

Katarzyna Rybarczyk–Krzywdzińska

**Losowe grafy przecięć.
Modelowanie sieci i ich analiza.**

Rozprawa doktorska napisana
na Wydziale Matematyki i Informatyki
Uniwersytetu im. Adama Mickiewicza w Poznaniu

Promotor: prof. UAM dr hab. Jerzy Jaworski

Poznań 2009

Składam serdeczne podziękowania
panu profesorowi doktorowi habilitowanemu
Jerzemu Jaworskiemu za jego
nieocenioną pomoc i dobre rady
dotyczące nie tylko spraw naukowych
oraz współautorom moich prac
za owocną współpracę.

Dziękuję także serdecznie mojej rodzinie,
bez której wsparcia i pomocy
praca ta nie mogłaby powstać.

Część prezentowanych w rozprawie doktorskiej wyników uzyskana została podczas badań związanych z grantem Ministerstwa Nauki i Szkolnictwa Wyższego „Teoretyczne aspekty bezpieczeństwa informacji, komunikacji oraz protokołów kryptograficznych” NN206270133.

Część prezentowanych w rozprawie doktorskiej wyników uzyskana została podczas badań związanych z grantem promotorskim Ministerstwa Nauki i Szkolnictwa Wyższego „Modelowanie i analiza sieci za pomocą losowych grafów przecięć” NN206258535.

Spis treści

Wstęp	6
1. Definicje i oznaczenia	9
1.1. Oznaczenia	9
1.2. Graf losowy	9
1.3. Losowe grafy przecięć – podstawowe modele	10
1.4. Własności grafów	12
1.4.1. Stopień wierzchołka, wierzchołki izolowane	12
1.4.2. Podgrafy i kliki	12
1.4.3. Spójność, składowe spójności	12
1.4.4. Średnica	13
1.4.5. Współczynnik skupienia	13
1.5. Własności asymptotyczne	15
1.5.1. Funkcje progowe	15
2. Sieci rzeczywiste i ich teoretyczne modele	17
2.1. Sieci złożone	17
2.1.1. Sieci złożone i ich podstawowe własności	17
2.1.2. Podstawowe pytania i modele grafów	19
2.1.3. Stan wiedzy	20
2.2. Sieci sensorowe	21
2.2.1. Sieć sensorowa z losową predystrybucją kluczy	21
2.2.2. Podstawowe pytania i modele grafów.	22
2.2.3. Stan wiedzy	23
3. Wyniki pomocnicze	26
3.1. Nierówności	26
3.1.1. Nierówności Markowa i Czebyszewa	26
3.1.2. Nierówność Chernoffa	27
3.2. Metody probabilistyczne	27
3.2.1. Metoda pierwszego i drugiego momentu	27
3.2.2. Metoda momentów	27
3.2.3. Metoda Steina	28
3.3. Oszacowania na prawdopodobieństwa powstania krawędzi	29
3.4. Prawdopodobieństwo warunkowe	30
4. Równoważność modeli	32
4.1. Równoważność losowych grafów przecięć.	32
4.2. Równoważność modelu jednostajnego i dwumianowego	34
4.3. Proste wnioski dla $G^{Jed}(\mathcal{V}, m, d)$	35
4.3.1. Spójność	35
4.3.2. Przejście fazowe	36
4.3.3. Krawędzie	36
4.3.4. Równoważność $G^{Jed}(\mathcal{V}, m, d)$ i $G(n, \hat{p})$	37

5. Rozkład stopni wierzchołków	38
5.1. Twierdzenie o stopniach wierzchołków	38
5.2. Wyniki pomocnicze	39
5.3. Dowód twierdzenia o stopniach wierzchołków	42
6. Współczynnik skupienia	44
6.1. Twierdzenie o lokalnym współczynniku skupienia	44
7. Kliki	49
7.1. Funkcja progowa dla własności posiadania kliku jako podgrafu	49
7.2. Zbieżność liczby klików do rozkładu Poissona	50
7.3. Dowód twierdzenia o liczbie klików	51
7.3.1. Zastosowanie metody Steina	51
7.3.2. Pokrycia klikowe	52
7.3.3. Modyfikacje pokryć klikowych	54
7.3.4. Optymalne pokrycia klikowe	57
7.3.5. Prawdopodobieństwo istnienia klików i ich sum jako podgrafów	60
7.3.6. Dowód twierdzenia 7.1	62
8. Średnica	65
8.1. Twierdzenie o długości średnicy w $G^{Jed}(\mathcal{V}, m, d)$	65
8.2. Drzewo BFS w grafie $G^{Jed}(\mathcal{V}, m, d)$	66
8.3. Procesy gałęzkowe	69
8.4. Przybliżenie BFS procesami gałęzkowymi	77
8.5. Dowód twierdzenia o długości średnicy	82
8.5.1. Własności $G^{Jed}(\mathcal{V}, m, d)$ na progu spójności	82
8.5.2. Dowód twierdzenia 8.1(i)	85
8.5.3. Własności $G^{Jed}(\mathcal{V}, m, d)$ na progu przejścia fazowego	85
8.5.4. Dowód twierdzenia 8.1(ii)	88
9. Wierzchołki izolowane	89
9.1. Twierdzenie o liczbie wierzchołków izolowanych	89
9.2. Wyniki pomocnicze	90
9.2.1. Podzbiory wierzchołków izolowanych	90
9.2.2. Wierzchołki izolowane w $G_s(\mathcal{V}, m, \bar{d})$	93
9.3. Dowód twierdzenia o liczbie wierzchołków izolowanych	97
9.4. Wnioski z twierdzenia o liczbie wierzchołków izolowanych	100
9.4.1. Jednostajny losowy graf przecięć	100
9.4.2. Dwumianowy losowy graf przecięć	100
9.4.3. Nieskończona liczba typów	100
9.4.4. Skończona liczba typów	102
10. Spójność i składowe spójności	104
10.1. Spójność $G^{Jed}(\mathcal{V}, m, d)$	104
10.2. Przejście fazowe	106
10.2.1. Prawdopodobieństwo wymarcia procesu gałęzkowego	106
10.2.2. Przejście fazowe w $G^{Jed}(\mathcal{V}, m, d)$	107
10.2.3. Przejście fazowe w $G_s^{Jed}(\mathcal{V}, m, d)$ i $G_s^{Bin}(\mathcal{V}, m, p)$	109
11. Wnioski i problemy otwarte	114
Notacja	116
Bibliografia	117

Wstęp

Naturalnym kierunkiem badań informatyki teoretycznej jest badanie sieci przy pomocy ich teoretycznych modeli grafowych. Okazuje się, że struktura sieci składającej się ze znacznej liczby elementów może być dobrze przybliżona strukturą grafu losowego o dużej liczbie wierzchołków, a co za tym idzie, własności takich sieci odpowiadają asymptotycznym własnościom dobrze dobranego modelu grafu losowego. Przykładem takiego zastosowania struktur losowych w informatyce jest modelowanie sieci internetowych, sieci stron WWW czy sieci sensorowych. Jednym z podstawowych pytań dotyczących tego zagadnienia jest odpowiedni dobór modelu teoretycznego. Wyniki eksperymentalne pokazują, że badane klasycznie grafy losowe, takie jak graf losowy z niezależnie powstającymi krawędziami czy graf losowy o danej liczbie krawędzi, nie oddają dobrze struktury wielu sieci rzeczywistych. Dlatego w ramach badań w tej tematyce poszukuje się nowych modeli grafów losowych, których struktura lepiej je przybliża a dla proponowanych modeli bada się ich własności asymptotyczne.

Celem rozprawy jest analiza modelu grafu losowego zwanego losowym grafem przecięć pod kątem wykorzystania go jako teoretycznego modelu sieci rzeczywistych. Badania nad losowym grafem przecięć w tym kontekście, zostały zapoczątkowane stosunkowo niedawno. Wszakże pierwsze wyniki dotyczące tego modelu ukazały się w połowie lat dziewięćdziesiątych, ale możliwość zastosowania go do analizy sieci została zauważona prawie dziesięć lat później. Było to istotne odkrycie dla teorii modelowania sieci rzeczywistych, gdyż losowy graf przecięć ma wiele ich cech. Dla porównania proponowane dotychczas modele posiadały tylko niektóre interesujące własności a ich struktura często nie miała związku z mechanizmami tworzenia sieci.

Bardzo ogólna definicja mówi, że losowy graf przecięć powstaje w wyniku procedury, w trakcie której każdy wierzchołek v należący do zbioru wierzchołków \mathcal{V} wybiera niezależnie zbiór własności $D(v)$ z ustalonego zbioru \mathcal{W} zgodnie z zadaniem rozkładem prawdopodobieństwa. Wierzchołki są połączone krawędzią w losowym grafie przecięć wtedy i tylko wtedy, gdy mają co najmniej s wspólnych własności, gdzie s jest z góry ustaloną liczbą (tzn. wierzchołki $v, v' \in \mathcal{V}$ łączą się, gdy zbiory $D(v)$ i $D(v')$ przecinają się na co najmniej s wierzchołkach). Istotnym parametrem tego grafu jest rozkład prawdopodobieństwa $\mathcal{P}_{(m)}$, który decyduje o mocach podzbiorów $D(v)$. Tak zdefiniowany losowy graf przecięć jest modelem bardzo elastycznym, gdyż dla różnych rozkładów $\mathcal{P}_{(m)}$ jego własności mogą zmieniać się diametralnie. W dodatku istnieją w nim zależności między powstającymi krawędziami, dlatego losowy graf przecięć o wiele lepiej niż klasyczny graf z niezależnymi krawędziami modeluje sieci rzeczywiste, w których elementy posiadające wspólnego sąsiada lub elementy o podobnych własnościach łą-

czą się z większym prawdopodobieństwem. Wyniki zawarte w rozprawie będą dotyczyć dwóch szczególnych rodzajów takich sieci. Pierwszy z nich to sieci złożone, a drugi to sieci sensorowe.

Przykładem sieci złożonych w informatyce są sieci stron WWW i sieci internetowe. O przynależności do tej szerokiej klasy sieci decydują charakterystyczne własności, z których najważniejsze są trzy. Pierwsza z nich to rozkład stopni wierzchołków spełniający prawo potęgowe. Mówimy o rozkładzie, że spełnia prawo potęgowe, gdy dla dużych k , liczba wierzchołków o stopniu k jest proporcjonalna do k^{-c} dla ustalonego c . Drugą z własności jest duży współczynnik skupienia. Stanowi on o tym, że elementy posiadające wspólnego sąsiada łączą się częściej niż wynikałoby to ze średniej gęstości sieci. Trzecia cecha to na tyle mała średnica, aby w sieci występował tak zwany efekt „małego świata”. Klasyczny graf losowy z niezależnymi krawędziami ma tylko ostatnią ze wspomnianych cech, natomiast dla odpowiednio dobranych parametrów losowy graf przecięć może mieć wszystkie trzy. Jest to spowodowane tym, że na strukturę sieci złożonych wpływa tendencja do łączenia elementów o podobnych własnościach. Natomiast z definicji wynika, że podobna zasada decyduje o własnościach losowych grafów przecięć.

