alpha \in (0; 1)$ . Przyjmujemy, że w każda runda odbywa się w dwóch etapach. W pierwszym z nich wierzchołki przekazują infekcję. W drugim z nich te, które były już wcześniej zainfekowane 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:

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

Tak jak poprzednio rozważmy zmienne

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

W kontraste dla modelu SI zachodzi  $\mathbb{P}[X_v = \infty] > 0$  ([Twierdzenie 4](#)). Z tego powodu wartość oczekiwana  $\mathbb{E}[X_v] = \infty$  niezależnie od struktury grafu. Dlatego też 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 oraz  $X_v$ . Z tego powodu tego nie będziemy jej rozważać. Zamiast rozważać liczbę tylko zainfekowanych lub tylko wyzdrowiałych wierzchołków będziemy rozważać liczbę niepodatnych wierzchołków po  $t$  krokach. Kładziemy więc

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

Mamy  $\mathbb{P}[\mathcal{I}_t = \emptyset] \rightarrow 1$  wraz z  $t \rightarrow \infty$  ([Twierdzenie 4](#)) a więc możemy zdefiniować zmienną  $Z$  oznaczającą czas wygaśnięcia infekcji. Dla uproszczenia będziemy rozważać moment w którym żaden nowy wierzchołek nie może być zarażony. Zatem

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

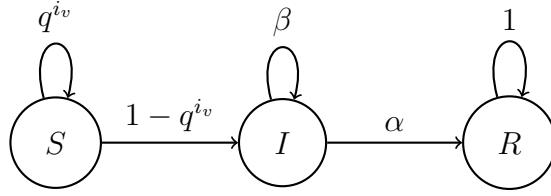
Dodatkowo definiujemy zmienną  $W$  będącą liczbą finalnie wyzdrowiałych wierzchołków.

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

Rozkłady  $Z, W$  jak i ich wartości oczekiwane będą naszym głównym obiektem zainteresowań 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 tak jak w SI



**Rysunek 3.2:** Diagram przejścia dla modelu SIR.

mamy  $\mathcal{Q} = \{S, I\}$ . W modelu SIS liczba zainfekowanych wierzchołków może oscylować w czasie i nie musi osiągnąć stanu pełnego zakażenia. Dla uproszczenia notacji kładziemy  $\beta = 1 - \alpha$ . Każda runda składa się z dwóch faz. W pierwszej z nich wierzchołki zarażone próbują przekazać infekcje. Natomiast w drugiej te zarażone wierzchołki, które były już zarażone przez fazą pierwszą, mogą powrócić w stan podatności. Jest to odzwierciedlenie naturalnego stanu rzeczy. Jeśli ktoś zostanie zakażony chorobą to nie może wyzdrowieć szybciej niż jego zakaziciel. Rozkład prawdopodobieństwa w tym modelu jest definiowany przez następujące zależności:

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

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

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

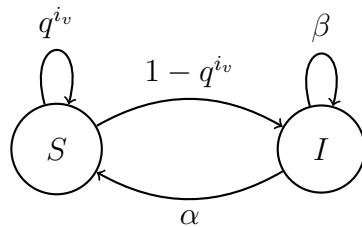
oraz dla  $t \in \mathbb{N}$  definiujemy

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

Wygaśnięcie infekcji zachodzi z prawdopodobieństwem 1 ([Twierdzenie 5](#)). A więc definiujemy zmienną losową opisującą czas wygaśnięcia infekcji:

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

Mamy  $\mathbb{P}[X_v = \infty] > 0$  a zatem interesować nas będzie  $\mathbb{E}[X_v | X_v < \infty]$ . Dalej wykażemy, że  $\mathbb{E}[Y_t]$  dążą do 0. ([Twierdzenie 5](#)) Dlatego też także nie będą nas interesować. Skupimy się tylko na rozkładzie  $Y_t$ . Jeśli chodzi o zmienną  $Z$  to przyjrzymy się zarówno jej rozkładowi jak i wartości oczekiwanej.



**Rysunek 3.3:** Diagram przejścia dla modelu SIS.

# Rozdział 4

## Scenariusz aplikacyjny

Modele propagacji SI, SIR i SIS mają szerokie zastosowanie w analizie rozprze-strzeniania się zjawisk w sieciach. Jednym z naturalnych zastosowań jest modelowa-nie rozprzestrzeniania się wirusa komputerowego w sieci. W tym rozdziale przedsta-wimy szczegółowy scenariusz aplikacyjny dla każdego z trzech modeli, pokazując jak zmienne losowe zdefiniowane wcześniej pozwalają na analizę różnych aspektów epidemii cyfrowej. Rozważamy sieć komputerową reprezentowaną przez graf  $G = (V, E)$ , gdzie wierzchołki odpowiadają poszczególnym komputerom, a krawędzie reprezentują bezpośrednie połączenia sieciowe (np. w sieci lokalnej, klastrze serwerów lub infrastrukturze chmurowej). Wirus rozpoczyna swoje rozprzestrzenianie od pojedynczego zainfekowa-nego komputera  $s \in V$ . W każdym cyklu czasowym zainfekowane komputery próbują zarazić bezpośrednio połączone z nimi komputery z prawdopodobieństwem  $p$ , które modeluje skuteczność transmisji wirusa przez połączenie sieciowe.

### 4.1 Model SI — Sieć bez zabezpieczeń

Rozważmy starszą infrastrukturę IT, która nie posiada żadnych aktywnych za-bezpieczeń antywirusowych ani mechanizmów automatycznego wykrywania zagrożeń. Może to być na przykład sieć przemysłowych systemów SCADA, które często działają na przestarzałym oprogramowaniu bez regularnych aktualizacji bezpieczeństwa, lub izolowana sieć wewnętrzna firmy z przestarzałymi systemami operacyjnymi. W takim środowisku, gdy komputer zostanie raz zainfekowany, pozostaje w tym stanie perma-nentnie — nie ma mechanizmu samoczynnego usunięcia wirusa. Model SI doskonale opisuje tę sytuację, gdzie  $S$  reprezentuje komputery podatne na infekcję, a  $I$  komputery już zainfekowane.

Dla każdego komputera  $v \in V$  zmienna  $X_v$  określa, w której godzinie cyklu ten konkretny komputer zostanie zainfekowany. Jest to kluczowa informacja dla admini-stratora sieci próbującego zrozumieć dynamikę ataku. Rozkład prawdopodobieństwa  $X_v$  pozwala odpowiedzieć na pytanie: jakie jest prawdopodobieństwo, że dany kom-puter zostanie zarażony w określonym czasie? Wartość oczekiwana  $\mathbb{E}[X_v]$  wskazuje średni czas, po którym można spodziewać się infekcji tego komputera. W kontekście praktycznym, znajomość rozkładu  $X_v$  pozwala na odpowiednią priorytetyzację działań obronnych — komputery o małym  $\mathbb{E}[X_v]$  powinny być zabezpieczone jako pierwsze, planowanie harmonogramu ręcznej interwencji technicznej a także oszacowanie okna czasowego na izolację krytycznych systemów.

Zmienna  $Y_t$  określa, ile komputerów jest zainfekowanych w chwili  $t$ . Rozkład  $Y_t$  po-

kazuje możliwe scenariusze rozwoju epidemii — czy atak rozwija się szybko, czy powoli. Wartość oczekiwana  $\mathbb{E}[Y_t]$  daje średnią liczbę zainfekowanych komputerów po czasie  $t$ . Z punktu widzenia zarządzania infrastrukturą,  $Y_t$  pozwala na monitorowanie tempa rozprzestrzeniania się wirusa czy ocenę skuteczności częściowych działań obronnych.

Zmienna  $Z$  opisuje, po ilu cyklach czasowych wszystkie komputery w sieci zostaną zainfekowane. Jest to moment, w którym cała infrastruktura zostaje skompromitowana. Rozkład  $Z$  określa prawdopodobieństwo różnych scenariuszy czasowych dla całkowitego przejęcia sieci. Wartość oczekiwana  $\mathbb{E}[Z]$  wskazuje średni czas, jaki administrator ma na reakcję zanim stracimy kontrolę nad całą siecią. Znajomość  $Z$  jest krytyczna dla określenia maksymalnego czasu reakcji zespołu bezpieczeństwa czy też oceny ryzyka związanego z konkretną architekturą sieci. Pozwala to na projektowania topologii sieci odpornej na szybkie rozprzestrzenianie się zagrożeń.

## 4.2 Model SIR — Sieć z automatycznymi łączkami bezpieczeństwa

Rozważmy nowoczesną infrastrukturę korporacyjną z aktywnym systemem zarządzania aktualizacjami i automatycznymi mechanizmami odpowiedzi na incydenty. Gdy komputer zostaje zainfekowany, system bezpieczeństwa może wykryć zagrożenie i zastosować łączkę zabezpieczającą, przenosząc komputer do stanu odpornego ( $R$ ). Prawdopodobieństwo  $\alpha$  modeluje skuteczność systemu automatycznych łączek — im większe  $\alpha$ , tym szybciej komputery są naprawiane. W modelu SIR komputer przechodzi przez stany: podatny ( $S$ )  $\rightarrow$  zainfekowany ( $I$ )  $\rightarrow$  odporny ( $R$ ). Stan  $R$  jest absorpcyjny — załączany komputer nie może być ponownie zainfekowany tym samym wariantem wirusa. To realistyczne założenie w kontekście nowoczesnych systemów, gdzie zaktualizowane oprogramowanie jest odporne na znane exploity.

W modelu SIR zmienna  $X_v$  nadal określa czas pierwszego zarażenia komputera  $v$ . Jednak w przeciwieństwie do modelu SI, nie każdy komputer musi zostać kiedykolwiek zainfekowany — istnieje możliwość, że wirus wygaśnie zanim dotrze do wszystkich węzłów. Zachodzi bowiem  $\mathbb{P}[X_v = \infty] > 0$ , co oznacza, że część komputerów może uniknąć infekcji całkowicie. Z tego powodu interesującą jest warunkowe oczekiwanie  $\mathbb{E}[X_v | X_v < \infty]$  — średni czas zarażenia dla tych komputerów, które rzeczywiście zostały zainfekowane. Rozkład  $X_v$  pozwala na identyfikację komputerów w strefie największego ryzyka.

W modelu SIR definiujemy  $Y_t$  jako liczbę komputerów, które nie są już w stanie podatnym. Pokazuje to postęp epidemii z innej perspektywy — ile systemów zostało już „dotkniętych” (zarażonych lub załączanych). Wraz z czasem  $Y_t$  stabilizuje się, osiągając wartość równą  $W$ .