Drugi rodzaj sieci, który będzie rozpatrywany w tej rozprawie, to sieci sensorowe z losową predystrybucją kluczy. Sieci sensorowe są utworzone z sensorów komunikujących się ze sobą radiowo i badających parametry otoczenia. W modelu sieci sensorowej z losową predystrybucją kluczy sensorom przydzielane są klucze zgodnie z danymi protokołami. Klucze te mają zapewnić bezpieczeństwo połączeń, w ten sposób, że dwa sensory kodują przekazywane informacje za pomocą wspólnych kluczy. Z tego wynika, że sensory komunikują się bezpośrednio tylko, gdy mają co najmniej s wspólnych kluczy, gdzie s oznacza minimalną liczbę kluczy niezbędną do bezpiecznego kodowania. Także w tym przypadku losowy graf przecięć jest naturalnym modelem sieci. Jeśli przyjmiemy, że wierzchołki grafu odpowiadają sensorom a zbiory własności $D(v)$ są podzbiorami przydzielanych sensorowi v kluczy, to struktura krawędzi losowego grafu przecięć odpowiada strukturze połączeń między sensorami.

Struktura rozprawy jest następująca. W pierwszym rozdziale przedstawione zostały podstawowe matematyczne definicje i oznaczenia, które będą wykorzystywane w dalszej części.

Tematem rozdziału drugiego jest omówienie zagadnień związanych z sieciami złożonymi i sieciami sensorowymi. Zaprezentowane tam zostały ich własności oraz omówione zostały modele grafów losowych dotychczas wykorzystywanych do modelowania tych sieci. Głównym celem tego rozdziału jest umotywowanie wyboru losowego grafu przecięć jako teoretycznego modelu badanych sieci oraz przedstawienie dotychczasowych wyników dotyczących dwóch zagadnień: doboru modeli sieci rzeczywistych oraz własności losowego grafu przecięć.

Rozdziały od trzeciego do dziewiątego poświęcone są zaprezentowaniu głównych wyników rozprawy.

Po rozdziale trzecim, w którym zebrane zostały twierdzenia i lematy pomocnicze, w rozdziale czwartym przedstawione zostały twierdzenia dotyczące równoważności różnych modeli losowych grafów przecięć. Rezultaty te pozwa-

lają niektóre wyniki dotyczące jednego modelu losowego grafu przecięć zastosować dla innych modeli. Będą one kilkakrotnie wykorzystywane w dalszej części rozprawy.

Rozdziały od piątego do siódmego poświęcone są analizie własności losowych grafów przecięć, które decydują o tym, że są one dobrymi modelami sieci rzeczywistych, a w szczególności sieci złożonych. W rozdziale piątym udowodnione zostanie bardzo ogólne twierdzenie dotyczące rozkładu stopni wierzchołków, z którego wynika, że można dobrać tak rozkład $\mathcal{P}_{(m)}$, aby stopnie wierzchołków losowego grafu przecięć spełniały prawo potęgowe. Kolejne dwa rozdziały będą dotyczyły współczynnika skupienia i związanego z nim rozkładu liczby klik w grafie. Wyniki w nich zawarte pokażą, że dla niektórych rozkładów $\mathcal{P}_{(m)}$ zależności między powstawaniem krawędzi w losowym grafie przecięć są dość duże, aby dobrze on modelował sieci rzeczywiste.

Rozdziały od ósmego do dziesiątego dotyczą własności związanych ze spójnością losowych grafów przecięć. Przedstawione tam zostały rezultaty, z których wynika między innymi jak dobrać parametry sieci sensorowej, aby komunikacja w niej była efektywna i bezpieczna. Ostatecznie zostanie rozwiązany problem spójności, składowych spójności oraz długości drogi, jaką musi przejść informacja w sieciach sensorowych, w których do komunikacji wykorzystywany jest jeden wspólny klucz. W dodatku niektóre rezultaty zawarte w tych rozdziałach będą dotyczyły o wiele szerszej klasy losowych grafów przecięć, a co za tym idzie mogą one być zastosowane do modelowania także innych sieci.

Ostatni rozdział jest poświęcony krótkiemu podsumowaniu wyników uzyskanych w rozprawie z naciskiem na przedstawienie kierunków dalszych badań nad podjętymi zagadnieniami. Zostaną w nim zatem zaprezentowane ciekawe otwarte problemy dotyczące tematyki rozprawy i hipotezy z nimi związane.

Część wyników przedstawionych w rozprawie zostało już opublikowanych. Niektóre rezultaty z rozdziału 4 i podrozdziału 10.2.3 są częścią pracy [12]. Natomiast główne twierdzenie z rozdziału 9 jest zawarte w pracy [28].

1. Definicje i oznaczenia

1.1. Oznaczenia

Wszystkie granice w rozprawie, o ile nie będzie inaczej wspomniane, będą brane przy $n \rightarrow \infty$. Będą też wykorzystywane standardowe oznaczenia asymptotyczne zdefiniowane jak w [32]:

- $a_n = O(b_n)$ jeśli istnieje stała $C > 0$ i n_0 takie, że $|a_n| \leq Cb_n$ dla każdego $n \geq n_0$;
- $a_n = \Omega(b_n)$ jeśli istnieje stała $c > 0$ i n_0 takie, że $a_n \geq cb_n$ dla każdego $n \geq n_0$;
- $a_n = o(b_n)$ jeśli $a_n/b_n \rightarrow 0$;
- $a_n \sim b_n$ jeśli $a_n/b_n \rightarrow 1$;
- $a_n \asymp b_n$ jeśli istnieją stałe $c > 0$, $C > 0$ i n_0 takie, że $cb_n \leq a_n \leq Cb_n$ dla każdego $n \geq n_0$.

Czasem, dla podkreślenia faktu, że wartości ciągu a_n są ujemne, będziemy wykorzystywali w notacji asymptotycznej oznaczenia $a_n = -O(b_n)$ w znaczeniu, że istnieje ciąg c_n liczb taki, że $a_n = c_n b_n$, $c_n \leq 0$ i $|c_n| \leq C$ dla $n \geq n_0$.

Przez $Po(\lambda)$ będzie oznaczany rozkład Poissona o wartości oczekiwanej λ oraz przez $Bin(n, p)$ rozkład dwumianowy z parametrami n i p . W dodatku, w niektórych przypadkach, parametr $\lambda = X$ lub parametr $p = Y$ będzie pewną zmienną losową. Będziemy wtedy mówili, odpowiednio, że rozpatrywany jest złożony rozkład Poissona $Po(X)$ lub złożony rozkład dwumianowy $Bin(n, Y)$.

Dla zmiennych losowych X i Y oznaczenie $X \prec Y$ będzie stosowane, gdy zmienna losowa X będzie stochastycznie zdominowana przez zmienną losową Y . Dla X i Y o wartościach naturalnych oznacza to, że dla dowolnego $a \in \mathbb{N}$ zachodzi $\Pr\{X \leq a\} \geq \Pr\{Y \leq a\}$.

1.2. Graf losowy

Definicja 1.1. *Nieskierowanym grafem prostym (**grafem**) nazywamy parę $G = (V, E)$, składającą się z niepustego zbioru wierzchołków $V = V(G)$ oraz zbioru krawędzi $E = E(G)$ będących nieuporządkowanymi parami wierzchołków z $V(G)$.*

Krawędź składającą się z wierzchołków v i v' będziemy oznaczać przez vv' .

Definicja 1.2. *Niech \mathcal{G} będzie pewnym zbiorem grafów. **Graf losowy** jest to graf wybrany w wyniku pewnego eksperymentu losowego spośród wszystkich grafów z \mathcal{G} .*

Na potrzeby tej rozprawy zwykle będziemy przyjmować, że $\mathcal{G} = \mathcal{G}_n$, gdzie \mathcal{G}_n jest zbiorem grafów o zbiorze wierzchołków $\mathcal{V} = \{v_1, \dots, v_n\}$.

Przykład 1.1. *Klasycznymi modelami grafów losowych są grafy $\mathbf{G}(n, M)$ i $\mathbf{G}(n, p)$. Pierwszy z nich, graf $\mathbf{G}(n, M)$, był wprowadzony przez Erdősa i Rényi'ego w pracach [21, 22], które zapoczątkowały zainteresowanie grafami losowymi. Jest on wybrany w sposób jednostajny spośród wszystkich grafów na n wierzchołkach i o dokładnie M krawędziach. Drugi z nich, $\mathbf{G}(n, p)$, jest aktualnie najczęściej rozpatrywany. Powstaje on z grafu pełnego na n wierzchołkach przez wyrzucenie z prawdopodobieństwem $1 - p$ każdej krawędzi niezależnie od innych krawędzi. W wielu przypadkach oba modele są sobie równoważne. Wyniki dotyczące równoważności tych modeli można znaleźć na przykład w [13, 32, 43].*

1.3. Losowe grafy przecięć – podstawowe modele

W rozprawie będziemy zajmować się własnościami grafu losowego zwanego losowym grafem przecięć. Niech \mathcal{V} będzie n -elementowym zbiorem, \mathcal{W} zbiorem m -elementowym oraz $\{D(v) : v \in \mathcal{V}\}$ rodziną losowych podzbiorów zbioru \mathcal{W} . Losowym grafem przecięć nazywamy graf o zbiorze wierzchołków \mathcal{V} , w którym dowolne dwa wierzchołki $v \in \mathcal{V}$ i $w \in \mathcal{V}$ są połączone krawędzią wtedy i tylko wtedy gdy zbiory $D(v)$ i $D(w)$ się przecinają. Definicję można uogólnić, jeśli dla pewnej liczby całkowitej $s \geq 1$, krawędź powstaje wtedy i tylko wtedy, gdy $D(v) \cap D(w)$ ma co najmniej s elementów. Uściślając:

Definicja 1.3. *Niech $m = m(n)$ będzie ciągiem liczb całkowitych, $s \geq 1$ będzie stałą całkowitą, $\mathcal{V} = \{v_1, v_2, \dots, v_n\}$ i $\mathcal{W} = \{w_1, w_2, \dots, w_m\}$ będą rozłącznymi zbiorami oraz $\mathcal{P}_{(m)} = (P_0, P_1, \dots, P_m)$ dyskretnym rozkładem prawdopodobieństwa. Załóżmy, że dla każdego $v \in \mathcal{V}$*

1. *najpierw niezależnie od innych elementów z \mathcal{V} , wybieramy moc zbioru $D(v)$ ($|D(v)| = Z_n$) zgodnie z rozkładem $\mathcal{P}_{(m)}$, tzn.*

$$\Pr\{Z_n = k\} = P_k, \quad k = 1, \dots, m;$$

2. *a następnie, jeśli $Z_n = k$, wybieramy zbiór $D(v)$ jednostajnie spośród wszystkich k -elementowych podzbiorów zbioru \mathcal{W} , tzn.*

$$(1.1) \quad \Pr\left\{D(v) = A \mid Z_n = k\right\} = \frac{1}{\binom{m}{k}},$$

dla dowolnego $A \subseteq \mathcal{W}$ takiego, że $|A| = k$.

Losowym grafem przecięć $G_s(\mathcal{V}, m, \mathcal{P}_{(m)})$ nazywamy graf o zbiorze wierzchołków \mathcal{V} , w którym dowolne dwa wierzchołki v i w są połączone krawędzią ($vw \in E(G(\mathcal{V}, m, \mathcal{P}_{(m)}))$) wtedy i tylko wtedy gdy $|D(v) \cap D(w)| \geq s$. W przypadku gdy $s = 1$ graf $G_s(\mathcal{V}, m, \mathcal{P}_{(m)})$ będziemy oznaczać przez $G(\mathcal{V}, m, \mathcal{P}_{(m)})$. Zbiór \mathcal{W} będziemy nazywać **zbiorem własności** losowego grafu przecięć a dla dowolnego $v \in \mathcal{V}$ zbiór $D(v)$ – **zbiorem własności wierzchołka** v .

Definicja tak ogólnego modelu losowego grafu przecięć została zaproponowana przez Godehardta i Jaworskiego w [26]. Wcześniej były rozpatrywane szczególne przypadki tych grafów. Mianowicie po raz pierwszy zagadnienia związane z losowymi grafami przecięć pojawiły się w pracy doktorskiej Singer-Cohen [58] i w pracy Karońskiego, Scheinermana i Singer-Cohen [40]. Zdefiniowany tam model jest losowym grafem przecięć, w którym dla pewnego $p = p(n) \in [0; 1]$ dla dowolnego $v \in \mathcal{V}$ dowolna własność $w \in \mathcal{W}$ zostaje dodana do zbioru $D(v)$ z prawdopodobieństwem p niezależnie od innych wierzchołków i innych własności. Z rozważań zawartych w [26] wynika, że w rozumieniu definicji 1.3, jest to losowy graf przecięć $G(\mathcal{V}, m, \mathcal{P}_{(m)})$, w którym rozkład $\mathcal{P}_{(m)}$ jest rozkładem dwumianowym $\text{Bin}(m, p)$. Drugim interesującym szczególnym przypadkiem jest losowy graf przecięć, w którym dla dowolnego $v \in V$ moc zbioru $|D(v)| = d = d(n)$ z prawdopodobieństwem równym 1, dla pewnej liczby całkowitej d . Własności takiego losowego grafu przecięć były po raz pierwszy badane przez Godehardta i Jaworskiego w [26] w kontekście analizy skupień a potem przez Di Pietro, Mancini, Mei, Panconesi i Radhakrishnan w [55] w kontekście zastosowań w informatyce. Dla tych dwóch szczególnych modeli wprowadzone zostaną dodatkowe nazwy i oznaczenia.

Definicja 1.4. Niech $s \geq 1$ będzie liczbą całkowitą, $p = p(n) \in [0; 1]$, $\alpha > 0$ i $m = n^\alpha$. Losowy graf przecięć $G_s(\mathcal{V}, m, \mathcal{P}_{(m)})$ nazywamy **dwumianowym losowym grafem przecięć** $G_s^{\text{Bin}}(\mathcal{V}, m, p)$ jeśli $\mathcal{P}_{(m)}$ jest rozkładem dwumianowym $\text{Bin}(m, p)$.

Definicja 1.5. Niech $s \geq 1$ i $d = d(n) \leq m(n)$ będą dodatnimi liczbami całkowitymi. Losowy graf przecięć $G_s(\mathcal{V}, m, \mathcal{P}_{(m)})$ nazywamy **jednostajnym losowym grafem przecięć** $G_s^{\text{Jed}}(\mathcal{V}, m, d)$ jeśli $\mathcal{P}_{(m)} = (P_0, P_1, \dots, P_m)$ jest rozkładem zdegenerowanym takim, że $P_d = 1$.

W przypadku, gdy $s = 1$ grafy $G_s^{\text{Bin}}(\mathcal{V}, m, p)$ i $G_s^{\text{Jed}}(\mathcal{V}, m, d)$ będziemy oznaczać przez $G^{\text{Bin}}(\mathcal{V}, m, p)$ i $G^{\text{Jed}}(\mathcal{V}, m, d)$, odpowiednio.

W ostatnim czasie jednym z nurtów badań losowych grafów przecięć jest poszukiwanie takich rozkładów $\mathcal{P}_{(m)}$, dla których rozkład stopni wierzchołków będzie interesujący z punktu widzenia zastosowań. W [19] Deijfen i Kets zdefiniowali i badali własności modelu grafu przecięć z wagami na wierzchołkach będącego uogólnieniem modelu $G^{\text{Bin}}(\mathcal{V}, m, p)$. Poniższa definicja różni się od tej z [19], ale jest jej równoważna. Definicja w poniżej podanej formie pokazuje, że zaproponowany w [19] model jest szczególnym przypadkiem ogólnego modelu losowego grafu przecięć zdefiniowanego w [26].

Definicja 1.6. Niech $\alpha, \beta, \gamma > 0$, $m = \beta n^\alpha$, $p = p(n)$ oraz niech W będzie zmienną losową o wartościach dodatnich. **Dwumianowym losowym grafem przecięć z wagami** $G^{\text{Bin}}(\mathcal{V}, m, W)$ nazywamy losowy graf przecięć, w którym $\mathcal{P}_{(m)}$ jest złożonym rozkładem dwumianowym $\text{Bin}(m, W')$, gdzie $W' = W'(n) = \min\{\gamma W p(n), 1\}$.

W pracy [19] graf ten był zdefiniowany i rozpatrywany dla przypadku $p(n) = n^{-(1+\alpha)/2}$. Ten przypadek był szczególnie interesujący ze względu na podobieństwa z modelem sieci złożonych.

1.4. Własności grafów

Z punktu widzenia zastosowań w analizie sieci będziemy badać niektóre własności grafowe losowych grafów przecięć. Interesujące nas własności definiujemy poniżej. Wszystkie podane definicje odnoszą się do dowolnego grafu $G = (V(G), E(G))$.

1.4.1. Stopień wierzchołka, wierzchołki izolowane

Definicja 1.7. Wierzchołki $v \in V(G)$ i $w \in V(G)$ nazywamy **przyległymi** w G , jeśli $vw \in E(G)$.

Definicja 1.8. **Sąsiedztwem wierzchołka** $v \in V(G)$ nazywamy zbiór wierzchołków do niego przyległych. Zbiór ten oznaczamy przez $N(v)$, to znaczy $N(v) = \{w \in V(G) : vw \in E(G)\}$. Wierzchołki z $N(v)$ nazywamy sąsiadami wierzchołka v .

Definicja 1.9. **Stopniem wierzchołka** $v \in V(G)$ w grafie G nazywamy liczbę wierzchołków do niego przyległych i oznaczamy go przez $\deg(v)$ (tzn. $\deg(v) = |N(v)|$).

Definicja 1.10. Wierzchołek $v \in V(G)$ stopnia zero ($\deg(v) = 0$) nazywamy wierzchołkiem **izolowanym**.

1.4.2. Podgrafy i klikki

Definicja 1.11. **Kliką (grafem pełnym)** na h wierzchołkach nazywamy graf K_h , w którym $|V(K_h)| = h$ oraz $E(K_h) = \{vw : v, w \in V(K_h)\}$. Dodatkowo, dla danego zbioru wierzchołków V' , przez $K_{V'}$ oznaczmy klikę o zbiorze wierzchołków V' .

Definicja 1.12. **Podgrafem grafu** $G = (V(G), E(G))$ nazywamy dowolny graf $G' = (V(G'), E(G'))$ taki, że $V(G') \subseteq V(G)$ i $E(G') \subseteq E(G)$.

Definicja 1.13. Niech $V' \subseteq V(G)$. **Podgrafem grafu G indukowanym** na zbiorze wierzchołków V' nazywamy podgraf grafu G o zbiorze wierzchołków V' i zbiorze krawędzi $\{v_1v_2 : v_1v_2 \in E(G) \text{ oraz } v_1, v_2 \in V'\}$ i oznaczamy go przez $G[V']$.

1.4.3. Spójność, składowe spójności

Definicja 1.14. Niech $v, w \in V(G)$. Niezerowy ciąg $(v = v_0, \dots, v_{k+1} = w)$ taki, że $v_0, \dots, v_{k+1} \in V$, dla każdego $1 \leq i \leq k$, $v_i v_{i+1} \in E(G)$ oraz wierzchołki są parami różne będziemy nazywać **ścieżką** łączącą wierzchołki v i w . **Długością ścieżki** będziemy nazywać liczbę jej krawędzi.

Definicja 1.15. Graf G nazywamy **spójnym**, gdy istnieje ścieżka między dowolną parą wierzchołków z $V(G)$.

Definicja 1.16. **Podzbiór** $S \subseteq V(G)$ nazywamy **spójnym**, gdy istnieje ścieżka między dowolną parą wierzchołków z S .

Definicja 1.17. **Podzbiór** $S \subseteq V(G)$ nazywamy **izolowanym**, gdy nie istnieje krawędź $vw \in E$ taka, że $v \in S$ i $w \in V \setminus S$.

Definicja 1.18. Niech $S \subseteq V(G)$. Podgraf $G[S]$ grafu G nazywamy **składową spójnością grafu G** , jeśli S jest spójny i izolowany.

1.4.4. Średnica

Definicja 1.19. Niech $v, w \in V(G)$. **Odległością** między v i w nazywamy długość najkrótszej ścieżki łączącej wierzchołki v i w i oznaczamy $d(v, w)$. Jeśli w G wierzchołki v i w nie są połączone żadną ścieżką, to $d(v, w) = \infty$.

Definicja 1.20. **Średnicą grafu G** nazywamy największą odległość między wierzchołkami i oznaczamy przez $\text{diam}(G)$. Tzn.

$$\text{diam}(G) = \max\{d(v, w) : v, w \in V(G)\}.$$

W przypadku grafu niespójnego średnica jest nieskończona. Ze względu na zastosowania, w grafie niespójnym, będziemy zainteresowani średnicą składowej spójności grafu o największej liczbie wierzchołków.

1.4.5. Współczynnik skupienia

Współczynnik skupienia (ang. *clustering coefficient*) jest współczynnikiem odzwierciedlającym lokalną gęstość grafu w okolicy danego wierzchołka $v \in V(G)$. Dokładniej, mierzy on, jak fakt, że dwa wierzchołki, powiedzmy w i u , mają wspólnego sąsiada, wpływa na powstanie krawędzi wu między nimi.