Zmienna  $W$  określa całkowitą liczbę komputerów, które zostały zainfekowane w trakcie całej epidemii. To kluczowa metryka dla oceny skali incydentu bezpieczeństwa. Rozkład  $W$  pokazuje możliwe scenariusze zasięgu ataku — czy zainfekowane zostaną tylko nieliczne komputery w pobliżu źródła, czy znaczna część sieci. Wartość oczekiwana  $\mathbb{E}[W]$  daje średnią liczbę skompromitowanych systemów, co bezpośrednio przekłada się na szacowanie kosztów naprawy i przywrócenia systemów, ocenę potencjalnego wycieku danych lub szkód biznesowych, planowanie zasobów zespołu reagowania na incydenty.

Zmienna  $Z$  określa moment, w którym epidemia wygasła — żaden nowy komputer

nie może już zostać zainfekowany, ponieważ wszystkie zainfekowane komputery albo wyzdrowiały, albo nie mają podatnych sąsiadów. To moment, w którym incydent bezpieczeństwa można uznać za opanowany. Rozkład  $Z$  i wartość  $\mathbb{E}[Z]$  są krytyczne dla planowania czasu trwania trybu kryzysowego w organizacji, oszacowania czasu, przez który zespół bezpieczeństwa musi działać w trybie podwyższonej gotowości. Warto jest też dokonać oceny, jak parametry  $p$  (skuteczność transmisji) i  $\alpha$  (skuteczność latek) wpływają na czas trwania incydentu.

### 4.3 Model SIS — Sieć z tymczasowymi zabezpieczeniami

Rozważmy sieć z mechanizmami tymczasowego oczyszczania, takimi jak regularne restarty systemów, okresowe czyszczenie pamięci podręcznej, lub rotacyjne odświeżanie obrazów systemowych. W takim środowisku zainfekowany komputer może wrócić do stanu podatnego z prawdopodobieństwem  $\alpha$  w każdym cyklu. Model ten jest szczególnie odpowiedni dla serwerów z krótkotrwałymi sesjami czy z automatycznymi resetami. W modelu SIS komputer może być zarażany wielokrotnie, a wirus może krażyć w sieci przez długi czas, powodując powtarzające się fale infekcji. W przeciwieństwie do modeli SI i SIR, tutaj nie ma stanu końcowego — sieć może oscylować między różnymi stanami zarażenia.

Podobnie jak w modelu SIR, nie każdy komputer musi zostać kiedykolwiek zainfekowany — wirus może wygasnąć zanim dotrze do odległych węzłów. Zachodzi  $\mathbb{P}[X_v = \infty] > 0$ . Interesuje nas  $\mathbb{E}[X_v | X_v < \infty]$  — średni czas do pierwszego zarażenia, jeśli do niego dojdzie. W kontekście praktycznym,  $X_v$  pozwala zrozumieć, które komputery są najbardziej narażone na początkową falę ataku.

W modelu SIS zmienna  $Y_t$  ma szczególne znaczenie, ponieważ liczba zainfekowanych komputerów może oscylować w czasie. Rozkład  $Y_t$  pokazuje możliwe stany sieci w różnych momentach — ile komputerów może być jednocześnie skompromitowanych. To pozwala na oszacowanie średniego obciążenia zespołu IT obsługą incydentów albo planowanie pojemności systemów monitoringu bezpieczeństwa. Warto zauważyć, że  $\mathbb{E}[Y_t] \rightarrow 0$  wraz z  $t \rightarrow \infty$  — średnio liczba zarażonych komputerów dąży do zera, co oznacza, że wirus ostatecznie wygasza. Jednak rozkład  $Y_t$  dla skończonych  $t$  może wykazywać znaczną wariancję.

Zmienna  $Z$  określa moment, w którym wirus całkowicie znika z sieci — wszystkie komputery są w stanie podatnym i żaden nie jest zainfekowany. W przeciwieństwie do modelu SIR, gdzie epidemia wygasza z braku podatnych celów, w modelu SIS wirus wygasza losowo z powodu równoczesnego „wyleczenia” wszystkich zainfekowanych komputerów. Rozkład  $Z$  i wartość  $\mathbb{E}[Z]$  informują o średnim czasie, przez który wirus będzie krażył w sieci czy skuteczności strategii „poczekaj aż wirus sam wygaśnie” w porównaniu z aktywną interwencją.



# Rozdział 5

## Analiza modelu SI

### 5.1 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,  $X_v \sim \text{Geo}(p)$ , co za tym idzie  $\mathbb{E}[X_v] = \frac{1}{p}$ .

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  $\mathbb{E}[Y_t] = 2 - q^t$ .

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

### 5.2 Trójkąt

Przyjrzyjmy się teraz nieco większemu grafowi — trójkątowi. Niech jeden z wierzchołków będzie źródłem  $s$ , a pozostałe oznaczmy  $u, v$ . Zmienna  $X_v$  ma taki sam rozkład jak  $X_u$ . 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

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

Policzmy teraz wartość oczekiwana.

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

Jeśli chodzi o liczbę zainfekowanych po  $t$  krokach, to 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  $p^2q^{2 \cdot (t-1)}$ . 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:

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

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 znaczaco komplikuje sytuację, jeśli chodzi o model SI.

### 5.3 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 propagacje jako łańcuch Markova. W danej chwili stanem jest  $\mathcal{I}_t$ . Stan absorbujący wynosi  $V$ . Jeśli  $\ell = \max\{d(u, v) : u \in \mathcal{I}_t, v \in \mathcal{S}_t\}$  to po  $\ell$  rundach możemy przejść do stanu całkowitej infekcji grafu z dodatnim prawdopodobieństwem.

Zmienna  $Z$  to dokładnie czas, w którym propagacja osiągnie ten stan absorbujący. A zatem z własności łańcuchów Markova 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

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

□

## 5.4 Grafy ścieżkowe

Jako pierwszą rodzinę grafów rozważmy grafy ścieżkowe  $P_n$ . Założmy, 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:

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

gdzie  $G_2, G_3, \dots, G_n \sim \text{Geo}(p)$  oraz  $G_2, G_3, \dots, G_n$  są niezależne. Widzimy zatem, że  $X_v = G_2 + G_3 + \dots + G_v$ , a więc z Faktu 6  $X_v$  ma rozkład ujemny dwumianowy:

$$X_v \sim \text{NegBin}(v-1, p), \quad \mathbb{E}[X_v] = \frac{v-1}{p}.$$

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  $Y_t$ :

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

Jeśli  $n \geq 1 + t$  to wówczas

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

Czas całkowitego zainfekowania grafu to  $Z = \max\{X_1, \dots, X_n\} = X_n$ . Zatem rozkład zmiennej  $Z$  oraz jej wartość oczekiwana są już nam znane i wynoszą

$$Z \sim \text{NegBin}(n-1, p), \quad \mathbb{E}[Z] = \frac{n-1}{p}.$$

## 5.5 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  $X_1, \dots, X_n$  są niezależne. Stąd dla każdego  $v \in \{1, \dots, n\}$  mamy

$$X_v \sim \text{Geo}(p), \quad \mathbb{E}[X_v] = \frac{1}{p}.$$

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 zmenna  $X_v$  osiągnie sukces w czasie co najwyżej  $t$ . Mamy  $\mathbb{P}[X_v \leq t] = 1 - q^t$  a więc prawdopodobieństwo tej próby wynosi  $1 - q^t$ . Niech  $B_t \sim \text{Bin}(n, 1 - q^t)$ . Wtedy

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

Przejdzmy teraz do zmiennej  $Z$ . Mamy  $Z = \max\{X_1, \dots, X_n\}$ . Ponieważ zmienne te są IID, z [Faktem 1](#) otrzymujemy  $\mathbb{P}[Z \leq t] = (1 - q^t)^n$ . Policzmy wartość oczekiwana całkowitego zainfekowania grafu:

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

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

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

to z [Nierównością 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:

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

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:

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

## 5.6 Uniwersalne 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.

**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  $Z'$  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 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  $\lambda = \log\left(\frac{1}{1-p}\right)$  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  $A_j = \{v \in V : d(s, v) = j\}$  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  $T_h = U_1 + \dots + U_h$  oraz  $Z = T_h$  a więc

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

gdzie w linijce pierwszej wykorzystujemy Lemat 1, w drugiej Nierówność 2 a w ostatniej 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:

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

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.

## 5.7 Grafy cykliczne

Przejdźmy teraz do grafów cyklicznych. W rozważaniach dla trójkąta, to jest  $C_3$ , mogliśmy zauważać, 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  $j \in \{1, 2, \dots, n-1\}$  położmy  $N_j \sim \text{NegBin}(j, p)$ . Zmienne te są niezależne. Mamy wtedy

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

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  $X_v \approx \min\{A, B\}$  oraz

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

Położmy  $C = A - B$ . Korzystając z [Faktu 7](#) 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  $z = \frac{x-\mu}{\sigma}$ . Wtedy  $x = \mu + \sigma z$ ,  $dx = \sigma dz$ . Dla  $x > 0$  mamy  $z > -c$  zaś dla  $x < 0$  mamy  $z < -c$ . Otrzymujemy więc

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

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 [Lemat 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  $\varphi\left(\frac{\eta}{\xi}\right) \rightarrow 0$  oraz  $\Phi\left(\frac{\eta}{\xi}\right) \rightarrow 0$  a więc  $\mathbb{E}[X_v] \rightarrow \mu_a = \frac{a}{p}$ . 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 dla dużych  $n$  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

$$\mathbb{E}[Y_t] = \sum_{j=0}^{2t} \min\{n, 1+j\} \cdot \binom{2t}{j} p^j q^{2t-j},$$

a dla  $n \geq 1 + 2t$  mamy

$$\mathbb{E}[Y_t] = 1 + 2tp.$$

Teraz możemy wyznaczyć rozkład i wartość oczekiwana zmiennej  $Z$ . Zauważmy, że

$$\mathbb{P}[Z > t] = \mathbb{P}[Y_t \neq n] = \mathbb{P}[L_t + R_t \leq n - 2] = \sum_{j=0}^{n-2} \binom{2t}{j} p^j q^{2t-j}.$$