Jedną z pierwszych definicji współczynnika skupienia podali Strogatz i Watts w [60]. Zdefiniowany przez nich współczynnik skupienia mierzy jak bardzo graf indukowany na wierzchołkach z sąsiedztwa wierzchołka v różni się od kliku.

Definicja 1.21. Niech $v \in V(G)$ **współczynnikiem skupienia wierzchołka v** nazywamy wielkość:

$$\text{Clus}_1(v) = \begin{cases} \frac{|E(G[N(v)])|}{\binom{|N(v)|}{2}} & \text{dla } |N(v)| \geq 2 \\ 0 & \text{dla } |N(v)| \leq 1 \end{cases}$$

Współczynnikiem skupienia grafu G nazywamy wielkość

$$\text{Clus}_1(G) = \frac{1}{|V(G)|} \sum_{v \in V(G)} \text{Clus}_1(v)$$

W [4] Barrat i Weigt oraz w [49] Newman, Strogatz i Watts podają inną definicję współczynnika skupienia. Zgodnie z tą definicją współczynnik skupienia mierzy jak fakt, że dwa wierzchołki mają wspólnego sąsiada wpływa na powstanie krawędzi między nimi. W definicji tej nie są jednak uwzględnione wierzchołki o stopniu co najwyżej jeden. Dla rozróżnienia będziemy nazywać ten współczynnik globalnym współczynnikiem skupienia, choć w literaturze obie definicje współczynnika skupienia występują pod tą samą nazwą.

Definicja 1.22. Niech G będzie grafem o co najmniej jednym wierzchołku stopnia co najmniej 2. **Globalnym współczynnikiem skupienia grafu** G nazywamy wielkość

$$\begin{aligned} \text{Clus}_2(G) &= \frac{3|\{\{v_1, v_2, v_3\} : v_1v_2, v_2v_3, v_1v_3 \in E(G)\}|}{\sum_{v \in V(G)} \binom{|N(v)|}{2}} = \\ &= \frac{3|\{\{v_1, v_2, v_3\} : v_1v_2, v_2v_3, v_1v_3 \in E(G)\}|}{|\{\{v_1, v_2, v_3\} : v_1v_2, v_2v_3 \in E(G)\}|}, \end{aligned}$$

gdzie $\binom{k}{l} = 0$ jeśli $k < l$.

Warto zauważyć, że $|\{\{v_1, v_2, v_3\} : v_1v_2, v_2v_3, v_1v_3 \in E(G)\}|$ jest liczbą trójkątów w grafie, natomiast $|\{\{v_1, v_2, v_3\} : v_1v_2, v_2v_3 \in E(G)\}|$ liczbą ścieżek długości 2.

W [19, 46] Deijfen i Kets oraz Newman, badając własności losowego grafu przecięć, zainteresowani byli współczynnikiem, który mierzy dokładnie prawdopodobieństwo powstania krawędzi wu w przypadku, gdy w i u mają wspólnego sąsiada.

Definicja 1.23. Niech G będzie grafem losowym. **Lokalnym współczynnikiem skupienia** $\text{Clus}_{loc}(v)$ wierzchołka v nazywamy wielkość:

$$\text{Clus}_{loc}(v) = \Pr \left\{ uw \in E(G) \mid vw \in E(G) \text{ i } vu \in E(G) \right\}.$$

W [19, 46] powyżej zdefiniowana wielkość występuje pod nazwą współczynnika skupienia. W rozprawie jednak w celu zachowania jednoznaczności będziemy stosować powyżej wprowadzone nazewnictwo.

Zauważmy, że w przypadku grafów losowych istnieje zależność między wartością oczekiwaną współczynnika skupienia wierzchołka a lokalnym współczynnikiem skupienia. Na przykład w grafie $G^{Jed}(\mathcal{V}, m, d)$ dla $k = 0, 1$

$$\mathbb{E}(\text{Clus}_1(v) \mid \deg(v) = k) = 0$$

a dla $k \geq 2$

$$\begin{aligned} &\mathbb{E}(\text{Clus}_1(v) \mid \deg(v) = k) = \\ &= \sum_{\substack{V' \subseteq \mathcal{V}, \\ |V'|=k}} \mathbb{E}(\text{Clus}_1(v) \mid N(v) = V') \Pr \{N(v) = V' \mid \deg(v) = k\} = \\ &= \sum_{\substack{V' \subseteq \mathcal{V}, \\ |V'|=k}} \frac{\mathbb{E} \left(\left| E(G^{Jed}(\mathcal{V}, m, d)[V']) \right| \mid N(v) = V' \right)}{\binom{k}{2}} \cdot \frac{1}{\binom{n}{k}} = \\ &= \frac{1}{\binom{n}{k}} \sum_{\substack{V' \subseteq \mathcal{V}, \\ |V'|=k}} \frac{\sum_{v', v'' \in V', v' \neq v''} \Pr \left\{ v'v'' \in E(G^{Jed}(\mathcal{V}, m, d)) \mid N(v) = V' \right\}}{\binom{k}{2}} = \\ &= \Pr \left\{ v'v'' \in E(G^{Jed}(\mathcal{V}, m, d)) \mid vv', vv'' \in E(G^{Jed}(\mathcal{V}, m, d)) \right\} = \\ &= \text{Clus}_{loc}(v). \end{aligned}$$

Ostatecznie z niezależności wyborów zbiorów własności, jeśli \hat{p} jest równe prawdopodobieństwu powstania krawędzi w $G^{Jed}(\mathcal{V}, m, d)$ oraz $n\hat{p} \sim c$ dla pewnej stałej c .

$$\begin{aligned} \mathbb{E}\text{Clus}_1(v) &= \sum_{k=0}^{n-1} \mathbb{E}(\text{Clus}_1(v) \mid \deg(v) = k) \Pr\{\deg(v) = k\} = \\ &= \text{Clus}_{loc}(v)(1 - (1 - \hat{p})^n - n\hat{p}(1 - \hat{p})^{n-1}) \sim \text{Clus}_{loc}(v)(1 - e^{-c} - ce^{-c}). \end{aligned}$$

1.5. Własności asymptotyczne

Przypomnijmy, że mając dany zbiór $\mathcal{G} = \mathcal{G}_n$ wszystkich grafów o zbiorze wierzchołków $\mathcal{V} = \{v_1, \dots, v_n\}$, przez graf losowy będziemy rozumieć graf wybrany spośród wszystkich grafów z \mathcal{G} w wyniku pewnego eksperymentu losowego. Zgodnie z definicją podaną w [32]:

Definicja 1.24. Podzbiór $\mathcal{A}_n \subseteq \mathcal{G}_n$ nazywamy **własnością grafu** jeśli dla dowolnego $G \in \mathcal{A}_n$ podzbiór \mathcal{A}_n zawiera wszystkie izomorficzne kopie grafu G . Mówimy, że graf G ma własność \mathcal{A}_n , jeśli $G \in \mathcal{A}_n$.

Definicja 1.25. Własność $\mathcal{A}_n \subseteq \mathcal{G}_n$ nazywamy **rosnącą** (odpowiednio **malejącą**), gdy jeśli $G = (\mathcal{V}, E) \in \mathcal{A}_n$, to $G' \in \mathcal{A}_n$ dla dowolnego grafu $G' \in \mathcal{G}_n$ takiego, że $E(G) \subseteq E(G')$ (odpowiednio $E(G') \subseteq E(G)$). Wszystkie własności rosnące i malejące nazywamy **własnościami monotonicznymi**.

Przykład 1.2. Przykładami własności monotonicznych są spójność (własność rosnąca), posiadanie największej składowej spójności o co najmniej k wierzchołkach (własność rosnąca) lub co najwyżej k wierzchołkach (własność malejąca), zawieranie klikki na h wierzchołkach jako podgrafu (własność rosnąca)

1.5.1. Funkcje progowe

Definicja 1.26. Mówimy, że graf losowy **asymptotycznie prawie na pewno, ma własność \mathcal{A} jeśli**

$$\Pr\{G_n \in \mathcal{A}\} \rightarrow 1 \text{ gdy } n \rightarrow \infty.$$

oraz **asymptotycznie prawie na pewno, własności \mathcal{A} nie ma jeśli**

$$\Pr\{G_n \in \mathcal{A}\} \rightarrow 0 \text{ gdy } n \rightarrow \infty.$$

Jednym z głównych odkryć Erdősa i Rényi'ego ([21, 22]) jest fakt, że dodanie niewielkiej (z asymptotycznego punktu widzenia) liczby krawędzi do grafu losowego $G(n, M)$ może bardzo zmienić jego własności. Podobnie w grafie $G(n, p)$ małe zwiększenie wartości p może zmienić strukturę tego grafu losowego. Takie graniczne wartości M i p nazywamy funkcjami progowymi.

Definicja 1.27. Funkcją progową własności \mathcal{A} w grafie $G(n, p)$ nazywamy funkcję $\tilde{p} = \tilde{p}(n)$ taką, że

Jeśli $\frac{n}{\tilde{p}} \rightarrow 0$, to $G(n, p)$ ma własność \mathcal{A} asymptotycznie prawie na pewno.

Jeśli $\frac{p}{\tilde{p}} \rightarrow \infty$, to $G(n, p)$ własności \mathcal{A} asymptotycznie prawie na pewno nie ma.

lub

Jeśli $\frac{p}{\tilde{p}} \rightarrow 0$, to $G(n, p)$ nie ma własności \mathcal{A} asymptotycznie prawie na pewno.

Jeśli $\frac{p}{\tilde{p}} \rightarrow \infty$, to $G(n, p)$ ma własność \mathcal{A} asymptotycznie prawie na pewno.

Funkcją progową własności \mathcal{A} w grafie $G(n, M)$ nazywamy taką funkcję $\tilde{M} = \tilde{M}(n)$, że

Jeśli $\frac{M}{\tilde{M}} \rightarrow 0$, to $G(n, M)$ ma własność \mathcal{A} asymptotycznie prawie na pewno.

Jeśli $\frac{M}{\tilde{M}} \rightarrow \infty$, to $G(n, M)$ własności \mathcal{A} asymptotycznie prawie na pewno nie ma.

lub

Jeśli $\frac{M}{\tilde{M}} \rightarrow 0$, to $G(n, M)$ nie ma własności \mathcal{A} asymptotycznie prawie na pewno.

Jeśli $\frac{M}{\tilde{M}} \rightarrow \infty$, to $G(n, M)$ ma własność \mathcal{A} asymptotycznie prawie na pewno.

W [32] można znaleźć prosty dowód, że każda własność monotoniczna ma funkcję progową w $G(n, p)$ i $G(n, M)$. W powyższej definicji funkcja progowa jest określona z dokładnością do rzędu, to znaczy, że jeśli $\tilde{p} \asymp \tilde{p}'$ i \tilde{p} jest funkcją progową, to \tilde{p}' też jest funkcją progową. W przypadku wielu własności można określić funkcje progowe dokładniej.