W takim razie wartość oczekiwana wynosić będzie

$$\mathbb{E}[Z] = \sum_{t=0}^{\infty} \mathbb{P}[Z > t] = \sum_{t=0}^{\infty} \sum_{j=0}^{n-2} \binom{2t}{j} p^j q^{2t-j} = \sum_{j=0}^{n-2} p^j q^{-j} \sum_{t=0}^{\infty} \binom{2t}{j} q^{2t}.$$

Mamy

$$\begin{aligned} \sum_{t=0}^{\infty} \binom{2t}{j} q^{2t} &= \frac{1}{2} \left( \sum_{t=0}^{\infty} \binom{t}{j} q^t + \sum_{t=0}^{\infty} \binom{t}{j} (-q)^t \right) \\ &= \frac{1}{2} \left( \frac{q^j}{(1-q)^{j+1}} + \frac{(-q)^j}{(1+q)^{j+1}} \right) = \frac{q^j}{2} \cdot \left( p^{-(j+1)} + \frac{(-1)^j}{(2-p)^{j+1}} \right). \end{aligned}$$

Dalej

$$\begin{aligned} \mathbb{E}[Z] &= \sum_{j=0}^{n-2} p^j q^{-j} \frac{q^j}{2} \cdot \left( p^{-(j+1)} + \frac{(-1)^j}{(2-p)^{j+1}} \right) = \frac{1}{2} \sum_{j=0}^{n-2} \frac{1}{p} + \frac{(-p)^j}{(2-p)^{j+1}} \\ &= \frac{n-1}{2p} + \frac{1}{2(2-p)} \sum_{j=0}^{n-2} \left( \frac{-p}{2-p} \right)^j = \frac{n-1}{2p} + \frac{1}{2(2-p)} \cdot \frac{1 - \left( \frac{-p}{2-p} \right)^{n-1}}{1 - \frac{-p}{2-p}} \\ &= \frac{n-1}{2p} + \frac{1}{4} \cdot \left( 1 + (-1)^n \cdot \left( \frac{p}{2-p} \right)^{n-1} \right). \end{aligned}$$

Zauważmy, że  $\left| \frac{p}{2-p} \right| < 1$  a więc  $\mathbb{E}[Z] \sim \frac{n-1}{2p}$ , również podobnie jak dla grafów ścieżkowych. Jeśli  $n = 3$  to wynik ten pokrywa się ze wzorem uzyskanym dla grafu trójkątnego.

## 5.8 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  $B \sim \text{Bin}(n-a, 1-q^a)$ . Zatem

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

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. 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 asymptotyke  $\mathbb{E}[Y_2]$ . Oznaczamy  $Y_1 = 1 + U$  gdzie  $U \sim \text{Bin}(n-1, p)$ . Z prawa całkowej wartości oczekiwanej mamy

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

Musimy wyznaczyć  $\mathbb{E}[q^{Y_1}]$  jak i  $\mathbb{E}[Y_1 q^{Y_1}]$ . Korzystając z [Faktu 5](#) dostajemy

$$\mathbb{E}[q^{Y_1}] = \mathbb{E}[q^{1+U}] = q \cdot \mathbb{E}[q^U] = q(q + pq)^{n-1} = q^n(1 + p)^{n-1}.$$

Dla drugiej wartości mamy zaś

$$\begin{aligned} \mathbb{E}[Y_1 q^{Y_1}] &= \mathbb{E}[(1 + U)q^{1+U}] = q(\mathbb{E}[q^U] + \mathbb{E}[Uq^U]) \\ &= q(q^{n-1}(1 + p)^{n-1} + (n-1)pq^{n-1}(1 + p)^{n-2}) \\ &= q^n(1 + p)^{n-2}(1 + p + (n-1)p) = q^n(1 + p)^{n-2}(1 + np). \end{aligned}$$

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

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

Położymy  $\varepsilon_n = (n-1)(1 + p)^{-2}(1 - p^2)^n$ . 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 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  $\mathbb{E}[(Z')^2] = \frac{(n-1)^2 + (n-1)q}{p^2}$ . Mamy zatem

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

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.

## 5.9 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 wierzchołek  $v$ . 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

$$X_v \sim \text{NegBin}(d_v, p), \quad \mathbb{E}[X_v] = \frac{d_v}{p}.$$

**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:

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

□

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

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

Położmy  $a_j = |\{v \in V : d_v = j\}|$  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  $F(t; j, p) = 0$ , 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  $L$  będzie zbiorrem liści w  $G$ . Wtedy mamy  $Z = \max_{u \in L} X_u$ . Z nierówności Jensaena (7) otrzymujemy

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

Zauważmy, że  $\epsilon(s) = \max_{u \in L} d_u$  i jest to wysokość drzewa. Oznaczmy ją przez  $h$ . 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 drzew o  $n$  wierzchołkach i wysokości  $h$  osiągających zarówno dolne jak i górne ograniczenie na  $\mathbb{E}[Z]$ . Dla dolnej nierówności możemy взять 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. Położmy  $a_1 = b$ . 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 drzewa.

## 5.10 Eksperymenty

Przeprowadzimy teraz serie symulacji w celu sprawdzenia, czy rzeczywiście wyniki teoretyczne pokrywają się z praktycznymi. Dla wszystkich symulacji ustalmy  $p = 0.2$ . Rozkłady zmiennej  $Y_t$  oraz  $Z$  będziemy ustalać dla grafów mających 100 wierzchołków.

Dla  $Y_t$  będziemy wybierać  $t$  zależne od konkretnej rodziny. Wartość oczekiwana  $Z$  będziemy wyznaczać dla liczby wierzchołków należących do  $\{1, 2, \dots, 1000\}$ . Korzystamy z [Algorytm 1](#)

Na początku oczywiście analizujemy grafy ścieżkowe. Dla tej rodziny zmiennej  $Y_t$  zachowuje się inaczej jeśli  $t < n$  a inaczej gdy  $t \geq n$ . Przewidywany czas całkowitej infekcji to  $\frac{n}{p}$ . Wyznaczmy zatem rozkład dla wartości  $t = r \cdot \frac{n}{p}$  dla  $r \in \{0.1, 0.5, 0.8, 0.9\}$ . Widzimy, że rozkłady są coraz bardziej skoncentrowane na lewo, zgodnie z przewidywaniami (patrz [Wykres 5.1](#)). Nie będziemy już wyznaczać wartości oczekiwanej  $Y_t$ . Rozkład  $Z$  jest rozkładem ujemnym dwumianowym i jest dość dobrze znany. Nie będziemy zatem go sprawdzać.

Teraz przyjrzyjmy się grafom  $S_n$ . Dla rozkładu  $Y_t$  ustalmy  $t = \lfloor \frac{\log(n)}{2p} \rfloor$ . Widzimy, że wyniki niemal idealnie pokrywają się z oczekiwaniami teoretycznymi dla grafów gwiazdowych. (patrz [Wykres 5.2](#)). Również wartości  $\mathbb{E}[Z]$  są zbliżone do teoretycznej asymptotyki ([Wykres 5.3](#)).

Następnie popatrzmy na grafy cykliczne. Skoro zachowują się one podobnie do grafów ścieżkowych to zasymulujemy jedynie wartości oczekiwane zmiennej  $Y_t$  dla  $t = 0.4 \frac{n}{p}$  oraz  $Z$ . Wyniki te przedstawiono na [Wykresie 5.4](#).

Na sam koniec popatrzmy na grafy pełnie. Wartość  $\mathbb{E}[Z]$  powinna wraz z rozmiarem grafu zbiegać do liczby 2. Aby się upewnić, że na pewno tak jest sprawdźmy to dla  $p \in \{0.2, 0.1\}$ . Jak widać na [Wykresie 5.5](#) rzeczywiście dla większych  $n$  czas całkowitego poinformowania grafu zbiega do 2.