W sposób analogiczny można zdefiniować funkcje progowe \tilde{d} dla własności grafu $G_s^{Jed}(\mathcal{V}, m, d)$ i funkcje progowe \tilde{p} dla własności grafu $G_s^{Bin}(\mathcal{V}, m, p)$. Chcielibyśmy także zwrócić uwagę, że dla własności ogólnego modelu grafu $G_s(\mathcal{V}, m, \mathcal{P}_{(m)})$ można określić funkcję progową która jest rodziną rozkładów. Jak dokładnie będziemy to rozumieć będzie widoczne w momencie, gdy te funkcje będziemy wyznaczać.

2. Sieci rzeczywiste i ich teoretyczne modele

W rozdziale poniższym znajdują się rozważania na temat praktycznych zastosowań wyników teoretycznych przedstawionych w dalszej części rozprawy. Głównym jego celem jest zaprezentowanie argumentów na to, że losowy graf przecięć najlepiej ze znanych modeli grafów losowych odzwierciedla strukturę sieci złożonych i sieci sensorowych oraz przedstawienie motywacji doboru zagadnień teoretycznych rozpatrywanych w dalszej części rozprawy.

2.1. Sieci złożone

2.1.1. Sieci złożone i ich podstawowe własności

Z badań eksperymentalnych nad sieciami rzeczywistymi wynika, że wiele z nich posiada podobne cechy. Podstawowe trzy z nich to:

- rozkład stopni wierzchołków spełniający prawo potęgowe (ang. *power law distribution*);
- wysoki współczynnik skupienia (ang. *clustering coefficient*);
- efekt małego świata (ang. *small world*).

Sieci posiadające wymienione własności nazywa się sieciami złożonymi. Zostały one szeroko opisane w przeglądowych pracach na ich temat, na przykład [2, 47]. Przykładami obiektów z tej klasy są:

- sieć współpracy autorów opracowań naukowych,
- sieć połączeń telefonicznych,
- sieć cytowań,
- sieć znajomości ludzi w danym kraju,
- sieć współwystępowania aktorów w filmach;

natomiast w informatyce:

- sieć internetowa,
- sieć stron WWW.

Wyniki zawarte w tej rozprawie odnoszą się do wszystkich wymienionych powyżej sieci, więc w szczególności do dwóch ostatnich sieci złożonych występujących w informatyce. Sieć wymieniona jako przedostatnia, czyli sieć internetowa, jest modelowana grafem, w którym wierzchołki odpowiadają routery a krawędzie reprezentują, zarówno przewodowe jak i bezprzewodowe, połączenia między nimi. Natomiast wierzchołkom grafu modelującego ostatnią z sieci odpowiadają strony WWW a krawędziom hiperlinki. Ponieważ w naturalny sposób można zdefiniować skierowanie krawędzi łączącej dwie strony WWW, dlatego często przyjmuje się za model tej sieci graf skierowany. W rozprawie będziemy jednak analizować jego nieskierowany graf podległy.

Wymienione cechy sieci złożonych przekładają się na pożądane własności ich teoretycznych modeli grafowych. Kolejne trzy krótkie podrozdziały poświęcone są ich przedstawieniu.

Rozkład spełniający prawo potęgowe

W każdej pracy dotyczącej sieci bezskalowych lub sieci złożonych znajduje się krótkie omówienie definicji rozkładu spełniającego prawo potęgowe. Szczegółowiej jego własności zostały opisane na przykład w [48]. Rozkład spełniający prawo potęgowe różni się znacznie od rozkładu Poissona charakterystycznego dla rozkładu stopni w klasycznym grafie $G(n, p)$. Przyjmuje się, że zmienna losowa X ma rozkład, który spełnia **prawo potęgowe** jeśli

$$\Pr\{X \geq x\} \sim cx^{-\alpha}$$

dla pewnych stałych $c > 0$ i $\alpha > 0$. Dla dużych x jest to o wiele większa wartość niż wartość $\Pr\{Y \geq x\}$ dla zmiennej Y o rozkładzie Poissona. Przykładem zmiennej losowej spełniającej prawo potęgowe jest zmienna losowa o **rozkładzie Pareto** przedstawionym poniżej. Czasami rozkładami Pareto nazywa się wszystkie rozkłady spełniające prawo potęgowe lub mówi się o nich, że spełniają **prawo Zipfa**. Przykładem ciągłego rozkładu spełniającego prawo potęgowe jest rozkład o funkcji gęstości

$$f(x) = \alpha k^\alpha x^{-\alpha-1} \mathbb{I}_{\geq k}.$$

Dla zmiennej losowej X o tym rozkładzie zachodzi, że z prawdopodobieństwem jeden $X \geq k$ oraz dla $x \geq k$

$$\Pr\{X \geq x\} = \left(\frac{x}{k}\right)^{-\alpha}.$$

Można podać też naturalny przykład rozkładu dyskretnego spełniającego prawo potęgowe. Niech $N \in \mathbb{N}$ oraz $\alpha > 0$. Oznaczmy przez $H_{N,\alpha}$ uogólnioną liczbę harmoniczną tzn.

$$H_{N,\alpha} = \sum_{k=1}^N \frac{1}{k^\alpha}.$$

Dla danych parametrów N i α oznaczmy przez $X_{N,\alpha}$ zmienną losową o rozkładzie podanym poniżej

$$\Pr\{X_{N,\alpha} = k\} = \frac{k^{-\alpha}}{H_{N,\alpha}}.$$

Wtedy

$$\mathbb{E}(X_{N,\alpha}) = \frac{H_{N,\alpha-1}}{H_{N,\alpha}}$$

oraz

$$\mathbb{E}(X_{N,\alpha}^2) = \frac{H_{N,\alpha-2}}{H_{N,\alpha}}.$$

Oczywiście $H_{N,\alpha} < \infty$ przy $N \rightarrow \infty$ wtedy i tylko wtedy gdy $\alpha > 1$. Można zatem zdefiniować dla $\alpha > 1$ zmienną X_α przez

$$\Pr\{X_\alpha = k\} = \frac{k^{-\alpha}}{\lim_{N \rightarrow \infty} H_{N,\alpha}}.$$

Dla $2 < \alpha < 3$ zmienna ta ma skończony pierwszy i nieskończony drugi moment.

Wysoki współczynnik skupienia

Współczynnik skupienia w strukturze sieci mierzy liczbę trójkątów w teoretycznym modelu sieci w stosunku do liczby krawędzi w tym modelu (patrz definicje w podrozdziale 1.4.5). Wysoki, w stosunku do wielkości sieci, współczynnik skupienia jest cechą charakterystyczną sieci złożonych. Najlepszym przykładem sieci złożonej, w której współczynnik skupienia jest duży jest sieć znajomości, gdyż jeśli dwóch ludzi zna pewną trzecią osobę, to jest bardzo prawdopodobne, że tych dwóch ludzi też jest znajomymi. Zakłada się, że w dobrym modelu sieci złożonych współczynnik skupienia jest asymptotycznie stały, to znaczy niezależny od wielkości sieci.

Efekt małego świata

Efekt małego świata polega na tym, że mimo dużego rozmiaru sieci, między dowolnymi dwoma wierzchołkami grafu reprezentującego sieć istnieje krótka ścieżka. W sieciach rzeczywistych może to być na przykład ścieżka długości 3 – jak w przypadku sieci aktorów występujących we wspólnym filmie lub 6 – w przypadku sieci znajomości w USA (patrz [2]). Z definicji średnicy grafu wynika, że dana sieć podlega efektowi małego świata, gdy jej teoretyczny model ma małą średnicę. W przypadku teoretycznego modelu sieci złożonej zakłada się, że średnica rzędu równa $\ln n$, gdzie n jest liczbą elementów sieci, zapewnia istnienie efektu małego świata.

2.1.2. Podstawowe pytania i modele grafów

Pierwotnie duże sieci próbowano modelować klasycznym grafem losowym $G(n, p)$. Wraz z rozwojem badań eksperymentalnych i odkryciem, że własności wielu sieci rzeczywistych znacznie odbiegają od własności modelu grafu $G(n, p)$, zaczęto poszukiwania modeli grafów losowych, które bardziej przypominają strukturą sieci rzeczywiste.

W badaniach dotyczących różnych teoretycznych modeli sieci rzeczywistych, w tym sieci złożonych, przede wszystkim poszukuje się ich nowych modeli, które zastąpią graf $G(n, p)$. W tym przypadku proponuje się najpierw model i próbuje się zanalizować jego własności pod kątem podobieństw do sieci złożonych. Bada się wtedy rozkład stopni wierzchołków, współczynnik skupienia, średnicę lub średnie odległości między wierzchołkami. Następnym krokiem jest rozpatrzenie modelu o interesujących własnościach i zbadanie jego najważniejszych cech z punktu widzenia sieci. A mianowicie, badanie spójności i przejścia fazowego albo procesów epidemicznych.

W związku z tym, że trudno jest odnaleźć model grafu, który posiadałby wszystkie trzy cechy sieci złożonych rozważa się sieci, które mają co najmniej jedną z tych cech.

Pierwszym przykładem są *sieci bezskalowe*, w których wierzchołki mają stopnie zgodne z rozkładem spełniającym prawo potęgowe. Najprostszym modelem grafu losowego, który może modelować sieci bezskalowe jest graf wybrany w sposób jednostajny spośród rodziny wszystkich grafów o danym ciągu stopni. Grafy takie były badane między innymi przez Molloy'a i Reeda w [44, 45]. Innym sposobem generowania grafów modelujących sieci bezskalowe jest tak zwane uprzywilejowane dołączanie (ang. *preferential attachment*)

zgodnie z którym generujemy graf dodając kolejne wierzchołki w ten sposób, że dodawane wierzchołki z większym prawdopodobieństwem zostają sąsiadami wierzchołków o dużym stopniu. Jak podaje Newmann w [47] taki sposób generowania grafów o ciągu stopni spełniającym prawo potęgowe pojawiał się we wczesnych pracach dotyczących sieci, ale stał się powszechnie wykorzystywany i cytowany dzięki pracom Alberta i Barabási'ego nad tym modelem (patrz [2, 1]). Jednakże, choć współczynnik skupienia w tak skonstruowanym modelu sieci jest wyższy niż współczynnik skupienia w $G(n, p)$, to maleje on wraz ze wzrostem wielkości sieci (patrz [2]).

Alternatywnym sposobem jest skupienie się na fenomenie małego świata. Klasyczny model, który oddaje charakter sieci podlegającej efektowi małego świata, opiera się na regularnej siatce a dokładniej na cyklu w którym dodatkowe krawędzie łączą każdy wierzchołek z wierzchołkami w odległości k od niego w cyklu. Losowość w tym modelu uzyskuje się poprzez losową modyfikację krawędzi pierwotnego grafu lub dodawanie w sposób losowy nowych krawędzi. Model ten został zaproponowany przez Watts'a i Strogatza w [60] i zmodyfikowany został przez Newmana i Watts'a w [50, 51]. O wiele lepiej niż wymienione powyżej modele sieci bezskalowych obrazuje on własności skupienia i odległości między wierzchołkami. Niestety stopnie wierzchołków w tym modelu nie spełniają prawa potęgowego.

Żaden z powyżej wspomnianych modeli nie oddaje charakteru sieci złożonych w całości. Tym bardziej, w konstrukcji modelu, nie ma ukrytej własności decydującej o specyficznej strukturze sieci. Dopiero praca [29] Guillaume'a i Latapy'iego próbuje odpowiedzieć na pytanie, co wpływa na to, że wiele sieci ma takie a nie inne cechy. Zwracają oni uwagę na to, że pewne własności elementów sieci mogą decydować o jej strukturze. Sugerują oni, że istnieje pewna, ukryta lub nie, dwudzielna struktura, w której mamy dwa zbiory: zbiór elementów sieci i zbiór ich własności. Elementy z tego drugiego zbioru decydują o połączeniach między elementami sieci. Warto wspomnieć, że wcześniej także Newman w [46] zaproponował model sieci złożonej bazujący na dwudzielnej strukturze elementów sieci i własności. Po tych pracach ukazały się kolejne, dotyczące już losowych grafów przecięć np. [9, 11, 19], w których próbowano analizować znany model ogólnego losowego grafu przecięć pod kątem analizy sieci złożonych.

2.1.3. Stan wiedzy

Spostrzeżenie przez Guillaume'a i Latapy'iego w [29] dwudzielnej struktury kryjącej się za sieciami złożonymi zapoczątkowało poszukiwania w tym kierunku. Pojawiły się wyniki takie jak algorytm Behrischa i Taraza wyszukujący ukrytą dwudzielną strukturę w danej sieci (patrz [6]). Z punktu widzenia analizy teoretycznych modeli pojawiły się pierwsze prace Deijfena i Ketsa [19] oraz Bloznelisa [19] poszukujące szczególnych przykładów rozkładów $\mathcal{P}_{(m)}$, dla których losowy graf przecięć miałby rozkład stopni wierzchołków spełniające prawo potęgowe. W [19] Deijfen i Kets pokazali także, że zaproponowany przez nich model, przy odpowiednim doborze parametrów ma współczynnik skupienia niezależny od wielkości grafu. Dla losowych grafów przecięć prace były kontynuowane przez Bloznelisa w [9, 10] oraz Lageråsa i Lindholma

w [42] w kierunku badania własności grafów związanych z przejściem fazowym [9, 42] i średnimi odległościami między wierzchołkami [10]. Jednak nadal wiele pytań zostaje otwartych. Wyniki Bloznelisa w pracach [9, 11, 10] dotyczą przypadku gdy $n = o(m)$ oraz nie pokazują współczynnika skupienia dla rozpatrywanego modelu. Natomiast w [19] przedstawione są rezultaty pokazujące przykład rozkładu $\mathcal{P}_{(m)}$ dla którego stopnie wierzchołków spełniają prawo potęgowe i dla tego rozkładu obliczony zostaje współczynnik skupienia, ale średnica i spójność nie są analizowane. Wynik z [42] dotyczy grafu o dobrym współczynniku skupienia, ale stopnie wierzchołków w rozpatrywanym modelu nie spełniają prawa potęgowego.

W tej rozprawie zostanie dokładniej zanalizowany model rozpatrywany przez Bloznelisa w [9, 11, 10]. Pokazany zostanie alternatywny dowód dotyczący ciągu stopni wierzchołków, dzięki któremu udowodnione zostanie twierdzenie z [11] dla wcześniej nierozpatrywanego przypadku $n \asymp m$. Wyliczona także zostanie wartość lokalnego współczynnika skupienia dla tego modelu, która okaże się być niezależna od wielkości grafu tylko dla $n \asymp m$.

2.2. Sieci sensorowe

Bezprzewodowa sieć sensorowa jest to sieć składająca się z rozrzuconych na pewnym obszarze małych urządzeń zwanych dalej dla uproszczenia sensorami. Posiadają one zdolność do komunikacji bezprzewodowej oraz są wyposażone w czujnik mierzący interesujące cechy otoczenia takie jak na przykład poziom hałasu, ciśnienie, temperaturę, wilgotność. Bezprzewodowe sieci sensorowe służą do zbierania i przekazywania informacji na temat otoczenia. W ostatnim czasie okazały się mieć one szerokie zastosowania w monitoringu środowiska, rolnictwie, medycynie, kontroli dystrybucji, monitoringu przemysłowym a także wojskowości. Ze względu na potrzebę umieszczenia wielu sensorów na monitorowanym obszarze są to urządzenia małe i niezbyt drogie. Z tego powodu są one zasilane bateriami o ograniczonej wydajności oraz mają procesor o niewielkiej mocy obliczeniowej. Nie posiadają one także dodatkowych zabezpieczeń, czyli jeśli ktoś przejmie fizyczną kontrolę nad sensorem może odczytać wszystkie informacje zawarte w jego pamięci. W związku z tymi ograniczeniami sensory nie mają możliwości stosowania standardowo używanych protokołów kryptograficznych zapewniających bezpieczeństwo przekazywania informacji. W dodatku atakujący sieć sensorową przechwytyjąc sensory uzyskuje informację o wszystkich kluczach w nich zawartych.

2.2.1. Sieć sensorowa z losową predystrybucją kluczy

Jednym ze sposobów przewyciężenia powyżej wspomnianych problemów związanych z fizycznymi ograniczeniami sensorów jest zastosowanie tak zwanej losowej predystrybucji kluczy (ang. *random key predistribution*) zaproponowanej przez Eschenauera i Gligora w [24]. Losowa predystrybucja kluczy przebiega w trzech etapach. W pierwszym generowana jest duża pula kluczy (np. m kluczy), którym przyporządkowuje się identyfikatory a następnie z tej puli losuje się niezależnie tą samą liczbę kluczy (np. d kluczy) dla każdego

sensora. Drugi etap odbywa się po rozrzuceniu sensorów w terenie. Wtedy każdy sensor odkrywa swoich sąsiadów i próbuje ustalić z każdym z nich wspólny klucz. Opcjonalny trzeci etap ma na celu zapewnienie sensorom znajdującym się w zasięgu komunikacji bezprzewodowej wspólnych kluczy. W naszych rozważaniach będziemy jednak zakładać, że wszystkie sensory znajdują się w zasięgu komunikacji i będziemy zainteresowani dwoma pierwszymi etapami.

Podsumowując, w rozpatrywanym przez nas modelu sieci sensorowej z losową predystrybucją kluczy, każdemu sensorowi przyporządkowuje się niezależnie d kluczy spośród m kluczy z puli oraz dwa sensory mogą bezpośrednio sobie przekazywać informacje wtedy i tylko wtedy gdy ich zbiory kluczy mają co najmniej jeden element wspólny.

2.2.2. Podstawowe pytania i modele grafów.

Modelem teoretycznym sieci sensorowej jest graf, w którym wierzchołki odpowiadają sensorom a krawędzie przedstawiają możliwość bezpośredniej komunikacji między sensorami. W pierwszej pracy dotyczącej losowej predystrybucji kluczy [24] Eschenauer i Gligor zaproponowali jako model teoretyczny do analizy sieci sensorowej klasyczny graf losowy $G(n, p)$. W [55] Di Pietro, Mancini, Mei, Panconesi i Radhakrishnan stawiają tezę, że graf $G(n, p)$, ze względu na niezależność krawędzi, nie jest modelem teoretycznym dobrze odzwierciedlającym własności sieci sensorowej. Zauważają oni, że to losowy graf przecięć najlepiej pokazuje zależności między połączeniami komunikacyjnymi między sensorami a kluczami, które im zostały przypisane.

W losowym grafie przecięć odzwierciedlającym strukturę sieci sensorowej, każdemu wierzchołkowi przyporządkowany jest sensor a każdemu dostępnemu kluczowi dokładnie jeden element ze zbioru własności. Każdemu z wierzchołków (sensorowi) przyporządkowuje się d własności (d kluczy) a krawędzie powstają między tymi wierzchołkami, które mają co najmniej jedną własność wspólną. Podsumowując idealnym modelem do teoretycznej analizy własności sieci sensorowych jest jednostajny losowy graf przecięć $G^{Jed}(\mathcal{V}, m, d)$.

Podstawowymi zagadnieniami dotyczącymi bezprzewodowych sieci sensorowych są spójność sieci, odporność na ataki oraz liczba skoków jakie musi przejść informacja, żeby została dostarczona z jednego sensora do dowolnego innego sensora. Przez skok rozumiemy tutaj przekazanie informacji bezpośrednio z jednego sensora do innego sensora za pomocą wspólnych kluczy. Celem wielu prac teoretycznych dotyczących sieci sensorowych w tym [12, 24, 30, 55] jest określenie w jakim stosunku muszą być parametry n , m i d , żeby informacja wysłana przez jeden sensor mogła dotrzeć bezpiecznie do dowolnego innego sensora lub prawie wszystkich innych sensorów w małej liczbie skoków.

Problem dostarczenia informacji do prawie wszystkich sensorów sprowadza się do wyznaczenia wielkości największej składowej spójności w jednostajnym losowym grafie przecięć $G^{Jed}(\mathcal{V}, m, d)$. Natomiast maksymalna liczba skoków, jakie musi zrobić informacja, jest równa średnicy największej składowej spójności grafu $G^{Jed}(\mathcal{V}, m, d)$. Zagadnienie bezpieczeństwa wiąże się z określeniem jak zebranie przez atakującego małej frakcji sensorów wpływa

na liczbę skompromitowanych połączeń, to znaczy połączeń kodowanych za pomocą kluczy, które atakujący zdobył z zebranych sensorów. Dokładniej zagadnienie sprowadza się do sprawdzenia, czy zebranie kluczy z małej liczby sensorów ($o(|V(G^{Jed}(\mathcal{V}, m, d))|)$ sensorów) może pozwolić atakującemu na odczytanie wiadomości z dużej liczby połączeń ($\Omega(|E(G^{Jed}(\mathcal{V}, m, d))|)$).

Należy wspomnieć, że model losowej predystrybucji kluczy z [24] został uogólniony przez Chana, Perriga i Songa w pracy [15]. Uogólniony model został nazwany q -kompozytowym modelem predystrybucji kluczy. W modelu tym dwa sensory mogą bezpośrednio wymieniać informacje, jeśli mają co najmniej $q \geq 1$ wspólnych kluczy i informacje przekazują wykorzystując wszystkie klucze, które ze sobą dzielą. W [15] Chan, Perrig i Song przedstawiają tezę, że w tym modelu zwiększa się odporność na „małe” ataki, tzn. ataki w których liczba przechwyconych sensorów jest mała w porównaniu do liczby sensorów w całej sieci. Natomiast w przypadku „dużych” ataków model przedstawiony w [24] jest zdecydowanie odporniejszy. Dla modelu q -kompozytowego najodpowiedniejszym teoretycznym odpowiednikiem sieci jest jednostajny losowy graf przecięć $G_s^{Jed}(\mathcal{V}, m, d)$ z parametrem przecięcia $s = q$.