---

### Algorytm 1 Propagacja SI

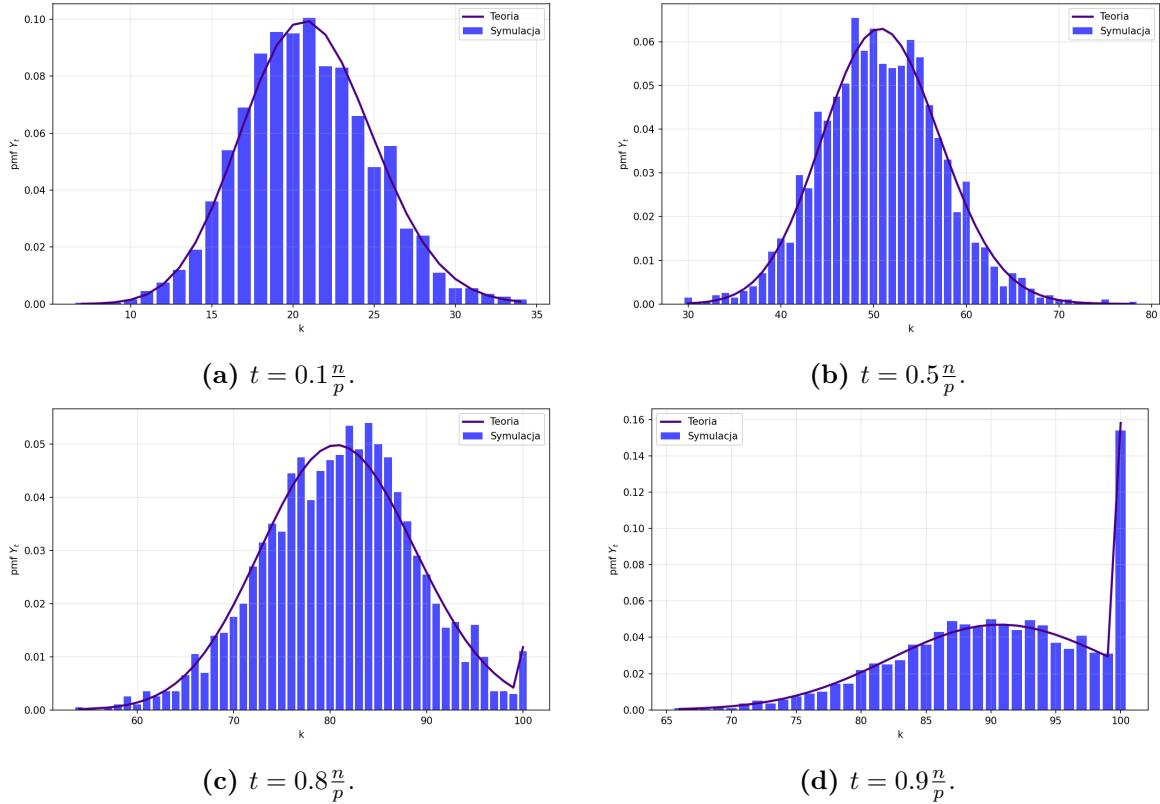
---

```

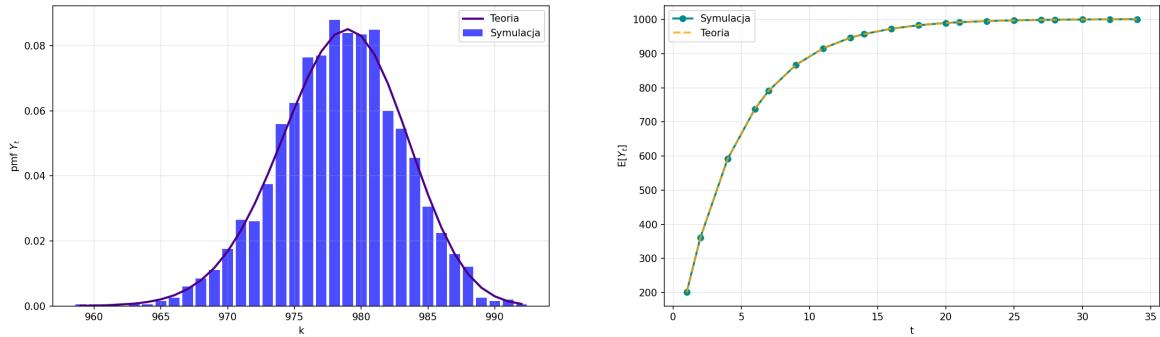
1: Input: Graf  $G = (V, E)$ , prawdopodobieństwo infekcji  $p$ , źródło  $s \in V$ 
2: Output: Zbiór zarażonych wierzchołków ( $\mathcal{I}_t$ ), czas trwania propagacji  $Z$ 
3:  $\mathcal{I}_0 \leftarrow \{s\}$ 
4:  $t \leftarrow 0$ 
5: while  $\mathcal{I}_t \neq V$  do
6:    $\mathcal{I}' \leftarrow \emptyset$ 
7:   for each  $u \in \mathcal{I}_t$  do
8:     for each  $v \in N(u)$  do
9:       if  $v \notin \mathcal{I}_t$  and  $\text{random()} < p$  then
10:         $\mathcal{I}' \leftarrow \mathcal{I}' \cup \{v\}$ 
11:      end if
12:    end for
13:  end for
14:   $\mathcal{I}_{t+1} \leftarrow \mathcal{I}_t \cup \mathcal{I}'$ 
15:   $t \leftarrow t + 1$ 
16: end while
17:  $Z \leftarrow t$ 
18: return ( $\mathcal{I}_t$ ),  $Z$ 

```

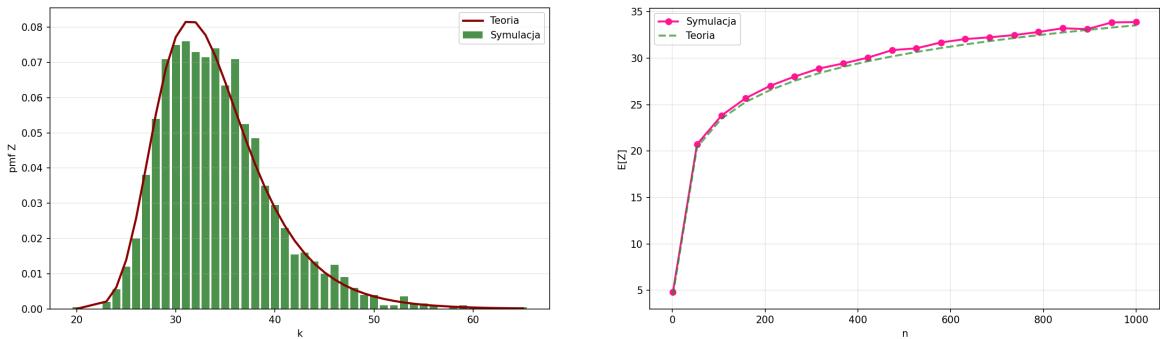
---



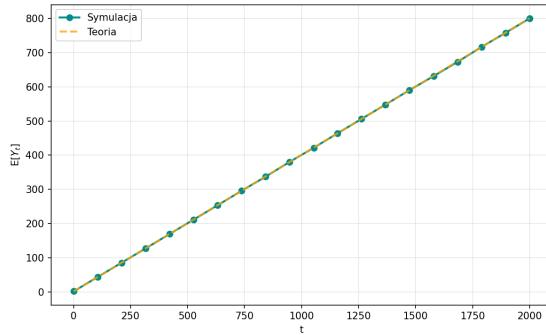
**Rysunek 5.1:** Rozkład zmiennej  $Y_t$  dla  $P_n$ .



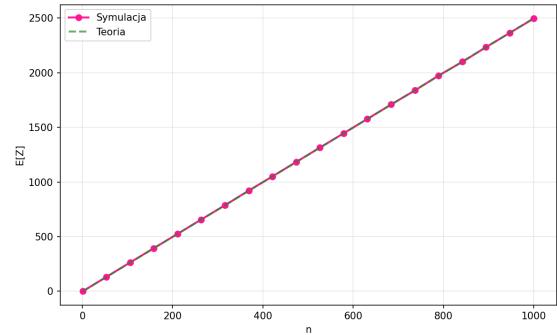
**Rysunek 5.2:** Rozkład i wartość oczekiwana zmiennej  $Y_t$  dla  $S_n$ .



**Rysunek 5.3:** Rozkład i wartość oczekiwana zmiennej  $Z$  dla  $S_n$ .

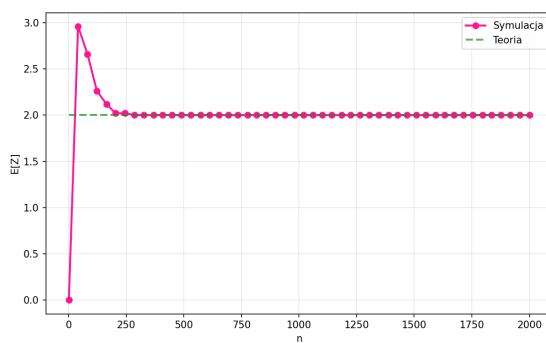


(a)  $\mathbb{E}[Y_t]$ .

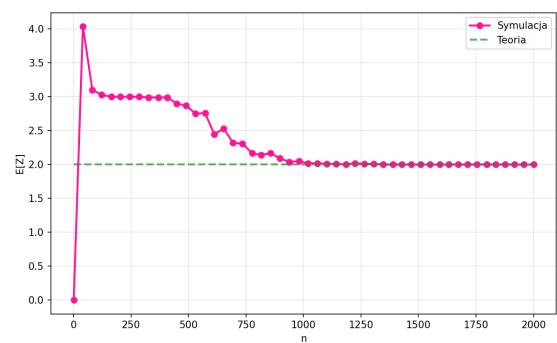


(b)  $\mathbb{E}[Z]$ .

**Rysunek 5.4:** Wartości oczekiwane zmiennych  $Y_t$  oraz  $Z$  dla  $C_n$ .



(a)  $p = 0.2$ .



(b)  $p = 0.1$ .

**Rysunek 5.5:** Wartości oczekiwane  $Z$  dla  $K_n$ .



# Rozdział 6

## Rozkłady umierające

W celu ułatwienia analizy modeli SIR oraz SIS wprowadzimy nowe rozkłady prawdopodobieństwa uwzględniające możliwość przerwania propagacji. Zmienne losowe mające te rozkłady będą przyjmować wartość nieskończoność w sytuacji, w której nastąpi przerwanie przed pożdanym 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  $\mathbb{E}[X|X < \infty]$ .

### 6.1 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 przerwaniem eksperymentu. Jeśli natomiast proces zostanie zabity szybciej niż pierwszy sukces to wtedy  $X = \infty$ . Oznaczamy  $X \sim KGeo(p, \alpha)$ . 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

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

oraz

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

Wartość oczekiwana wynosi nieskończoność. Natomiast

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

## 6.2 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  $X = \infty$ . Oznaczamy  $X \sim \text{KNegBin}(m, p, \alpha)$ . Dla wygody oznaczmy  $q = 1 - p$ ,  $\beta = 1 - \alpha$ . Alternatywnie istnieją niezależne zmienne  $Y_1, \dots, Y_m \sim \text{KGeo}(p, \alpha)$  takie, że  $X = Y_1 + \dots + Y_m$ . Aby wyznaczyć rozkład  $X$  ustalmy  $k \geq m$  i niech  $K = \{\mathbf{y} \in \mathbb{N}_+^m : y_1 + \dots + y_m = k\}$ . Wtedy

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

Łatwym kombinatorycznym argumentem można pokazać, że  $|K| = \binom{k-1}{m-1}$ . 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

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

# Rozdział 7

## Analiza modelu SIR

### 7.1 Dwa wierzchołki, jedna krawędź

W celu oswojenia się z bardziej skomplikowanym modelem, jakim jest SIR, przeanalizujmy na początku 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:

$$X_v \sim \text{KGeo}(p, \alpha), \quad \mathbb{E}[X_v | X_v < \infty] = \frac{1}{1 - q\beta}.$$

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

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

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. Sytuacja ta zachodzi z prawdopodobieństwem  $1 - q\beta$ . Czekamy na pierwszą taką chwilę, a więc

$$Z \sim \text{Geo}(1 - q\beta), \quad \mathbb{E}[Z] = \frac{1}{1 - q\beta}.$$

## 7.2 Pewne wygaśnięcie

W modelu SIR wygaszenie 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  $\mathbb{P}[X_v = \infty] \geq \alpha q^{i_s} > 0$ . Jeśli rozważamy propagacje jako łańcuch Markova to prawdopodobieństwo przejścia do stanu absorbowującego, 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 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  $W$ .  $\square$

## 7.3 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  $U_2, U_3, \dots, U_n \sim \text{KGeo}(p, \alpha)$ . Zmienne te są niezależne. Mamy  $X_v = U_2 + U_3 + \dots + U_v$  a zatem  $X_v$  ma rozkład umierający ujemny dwumianowy:

$$X_v \sim \text{KNegBin}(v-1, p, \alpha), \quad \mathbb{E}[X_v | X_v < \infty] = \frac{v-1}{1-q\beta}.$$