2.2.3. Stan wiedzy

W pierwszych pracach dotyczących predystrybucji kluczy w bezprzewodowych sieciach sensorowych (np. [15, 24, 30]) jako graf modelujący sieć wykorzystywano graf $G(n, p)$ z odpowiednimi parametrami. Pierwszym artykułem, w którym jako model teoretyczny został zastosowany losowy graf przecięć była praca [55] Di Pietro, Mancini’ego, Mei, Panconesi’ego i Radhakrishnana. Zauważono wtedy, że istnieją zależności między bezpiecznymi połączeniami między sensorami oraz że stosownym modelem do modelowania bezprzewodowych sieci sensorowych jest graf $G^{Jed}(\mathcal{V}, m, d)$. W pracy tej rozważano problem spójności i bezpieczeństwa sieci. Udowodniono zostało, że dla $c > 8$ i $m \geq n$ jeśli

$$(2.1) \quad \frac{d^2}{m} \sim \frac{c \ln n}{n},$$

to asymptotycznie prawie na pewno bezprzewodowa sieć sensorowa jest spójna. Pokazane także zostało, że dla

$$\frac{d}{m} \succ \frac{1}{n},$$

sieć ma własność *redoubtability*, tzn. przechwycenie $o(n)$ sensorów, gdzie n jest liczbą sensorów w sieci, nie pozwoli rozszyfrować stałej frakcji wszystkich komunikatów.

Problem spójności jednostajnego losowego grafu przecięć, nie tylko w kontekście bezprzewodowych sieci sensorowych, był rozważany w [56],[27] i [8]. W pracy [27] Godehardt, Jaworski i Rybarczyk uzyskali pewne wyniki dotyczące spójności $G^{Jed}(\mathcal{V}, m, d)$ dla d będącego stałą. Natomiast w [8] Blackburn i Gerke badali spójność tego grafu dla $m = n^\alpha$, gdzie α jest pewną stałą dodatnią. Blackburn i Gerke udowodnili, że jeśli w (2.1) $c > 1$, to

$G^{Jed}(\mathcal{V}, m, d)$ jest spójny asymptotycznie prawie na pewno oraz dla $c < 1$ asymptotycznie prawie na pewno jest niespójny. Postawili także hipotezę, że można określić dokładniej funkcję progową dla spójności przy $c \sim 1$.

W [30] Hwang i Kim zauważają, że czasami w zastosowaniach może wystarczyć, że tylko stała frakcja zbioru sensorów (np. 99%) może się komunikować. Wtedy godzimy się na to, że dane z niektórych sensorów nie będą przekazywane w sieci, ale zyskujemy na bezpieczeństwie przekazywania informacji. Problem określenia stosunku parametrów n , m i d w $G^{Jed}(\mathcal{V}, m, d)$, dla których tylko mała frakcja zbioru wierzchołków będzie poza częścią sieci, która może bez problemu się komunikować, jest związane z tak zwanym przejściem fazowym. Przejście fazowe jest to „nagle” pojawienie się w grafie losowym składowej spójności zawierającej stałą frakcję zbioru wierzchołków. Dla klasycznego grafu losowego problem ten rozpatrywali Erdős i Rényi w [22]. Zauważyli oni, że jeśli $p = cn^{-1}$, (funkcja progowa dla przejścia fazowego), wtedy dla $c < 1$ asymptotycznie prawie na pewno graf losowy ma tylko składowe o wielkości $O(\ln n)$, natomiast dla $c > 1$ graf ma dokładnie jedną składową o $\Omega(n)$ wierzchołkach a pozostałe składowe są wielkości $O(\ln n)$. Dodatkowo byli oni w stanie określić jak liczba wierzchołków zawartych w największej składowej zależy od c .

W kontekście bezprzewodowych sieci sensorowych pierwszą próbą określenia wielkości parametrów grafu $G^{Jed}(\mathcal{V}, m, d)$ dla których zachodzi przejście fazowe była praca [12] Bloznelisa, Jaworskiego i Rybarczyk. Wynik tam uzyskany dotyczył d takiego, że $\ln n = o(d)$ oraz

$$\frac{d^2}{m} \sim \frac{c}{n}.$$

Pokazane zostało, że przejście fazowe zachodzi dla $c = 1$. To znaczy, gdy $c < 1$, to wszystkie składowe są wielkości $O(\ln n)$, natomiast dla $c > 1$ istnieje dokładnie jedna składowa o rozmiarze $\Omega(n)$. Wcześniej w [27, 56] Godehard, Jaworski i Rybarczyk pokazali analogiczny wynik dotyczący stałych wartości d . Dla stałego d funkcją progową dla przejścia fazowego jest

$$\frac{d(d-1)}{m} \sim \frac{c}{n}.$$

dla $c = 1$.

Jedynym wynikiem dotyczącym modelu q -kompozytowego, tzn. grafu $G_s^{Jed}(\mathcal{V}, m, d)$ dla $s = q$, jest wynik, który uzyskali Bloznelis, Jaworski i Rybarczyk zawarty w [12].

W rozprawie tej zostanie udowodniony wynik, który ostatecznie odpowie na pytanie dotyczące funkcji progowej dla spójności i przejścia fazowego w grafie $G^{Jed}(\mathcal{V}, m, d)$ dla wszystkich $d = d(n)$ i $m = m(n)$, zostanie określona też średnica największej składowej spójności w $G^{Jed}(\mathcal{V}, m, d)$ w momencie pojawienia się składowej spójności o $\Omega(n)$ wierzchołkach i w momencie uspojnienia oraz wynik dotyczący przejścia fazowego w $G_s^{Jed}(\mathcal{V}, m, d)$. W kontekście badania spójności modelu q -kompozytowego przedstawione będzie twierdzenie dotyczące liczby wierzchołków izolowanych w $G_s^{Jed}(\mathcal{V}, m, d)$ i $G_s(\mathcal{V}, m, \mathcal{P}_{(m)})$. Twierdzenie to jest pierwszym krokiem w kierunku uogólnienia wyników dotyczących spójności $G^{Jed}(\mathcal{V}, m, d)$ na model $G_s^{Jed}(\mathcal{V}, m, d)$.

Zostanie także pokazany wynik dotyczący liczby trójkątów w $G^{Jed}(\mathcal{V}, m, d)$.
Będzie on obrazował jak różnią się zależności między połączeniami w grafach
 $G^{Jed}(\mathcal{V}, m, d)$ i $G(n, p)$.

3. Wyniki pomocnicze

3.1. Nierówności

3.1.1. Nierówności Markowa i Czebyszewa

W celu oszacowania prawdopodobieństwa odchylenia zmiennej losowej od wartości oczekiwanej będziemy wykorzystywać nierówności Markowa i Czebyszewa. Nierówności te można znaleźć w każdym klasycznym podręczniku do rachunku prawdopodobieństwa (w książce [31] autorzy nazywają je nierównościami Czebyszewa i Czebyszewa–Bienaymé, odpowiednio). Jako pierwszą podajemy nierówność Markowa.

Lemat 3.1. *Niech X będzie nieujemną zmienną losową. Wtedy dla każdego $\varepsilon > 0$*

$$\Pr\{X \geq \varepsilon\} \leq \frac{\mathbb{E}X}{\varepsilon}.$$

Przydatne też będzie uogólnienie powyższej nierówności będące wersją twierdzenia 5.7.5 z [31].

Lemat 3.2. *Niech X będzie nieujemną zmienną losową oraz $a \in (0; +\infty)$.*

$$\Pr\{X \geq a\} \leq \frac{\mathbb{E}(X\mathbb{I}_{\geq a}(X))}{a},$$

gdzie

$$\mathbb{I}_{\geq a}(x) = \begin{cases} 0 & \text{dla } x < a; \\ 1 & \text{dla } x \geq a. \end{cases}$$

Dowód.

$$a \Pr\{X \geq a\} \leq \int_{X \geq a} \mathbb{I}_{\geq a}(X) X dP = \mathbb{E}(X\mathbb{I}_{\geq a}(X)).$$

□

Nierówność Czebyszewa daje dokładniejsze ograniczenie górne na prawdopodobieństwo odchylenia od wartości oczekiwanej.

Lemat 3.3. *Niech X będzie zmienną losową i $\varepsilon > 0$, wtedy*

$$\Pr\{|X - \mathbb{E}X| \geq \varepsilon\} \leq \frac{\text{Var}X}{\varepsilon^2}.$$

3.1.2. Nierówność Chernoffa

Zmienna losowa o rozkładzie dwumianowym charakteryzuje się silniejszym skoncentrowaniem wokół wartości oczekiwanej niż wynikałoby to z nierówności Czebyszewa. Przetawione poniżej nierówności Chernoffa dają o wiele lepsze oszacowanie na prawdopodobieństwo odchylenia od wartości oczekiwanej. Dowody tych nierówności można znaleźć na przykład w [32] (patrz twierdzenie 2.1 i wniosek 2.2).

Lemat 3.4. *Niech X będzie zmienną losową o rozkładzie dwumianowym oraz niech $t > 0$, wtedy*

$$(3.1) \quad \Pr\{X \leq \mathbb{E}X - t\} \leq \exp\left(-\frac{t^2}{2\mathbb{E}X}\right),$$

$$(3.2) \quad \Pr\{X \geq \mathbb{E}X + t\} \leq \exp\left(-\frac{3t^2}{2(3\mathbb{E}X + t)}\right) \leq \\ \leq \exp\left(-\frac{t^2}{2\mathbb{E}X} + \frac{t^3}{6\mathbb{E}X^2}\right).$$

3.2. Metody probabilistyczne

3.2.1. Metoda pierwszego i drugiego momentu

Klasycznymi metodami dowodowymi stosowanymi w celu określenia funkcji progowych własności w grafach losowych są metody pierwszego i drugiego momentu. Przedstawione one są w wielu monografiach dotyczących metod probabilistycznych w kombinatoryce, na przykład w [23, 32, 54]. Metody te są bezpośrednim zastosowaniem nierówności przedstawionych w lematkach 3.1 i 3.3 do zmiennych losowych o wartościach całkowitych nieujemnych.