Jeśli chodzi o zmienną  $Y_t$  to zauważmy, że

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

A więc dla  $k \in \{1, \dots, n\}$  mamy

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

Dalej

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

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  $\theta = \frac{p}{1-q\beta}$ . Wtedy dla  $k \in \{1, \dots, n-1\}$  mamy

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

Ponadto

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

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. Zauważmy, że jeśli  $W < n$  to  $Z = T_1 + \dots + T_W$  a gdy zaś  $W = n$  to  $Z = T_1 + \dots + T_{n-1}$ . Zatem  $Z = T_1 + T_2 + \dots + T_Q$  gdzie  $Q = \min\{W, n-1\}$ . Będziemy potrzebować rozkładu i wartości oczekiwanej zmiennej  $Q$ . Możemy zapisać  $Q = W - \mathbf{1}_{W=n}$  a wtedy

$$\mathbb{E}[Q] = \mathbb{E}[W] - \mathbb{P}[W = n] = \frac{1-\theta^n}{1-\theta} - \theta^{n-1} = \frac{1-\theta^{n-1}}{1-\theta}.$$

Dalej  $\mathbb{P}[Q = k] = \mathbb{P}[W = k] = (1-\theta)\theta^{k-1}$  dla  $k \in \{1, \dots, n-2\}$  oraz  $\mathbb{P}[Q = n-1] = \mathbb{P}[W \geq n-1] = \theta^{n-2}$ . Zachodzi zatem  $\mathbb{P}[Z = t | Q = m] = \mathbb{P}[T_1 + \dots + T_m = t]$  ale  $T_1 + \dots + T_m \sim \text{NegBin}(m, 1-q\beta)$  ([Fakt 6](#)). Daje nam to  $\mathbb{P}[Z = t | Q = m] = \binom{t-1}{m-1} (1-q\beta)^m (q\beta)^{t-m}$ . Z prawdopodobieństwa całkowitego dostajemy

$$\mathbb{P}[Z = t] = \sum_{m=1}^{n-1} \binom{t-1}{m-1} (1-q\beta)^m (q\beta)^{t-m} \cdot \mathbb{P}[Q = m].$$

Natomiast wartość oczekiwana zmiennej  $Z$  wynosi

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

## 7.4 Grafy gwiazdne

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

$$X_v \sim \text{KGeo}(p, \alpha), \quad \mathbb{E}[X_v | X_v < \infty] = \frac{1}{1-q\beta}.$$

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

Przyjrzyjmy się teraz zmiennej  $Y_t$ . Położmy  $C = \min\{\tau \in \mathbb{N} : 0 \in \mathcal{R}_\tau\}$ . 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 gwiazdowych. Przypomnijmy, że rozkład ten wynosi  $1 + B_{j,t}$  gdzie  $B_{j,t} \sim \text{Bin}(n, 1 - q^{\min\{j,t\}})$ . A więc

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

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

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

Propagacja zakończy się 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 zwiezlej 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

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

gdzie  $\Gamma(x)$  jest funkcją Gamma,  $B(x, y)$  jest funkcją Beta oraz  $(x)_n = x(x+1) \dots (x+n-1)$  jest symbolem Pochhammer'a. Ostatecznie otrzymujemy

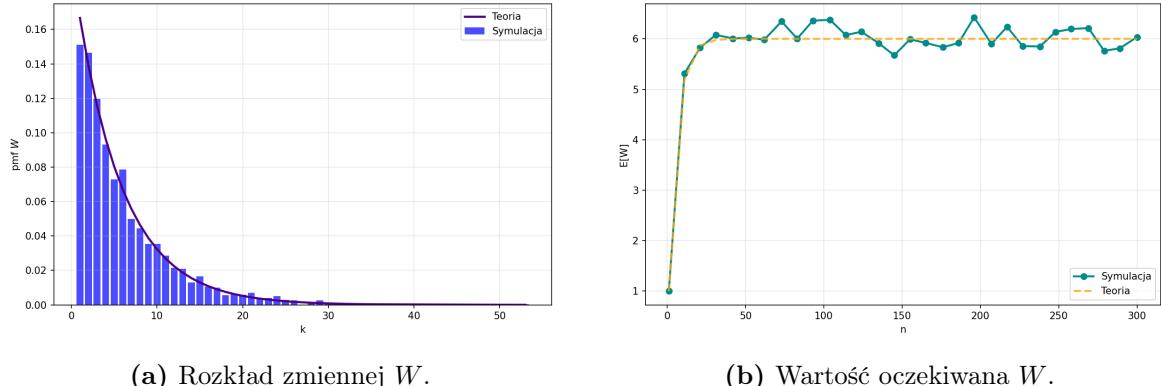
$$\mathbb{E}[Z] \approx \frac{1}{b} \cdot \left(1 - \frac{n!}{(\frac{b}{a}+1)_n}\right) = \frac{1}{\log(\beta)} \cdot \frac{n!}{\left(\frac{\log(\beta)}{\log(q)} + 1\right)_n} - \frac{1}{\log(\beta)}.$$

## 7.5 Eksperymenty

W modelu SIR symulować będziemy rozkłady zmiennych  $W, Z$  oraz ich wartości oczekiwane. Ustalmy  $p = 0.2$  oraz  $\alpha = 0.05$ . Do przeprowadzenia eksperymentu posłuży nam [Algorytm 2](#). Rozkłady wyznaczymy na grafach o  $n = 100$  wierzchołkach natomiast wartości oczekiwane dla  $n \in \{1, 2, \dots, 300\}$ .

Jeśli chodzi o rodzinę  $P_n$  to rozkład  $W$  pokrywa się idealnie z przewidywaniami. Widzimy nawet kumulację masy w ogonie. Wartości oczekiwane również zbiegają zgodnie z wyznaczonym wzorem. Podobnie mamy dla zmiennej losowej  $Z$ .

Dla grafów gwiazd rozkłady zarówno  $W$  jak i  $Z$  są dobrze przybliżone. Wartości oczekiwane  $W$  są niemal idealnie dopasowane. Możemy zauważać dobre przybliżenie na  $\mathbb{E}[W]$ .



(a) Rozkład zmiennej  $W$ .

(b) Wartość oczekiwana  $W$ .

**Rysunek 7.1:** Rozkład i wartość oczekiwana zmiennej  $W$  dla  $P_n$ .

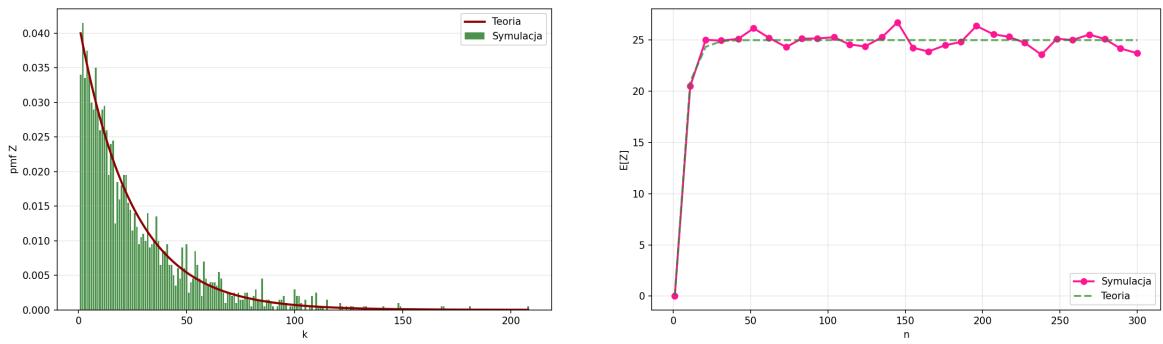
---

**Algorytm 2** Propagacja SIR

---

```
1: Input: Graf  $G = (V, E)$ , prawdopodobieństwo infekcji  $p$ , prawdopodobieństwo
   wyzdrowienia  $\alpha$ , źródło  $s \in V$ 
2: Output: Zbiór wyzdrowiałych wierzchołków ( $\mathcal{R}_t$ ), czas wymarcia infekcji  $Z$ 
3:  $\mathcal{S}_0 \leftarrow V \setminus \{s\}$ 
4:  $\mathcal{I}_0 \leftarrow \{s\}$ 
5:  $\mathcal{R}_0 \leftarrow \emptyset$ 
6:  $t \leftarrow 0$ 
7: while  $\mathcal{I}_t \neq \emptyset$  and  $\exists v \in \mathcal{I}_t : \mathcal{S}_t \cap N(v) \neq \emptyset$  do
8:    $\mathcal{I}' \leftarrow \emptyset$ 
9:   for each  $u \in \mathcal{I}_t$  do
10:    for each  $v \in N(u)$  do
11:      if  $v \in \mathcal{S}_t$  and  $\text{random}() < p$  then
12:         $\mathcal{I}' \leftarrow \mathcal{I}' \cup \{v\}$ 
13:      end if
14:    end for
15:   end for
16:    $\mathcal{R}' \leftarrow \emptyset$ 
17:   for each  $u \in \mathcal{I}_t$  do
18:     if  $\text{random}() < \alpha$  then
19:        $\mathcal{R}' \leftarrow \mathcal{R}' \cup \{u\}$ 
20:     end if
21:   end for
22:    $\mathcal{S}_{t+1} \leftarrow \mathcal{S}_t \setminus \mathcal{I}'$ 
23:    $\mathcal{I}_{t+1} \leftarrow (\mathcal{I}_t \cup \mathcal{I}') \setminus \mathcal{R}'$ 
24:    $\mathcal{R}_{t+1} \leftarrow \mathcal{R}_t \cup \mathcal{R}'$ 
25:    $t \leftarrow t + 1$ 
26: end while
27:  $\mathcal{R}_t \leftarrow \mathcal{R}_t \cup \mathcal{I}_t$ 
28:  $Z \leftarrow t$ 
29: return  $(\mathcal{R}_t), Z$ 
```

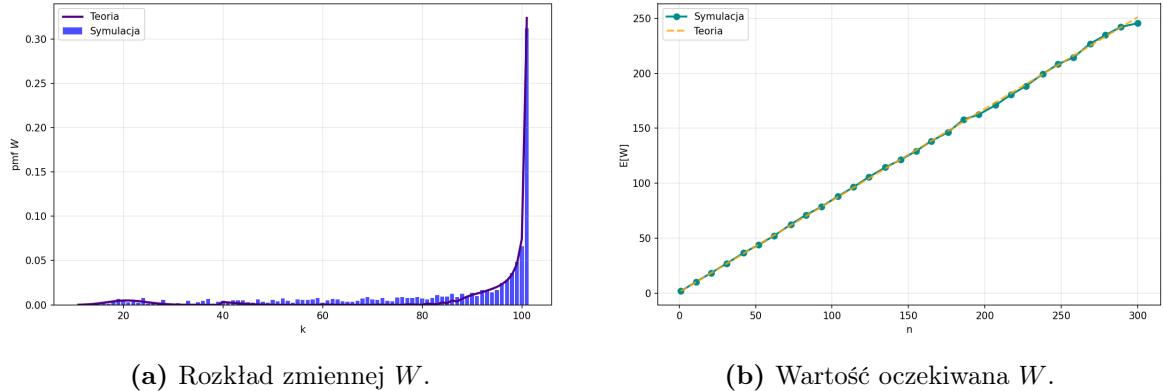
---



(a) Rozkład zmiennej  $Z$ .

(b) Wartość oczekiwana  $Z$ .

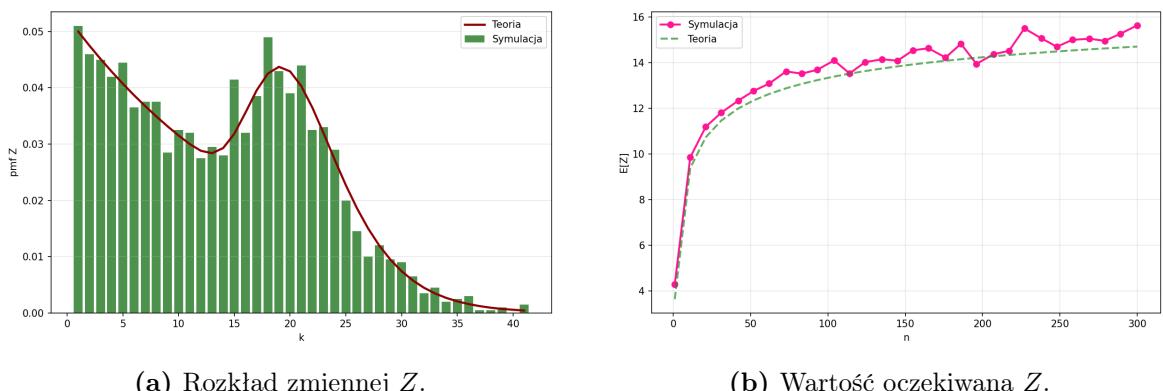
**Rysunek 7.2:** Rozkład i wartość oczekiwana zmiennej  $Z$  dla  $P_n$ .



(a) Rozkład zmiennej  $W$ .

(b) Wartość oczekiwana  $W$ .

**Rysunek 7.3:** Rozkład i wartość oczekiwana zmiennej  $W$  dla  $S_n$ .



(a) Rozkład zmiennej  $Z$ .

(b) Wartość oczekiwana  $Z$ .

**Rysunek 7.4:** Rozkład i wartość oczekiwana zmiennej  $Z$  dla  $S_n$ .



# Rozdział 8

## Analiza modelu SIS

### 8.1 Pewne wygaśnięcie

Analizę modelu SIS zacznijmy od formalnego udowodnienia pewności wygaśnięcia 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.* Tożsamości  $\mathbb{P}[X_v = \infty] > 0$  oraz  $\mathbb{P}[Z < \infty] = 1$  wynikają 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

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

□

### 8.2 Dwa wierzchołki, jedna krawędź

Standardowo na pierwszy przykład grafu rozważamy graf o wierzchołkach  $V = \{u, v\}$  i jednej krawędzi między nimi. Za wierzchołek początkowy przyjmijmy  $u$ . 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

$$X_v \sim \text{KGeo}(p, \alpha), \quad \mathbb{E}[X_v | X_v < \infty] = \frac{1}{1 - q\beta}.$$

Następnie spróbujmy wyznaczyć rozkład  $Y_t$ . Zauważmy, że  $Y_t \in \{0, 1, 2\}$ . Prawdo-

podobieństwa 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

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

Przyjrzymy 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:

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

### 8.3 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. 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 4](#) 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  $\mathcal{O}(n^3)$ . Wartość oczekiwana zmiennej  $Y_t$  możemy obliczyć ze wzoru

$$\mathbb{E}[Y_t] = \sum_{k=0}^n k \cdot \mathbf{y}_t^{(k)}.$$

Macierz  $\mathbf{P}$  jest kolumnowo stochastyczna a więc

$$\lim_{t \rightarrow \infty} \mathbf{P}^t = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix},$$

a co za tym idzie

$$\lim_{t \rightarrow \infty} \mathbf{y}_t = [1 \ 0 \ 0 \ \cdots \ 0]^\top.$$

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

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

## 8.4 Eksperymenty

Dla modelu SIS przeprowadzimy symulacje na grafie  $K_n$ . Na nasze potrzeby wystarczy nam  $n = 10$ . Aby mieć pewność, że infekcji umrze w sensownym czasie będziemy badać propagacje dla parametrów  $p = 0.1$ ,  $\alpha = 0.9$  ([Wykres 8.1](#)),  $p = 0.2$ ,  $\alpha = 0.8$  ([Wykres 8.2](#)) oraz  $p = 0.5$ ,  $\alpha = 0.9$  ([Wykres 8.3](#)). Położmy  $\tau = \max\{20, \mathbb{E}[Z]\}$ . Dalej wyznaczmy rozkład  $Y_t$  dla  $t \in \{0.2\tau, 0.4\tau, 0.6\tau, 0.8\tau\}$ . Wartość oczekiwana obliczmy dla  $t \in \{1, 2, \dots, \tau\}$  oraz  $t \in \{1, 2, \dots, 10\tau\}$ . Przeprowadzimy teraz serie symulacji w

celu sprawdzenia, czy rzeczywiście wyniki teoretyczne pokrywają się z praktycznymi. Symulacje dokonujemy korzystając z [Algorytmu 3](#).

Dla parametrów  $p = 0.1$ ,  $\alpha = 0.9$  widzimy, że infekcja dość szybko wymiera, większość masy prawdopodobieństwa  $Y_t$  jest w 0. Podobnie zachowuje się rozkład  $Z$ . Dalej dla  $p = 0.2$ ,  $\alpha = 0.8$  nadal spora część rozkładu jest w 0 ale spora część zbiera się w kształt rozkładu normalnego dla dodatnich wartości. Podobnie ma się sytuacja dla  $p = 0.5$ ,  $\alpha = 0.9$ . Warto również zauważyc, że w tym przypadku  $\mathbb{E}[Z]$  jest gigantycznie duża, rzędu kilkudziesięciu tysięcy.

---

### Algorytm 3 Propagacja SIS

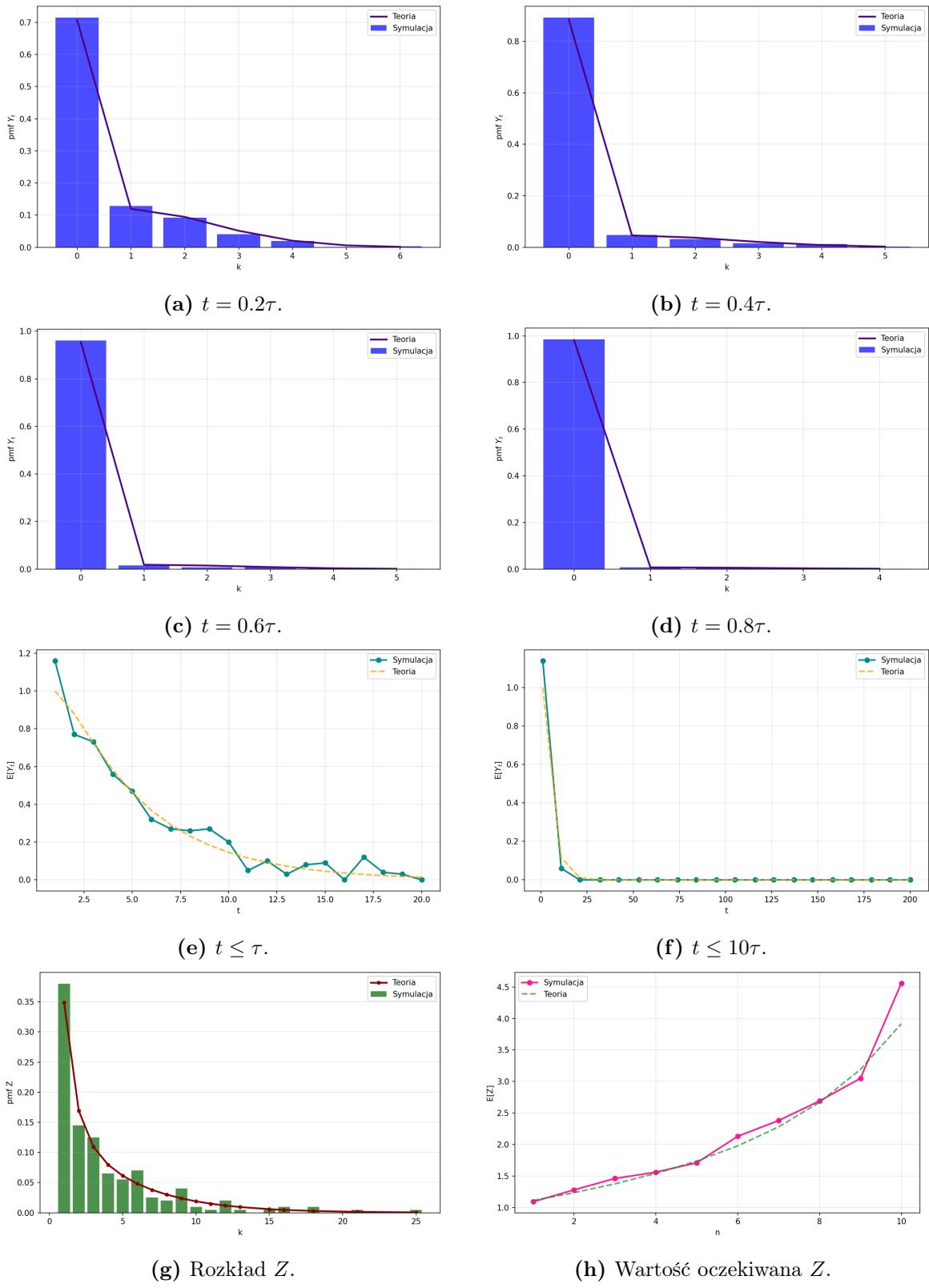
---

```

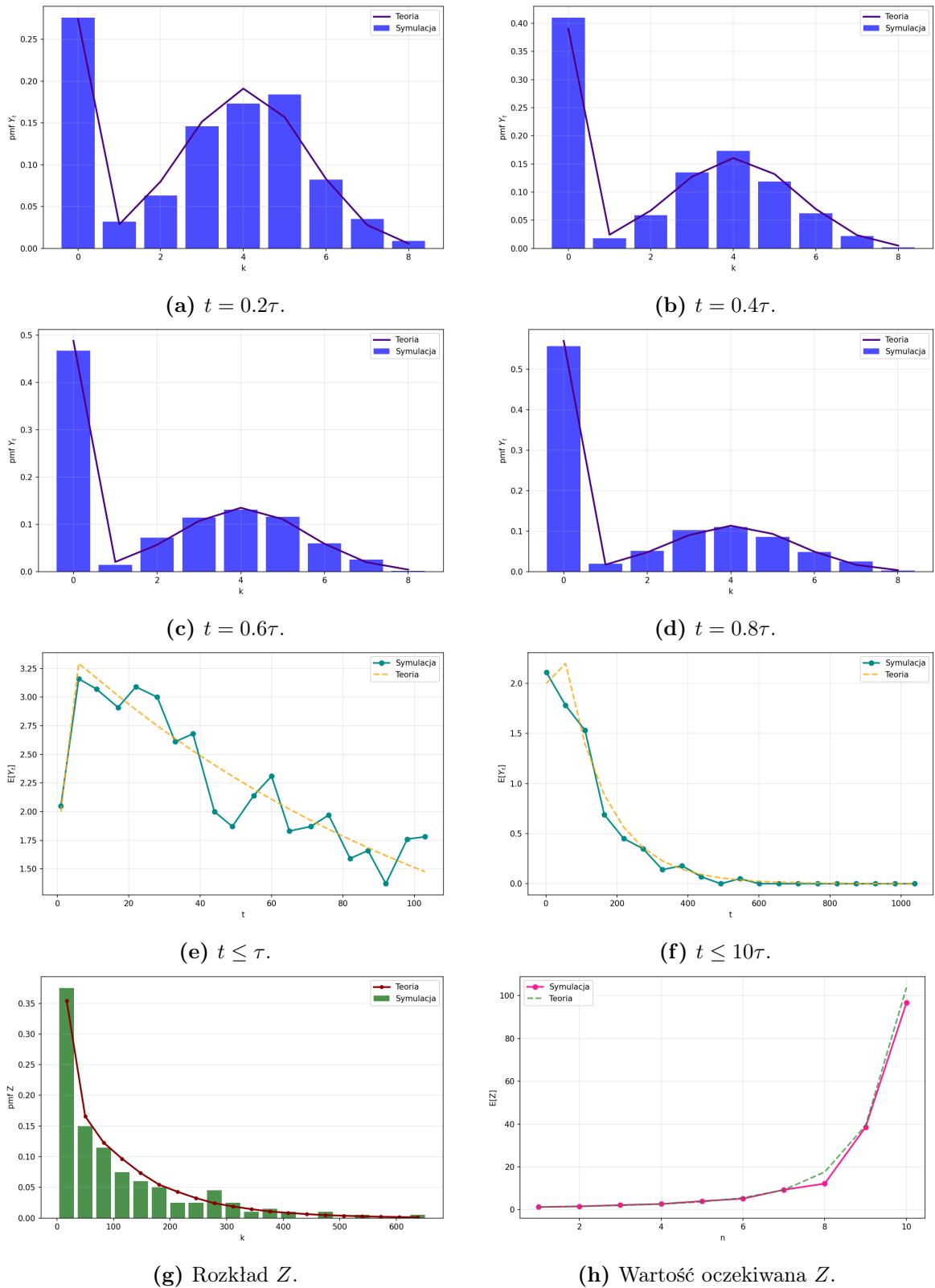
1: Input: Graf  $G = (V, E)$ , prawdopodobieństwo infekcji  $p$ , prawdopodobieństwo
   wyzdrowienia  $\alpha$ , źródło  $s \in V$ 
2: Output: Zbiór zarażonych wierzchołków ( $\mathcal{I}_t$ ), czas wymarcia infekcji  $Z$ 
3:  $\mathcal{I}_0 \leftarrow \{s\}$ 
4:  $t \leftarrow 0$ 
5: while  $\mathcal{I}_t \neq \emptyset$  do
6:    $\mathcal{I}' \leftarrow \emptyset$ 
7:   for each  $u \in \mathcal{I}_t$  do
8:     for each  $v \in N(u)$  do
9:       if  $v \notin \mathcal{I}_t$  and  $\text{random}() < p$  then
10:         $\mathcal{I}' \leftarrow \mathcal{I}' \cup \{v\}$ 
11:       end if
12:     end for
13:   end for
14:    $R \leftarrow \emptyset$ 
15:   for each  $u \in \mathcal{I}_t$  do
16:     if  $\text{random}() < \alpha$  then
17:        $R \leftarrow R \cup \{u\}$ 
18:     end if
19:   end for
20:    $\mathcal{I}_{t+1} \leftarrow (\mathcal{I}_t \cup \mathcal{I}') \setminus R$ 
21:    $t \leftarrow t + 1$ 
22: end while
23:  $Z \leftarrow t$ 
24: return  $(\mathcal{I}_t)$ ,  $Z$ 

```

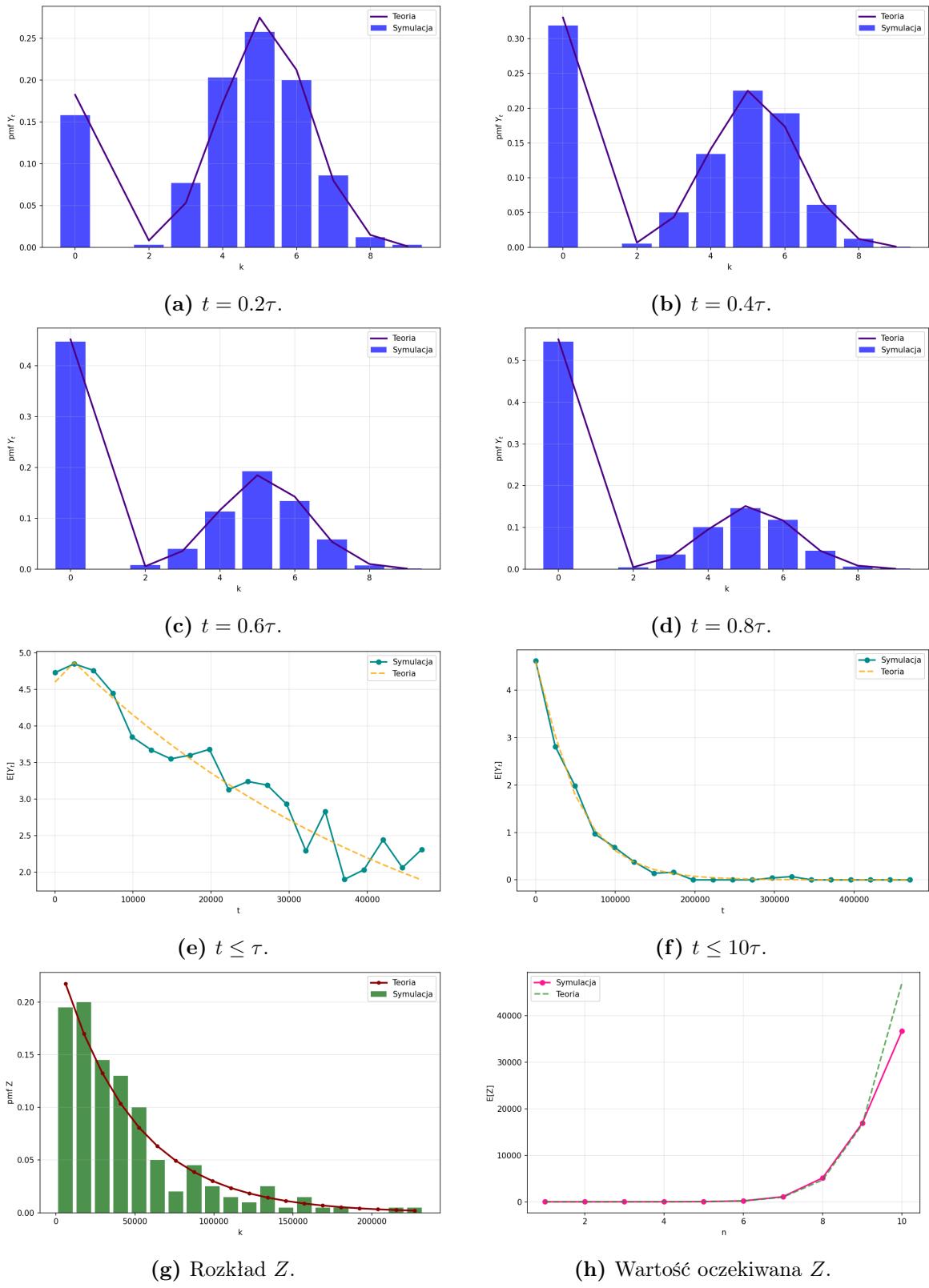
---



**Rysunek 8.1:** Rozkłady dla  $p = 0.1$ ,  $\alpha = 0.9$ .



**Rysunek 8.2:** Rozkłady dla  $p = 0.2$ ,  $\alpha = 0.8$ .



**Rysunek 8.3:** Rozkłady dla  $p = 0.5$ ,  $\alpha = 0.9$ .

# Rozdział 9

## Podsumowanie

### 9.1 Wnioski

W niniejszej pracy przeprowadziliśmy kompleksową analizę trzech klasycznych modeli propagacji w grafach: SI, SIR oraz SIS. Wyznaczenie rozkładów prawdopodobieństwa i wartości oczekiwanych kluczowych zmiennych losowych ( $X_v, Y_t, Z, W$ ) dla różnych topologii grafów pozwoliło na głębokie zrozumienie wpływu struktury grafu na dynamikę propagacji.

**Model SI.** Dla modelu SI udowodniliśmy, że całkowita infekcja grafu jest zdarzeniem pewnym oraz wyznaczyliśmy dokładne rozkłady dla wszystkich badanych rodzin grafów. Kluczowe wyniki obejmują wartości  $\mathbb{E}[Z]$ :

- Dla grafów ścieżkowych  $P_n$ :  $\mathbb{E}[Z] = \frac{n-1}{p}$ .
- Dla grafów gwiazdowych  $S_n$ :  $\mathbb{E}[Z] \sim \frac{H_n}{\log(\frac{1}{1-p})}$ .
- Dla grafów cyklicznych  $C_n$ :  $\mathbb{E}[Z] = \frac{n-1}{2p} + \frac{1}{4} + o(1)$ .
- Dla grafów pełnych  $K_n$ :  $\lim_{n \rightarrow \infty} \mathbb{E}[Z] = 2$ .

Wyprowadzone ogólne ograniczenie  $\mathbb{E}[Z] \leq h + \frac{h}{\lambda} \cdot (\log(\frac{n-1}{h}) + 1)$ , gdzie  $h$  jest ekscentrycznością źródła a  $\lambda = \log(\frac{1}{1-p})$ , pozwala na oszacowanie czasu propagacji w dowolnym grafie.

**Model SIR.** Wprowadzone rozkłady wymierające (KGeo, KNegBin) w celu ułatwienia analizy modelu SIR ale też SIS. Udowodniliśmy, że wygaśnięcie infekcji jest zdarzeniem pewnym oraz  $\mathbb{P}[X_v = \infty] > 0$  dla każdego wierzchołka  $v \neq s$ . Dla grafów ścieżkowych i gwiazdowych wyznaczyliśmy rozkłady zmiennych  $W$  oraz  $Z$ . Wartości oczekiwane  $Z$  wynoszą:

- Dla grafów ścieżkowych  $P_n$ :  $\mathbb{E}[Z] = \frac{1 - \theta^{n-1}}{q\alpha}$  gdzie  $\theta = \frac{p}{1 - q\beta}$ .

- Dla grafów gwiazdowych  $S_n$ :  $\mathbb{E}[Z] \approx \frac{1}{\log(\beta)} \cdot \frac{n!}{\left(\frac{\log(\beta)}{\log(q)} + 1\right)_n} - \frac{1}{\log(\beta)}$ .

a zmiennej  $W$ :

- Dla grafów ścieżkowych  $P_n$ :  $\mathbb{E}[W] = \frac{1 - \theta^n}{1 - \theta}$  gdzie  $\theta = \frac{p}{1 - q\beta}$ .
- Dla grafów gwiazdowych  $S_n$ :  $\mathbb{E}[W] = 1 + \frac{np}{1 - q\beta}$ .

**Model SIS.** Analiza modelu SIS dla grafu pełnego  $K_n$  ujawniła najbardziej złożoną dynamikę spośród trzech modeli. Wykorzystanie formalizmu łańcuchów Markova pozwoliło na wyznaczenie rozkładów  $Y_t$  i  $Z$  poprzez macierze przejść. Udowodniliśmy, że wyjaśnienie infekcji jest zdarzeniem pewnym mimo możliwości niezależnie od wartości parametrów  $p$  i  $\alpha$ . Symulacje pokazały, że rozkłady  $Y_t$  wykazują zarówno szybkie wygasanie jak i długotrwałą oscylację.

**Weryfikacja eksperymentalna.** Przeprowadzone symulacje komputerowe potwierdziły teoretyczne przewidywania dla wszystkich badanych modeli i topologii. Rozkłady empiryczne zmiennych losowych pokrywały się z teoretycznymi, a wartości oczekiwane zbiegały do wyznaczonych wzorów asymptotycznych. Szczególnie istotne było potwierdzenie, że dla grafu pełnego w modelu SI wartość  $\mathbb{E}[Z]$  rzeczywiście zbiega do 2 wraz ze wzrostem  $n$ .

**Scenariusz aplikacyjny.** Przedstawiony scenariusz rozprzestrzeniania się wirusa komputerowego ilustruje praktyczną użyteczność modeli:

- Model SI odpowiada infrastrukturze bez zabezpieczeń.
- Model SIR modeluje nowoczesne środowiska z automatycznymi łączkami bezpieczeństwa.
- Model SIS opisuje systemy z tymczasowymi mechanizmami oczyszczania.

## 9.2 Kierunki dalszych badań

Przeprowadzona analiza otwiera kilka interesujących kierunków kontynuacji badań:

**Rozszerzenie na inne rodziny grafów.** Analiza ograniczyła się do klasycznych rodzin grafów. Naturalne rozszerzenie obejmowałyby:

- Grafy dwudzielne pełne  $K_{m,n}$ , modelujące sieci z wyraźnym podziałem na dwie grupy (np. serwery i klienci).
- Grafy kratowe (grid graphs), istotne dla modelowania rozprzestrzeniania się epidemii w przestrzeni geograficznej.
- Grafy losowe Erdős-Rényiego  $G(n, p)$ , reprezentujące realistyczne sieci społeczne i technologiczne.

**Modele z heterogenicznymi parametrami.** W niniejszej pracy zakładaliśmy jednolite prawdopodobieństwa transmisji  $p$  i wyzdrowienia  $\alpha$  dla wszystkich wierzchołków i krawędzi. Realistyczne rozszerzenie obejmowałyby:

- Prawdopodobieństwa transmisji zależne od krawędzi:  $p_e$ , modelujące różną jakość połączeń sieciowych lub intensywność kontaktów społecznych.
- Prawdopodobieństwa wyzdrowienia zależne od wierzchołka:  $\alpha_v$ , uwzględniające zróżnicowaną odporność jednostek lub systemów.
- Modele z progową transmisją, gdzie wierzchołek zarażany jest dopiero po przekroczeniu określonej liczby zarażonych sąsiadów.

**Modele z opóźnieniami czasowymi.** Obecne modele zakładają, że zarażenie i wyzdrowienie następują po jednej jednostce czasu. Bardziej realistyczne modele uwzględniałyby:

- Losowe opóźnienia zarażenia, modelujące okres inkubacji w epidemiach biologicznych lub czas propagacji w sieciach fizycznych.
- Losowe czasy wyzdrowienia o rozkładzie bardziej ogólnym niż geometryczny, np. wykładniczym lub gamma.
- Model SEIR (Susceptible–Exposed–Infected–Recovered) z jawnym stanem narażenia.

Podsumowując, modele probabilistycznej propagacji w grafach stanowią bogate pole badawcze łączące teorię grafów, rachunek prawdopodobieństwa, procesy stochastyczne oraz zastosowania praktyczne. Niniejsza praca dostarcza solidnych fundamentów teoretycznych i metodologicznych dla kontynuacji badań w tym fascynującym obszarze.



# Bibliografia

- [1] Carlos Baquero i in. „Extrema Propagation: Fast Distributed Estimation of Sums and Network Sizes”. W: *IEEE Transactions on Parallel and Distributed Systems* 23.4 (2012), s. 668–675. DOI: [10.1109/TPDS.2011.209](https://doi.org/10.1109/TPDS.2011.209).
- [2] Dominik Bojko i Jacek Cichoń. „A Note on Leader Election Algorithms. Preliminary Report”. W: *2019 IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob)*. 2019, s. 81–87. DOI: [10.1109/APWiMob48441.2019.8964152](https://doi.org/10.1109/APWiMob48441.2019.8964152).
- [3] Jacek Cichon, Jakub Lemiesz i Marcin Zawada. „On Message Complexity of Extrema Propagation Techniques”. W: lip. 2012, s. 1–13. ISBN: 978-3-642-31637-1. DOI: [10.1007/978-3-642-31638-8\\_1](https://doi.org/10.1007/978-3-642-31638-8_1).
- [4] Jacek Cichoń, Dawid Dworzański i Karol Gotfryd. „On Reliability of the Extrema Propagation Technique in Random Environment”. W: *28th International Conference on Principles of Distributed Systems (OPODIS 2024)*. Red. Silvia Bonomi i in. T. 324. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2025, 29:1–29:16. ISBN: 978-3-95977-360-7. DOI: [10.4230/LIPIcs.OPODIS.2024.29](https://doi.org/10.4230/LIPIcs.OPODIS.2024.29). URL: <https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.OPODIS.2024.29>.
- [5] Saswata Das, Mohammad Hossein Samaei i Caterina Scoglio. *SIR Epidemics in Interconnected Networks: threshold curve and phase transition*. 2024. arXiv: [2307.10048 \[math.DS\]](https://arxiv.org/abs/2307.10048). URL: <https://arxiv.org/abs/2307.10048>.
- [6] Daniel Figueiredo, Giulio Iacobelli i Seva Shneer. „The End Time of SIS Epidemics Driven by Random Walks on Edge-Transitive Graphs”. W: *Journal of Statistical Physics* 179.3 (kw. 2020), s. 651–671. ISSN: 1572-9613. DOI: [10.1007/s10955-020-02547-7](https://doi.org/10.1007/s10955-020-02547-7). URL: <http://dx.doi.org/10.1007/s10955-020-02547-7>.
- [7] S. Gómez i in. „Discrete-time Markov chain approach to contact-based disease spreading in complex networks”. W: *EPL (Europhysics Letters)* 89.3 (lut. 2010), s. 38009. ISSN: 1286-4854. DOI: [10.1209/0295-5075/89/38009](https://doi.org/10.1209/0295-5075/89/38009). URL: <http://dx.doi.org/10.1209/0295-5075/89/38009>.
- [8] Yuanfei Huang i Adrian Röllin. *The SIR epidemic on a dynamic Erdős-Rényi random graph*. 2024. arXiv: [2404.12566 \[math.PR\]](https://arxiv.org/abs/2404.12566). URL: <https://arxiv.org/abs/2404.12566>.
- [9] Romualdo Pastor-Satorras i in. „Epidemic processes in complex networks”. W: *Reviews of Modern Physics* 87.3 (sierp. 2015), s. 925–979. ISSN: 1539-0756. DOI: [10.1103/revmodphys.87.925](https://doi.org/10.1103/revmodphys.87.925). URL: <http://dx.doi.org/10.1103/RevModPhys.87.925>.

- [10] Jiaojiao Wang, Qianqian Zhang i Sanyi Tang. „Global dynamic analyzes of the discrete SIS models with application to daily reported cases”. W: *Advances in Continuous and Discrete Models* 2024 (sierp. 2024). DOI: [10.1186/s13662-024-03829-0](https://doi.org/10.1186/s13662-024-03829-0).