Lemat 3.5. *Niech X_n będzie zmienną losową o wartościach całkowitych nieujemnych, wtedy*

- (i) jeśli $\mathbb{E}X_n \rightarrow 0$, to $\Pr\{X_n = 0\} \rightarrow 1$;
- (ii) jeśli $\mathbb{E}X_n \rightarrow \infty$ i $\text{Var}X_n = o((\mathbb{E}X_n)^2)$,
to $\Pr\{X_n \geq 1\} \rightarrow 1$ i $\forall_{\varepsilon > 0} \Pr\{(1 - \varepsilon)\mathbb{E}X \leq X \leq (1 + \varepsilon)\mathbb{E}X\} \rightarrow 1$.

3.2.2. Metoda momentów

Metoda momentów służy do określania, czy dany ciąg zmiennych losowych dąży do pewnego rozkładu. Można ją stosować w przypadku, gdy rozkład graniczny jest jednoznacznie wyznaczony przez swoje momenty. Klasyczny wynik (patrz na przykład twierdzenie 4.4.5 w [17]) zwany metodą momentów odnosi się do zwykłych momentów zmiennej losowej.

Twierdzenie 3.1. *Niech X będzie zmienną losową z dystrybucją wyznaczoną jednoznacznie przez swoje momenty. Jeśli X_1, X_2, \dots są zmiennymi losowymi o skończonych momentach takich, że $\mathbb{E}X_n^k \rightarrow \mathbb{E}X^k$ przy $n \rightarrow \infty$ dla dowolnego k naturalnego, wtedy X_n jest słabo zbieżna (zbieżna według rozkładu) do X .*

W rozprawie będzie wykorzystany wniosek z tego twierdzenia dotyczący zbieżności do rozkładu Poissona oraz momentów silniowych zmiennej losowej będącej sumą zmiennych losowych zero-jedynkowych.

Momentami silniowymi zmiennej losowej X o skończonych momentach nazywamy wyrażenia.

$$\mathbb{E}(X)_k = \mathbb{E}[X(X-1)\dots(X-k+1)] \text{ dla } k \geq 1 \text{ oraz } \mathbb{E}(X)_0 = 1.$$

W przypadku, gdy zmienna losowa X_n przelicza obiekty można ją zapisać w formie $X_n = \sum_{i=1}^n \xi_i$, gdzie ξ_i , $1 \leq i \leq n$, są zmiennymi zero-jedynkowych. Dla tak zdefiniowanej zmiennej losowej

$$(X_n)_k = \left(\sum_{i=1}^n \xi_i\right)_k = \sum_{i_1, \dots, i_k}^* \xi_{i_1} \xi_{i_2} \dots \xi_{i_k},$$

gdzie \sum_{i_1, \dots, i_k}^* oznacza sumę po wszystkich ciągach różnych indeksów i_1, \dots, i_k oraz

$$\mathbb{E}(X_n)_k = \sum_{i_1, \dots, i_k}^* \Pr\{\xi_{i_1} = 1, \dots, \xi_{i_k} = 1\}.$$

Poniższe twierdzenie jest to praktyczny wniosek z twierdzenia 3.1 (w [32] jest to wniosek 6.8).

Twierdzenie 3.2. *Niech X_n będzie sumą n zero-jedynkowych zmiennych losowych ξ_i (tzn. $X_n = \sum_{i=1}^n \xi_i$). Jeśli dla pewnego $\lambda > 0$*

$$\mathbb{E}(X_n)_k = \sum_{i_1, \dots, i_k}^* \Pr\{\xi_{i_1} = 1, \dots, \xi_{i_k} = 1\} \rightarrow \lambda^k$$

przy $n \rightarrow \infty$, to X_n jest słabo zbieżna (zbieżna według rozkładu) do zmiennej losowej o rozkładzie Poissona $Po(\lambda)$.

3.2.3. Metoda Steina

Metoda Steina pozwala określić zbieżność ciągu zmiennych losowych X_n do zmiennej losowej X poprzez oszacowanie odległości między nimi. Odległość między dwiema zmiennymi losowymi X i Y o wartościach w \mathbb{R} wprowadzimy za pomocą **normy całkowitego wahanía** (ang. *total variation distance*)

$$d_{TV}(X, Y) = \max_A |\Pr\{X \in A\} - \Pr\{Y \in A\}|,$$

gdzie maksimum jest wzięte po wszystkich zbiorach borelowskich $A \subseteq \mathbb{R}$. W przypadku, gdy X i Y przyjmują wartości całkowite wyżej określona odległość dana jest wzorem

$$(3.3) \quad d_{TV}(X, Y) = \frac{1}{2} \sum_{k=0}^{\infty} |\Pr\{X = k\} - \Pr\{Y = k\}|.$$

W rozprawie będziemy wykorzystywać metodę Steina w wersji z grafem zależności przedstawioną między innymi w [32]. Niech Γ będzie zbiorem indeksów rodziny zmiennych losowych indykatorowych $\{I_\beta : \beta \in \Gamma\}$. Załóżmy,

że $W = \sum_{\beta \in \Gamma} I_\beta$. **Grafem zależności** L nazywamy graf o zbiorze wierzchołków Γ , taki, że jeśli dla dowolnych podzbiorów $A \subseteq \Gamma$ i $B \subseteq \Gamma$ nie ma krawędzi między A i B w grafie L , to zmienne losowe ze zbioru $\{I_\beta : \beta \in A\}$ są niezależne od zmiennych losowych ze zbioru $\{I_\gamma : \gamma \in B\}$. Niech T_λ będzie zmienną losową o rozkładzie Poissona $\text{Po}(\lambda)$, gdzie $\lambda = \mathbb{E}W$ oraz niech $\pi_\beta = \mathbb{E}I_\beta$ dla $\beta \in \Gamma$. Z twierdzenia 6.23 z [32], mamy

$$(3.4) \quad d_{TV}(W, T_\lambda) \leq \min(\lambda^{-1}, 1) \left(\sum_{\beta \in V(L)} \pi_\beta^2 + \sum_{\beta\gamma \in E(L)} (\mathbb{E}(I_\beta I_\gamma) + \pi_\beta \pi_\gamma) \right),$$

gdzie suma $\sum_{\beta\gamma \in E(L)}$ oznacza sumowanie po wszystkich uporządkowanych parach (β, γ) takich, że $\beta\gamma \in E(L)$.

Łatwo sprawdzić, że jeśli dla ciągu zmiennych losowych X_n i zmiennej losowej X odległość $d_{TV}(X_n, X)$ dąży do zera, to X_n dąży według rozkładu do X . W dodatku jeśli Y_n jest zmienną losową o rozkładzie Poissona $\text{Po}(\mathbb{E}X_n)$ oraz $\mathbb{E}X_n \rightarrow \lambda$, to jeśli $d_{TV}(X_n, Y_n)$ dąży do zera, to X_n dąży według rozkładu do zmiennej losowej o rozkładzie Poissona $\text{Po}(\lambda)$.

3.3. Oszacowania na prawdopodobieństwa powstania krawędzi

W tym podrozdziale podane zostaną proste fakty dotyczące prawdopodobieństwa powstania krawędzi w różnych modelach losowych grafów przecięć. Rozważmy najpierw model losowego grafu przecięć $G_s(\mathcal{V}, m, \mathcal{P}_{(m)})$ (patrz definicja 1.3). Załóżmy, że dla danych dwóch wierzchołków v i w znamy moc zbiorów $D(v)$ i $D(w)$.

Lemat 3.6. *Niech $s \geq 1$ będzie stałą, $s \leq d \leq \Delta$, $s \leq d' \leq \Delta$ oraz niech $\Delta^2/m = o(1)$. Wtedy*

$$\begin{aligned} \Pr \left\{ vw \in E(G_s(\mathcal{V}, m, \mathcal{P}_{(m)})) \mid |D(v)| = d, |D(w)| = d' \right\} &= \\ &= \frac{(d)_s (d')_s}{m^s s!} \left(1 + O\left(\frac{\Delta^2}{m}\right) \right), \end{aligned}$$

Dowód. Załóżmy, że mamy dane $D(v) \subseteq \mathcal{W}$, $|D(v)| = d$. Niech A będzie rodziną podzbiorów zbioru \mathcal{W} o dokładnie d' elementach, które przecinają zbiór $D(v)$ na co najmniej s elementach. Wtedy

$$\binom{d}{s} \binom{m-d}{d'-s} \leq |A| \leq \binom{d}{s} \binom{m-s}{d'-s}.$$

Zauważmy, że dla $K = s$ lub $K = d$

$$\frac{\binom{m-K}{d'-s}}{\binom{m}{d'}} = \frac{(m-K)_{d'-s} d'^!}{(m)_{d'} (d'-s)!} = \frac{(d')_s (m-K)_{d'-s}}{m^s} \frac{m^{d'}}{m^{d'-s}} \frac{1}{(m)_{d'}}.$$

Co więcej, ponieważ dla takich K

$$1 - \frac{2\Delta^2}{m} \leq \left(1 - \frac{K+d'}{m} \right)^{d'} \leq \frac{(m-K)_{d'-s}}{m^{d'-s}} \leq 1$$

oraz

$$1 - \frac{\Delta^2}{m} \leq \left(1 - \frac{d'}{m}\right)^{d'} \leq \frac{(m)_{d'}}{m^{d'}} \leq 1,$$

więc

$$\begin{aligned} \Pr \left\{ vw \in E(G_s(\mathcal{V}, m, \mathcal{P}_{(m)})) \mid |D(v)| = d, |D(w)| = d' \right\} &= \\ &= \frac{|A|}{\binom{m}{d'}} = \frac{(d)_s (d')_s}{m^s s!} \left(1 + O\left(\frac{\Delta^2}{m}\right)\right). \end{aligned}$$

□

Dla przypadku, gdy $s = 1$ podamy wynik dokładniejszy. Jeżeli oznaczymy

$$\begin{aligned} \Pr_{dd'} &= \Pr \left\{ vw \in E(G(\mathcal{V}, m, \mathcal{P}_{(m)})) \mid |D(v)| = d, |D(w)| = d' \right\}; \\ (3.5) \quad \Pr_d &= \Pr \left\{ vw \in E(G(\mathcal{V}, m, \mathcal{P}_{(m)})) \mid |D(v)| = d \right\}, \end{aligned}$$

to prawdziwy jest poniższy lemat.

Lemat 3.7.

$$\begin{aligned} \frac{dd'}{m} - \frac{(dd')^2}{m^2} &\leq \Pr_{dd'} \leq \frac{dd'}{m}; \\ \frac{d \mathbb{E}(Z_n)}{m} - \frac{d^2 \mathbb{E}(Z_n^2)}{m^2} &\leq \Pr_d \leq \frac{d \mathbb{E}(Z_n)}{m}. \end{aligned}$$

Dowód.

$$\begin{aligned} \Pr \left\{ D(v) \cap D(w) \neq \emptyset \mid |D(v)| = d, |D(w)| = d' \right\} &\leq \\ &\leq \frac{d \binom{m-1}{d'-1}}{\binom{m}{d'}} = \frac{dd'}{m}; \\ \Pr \left\{ D(v) \cap D(w) = \emptyset \mid |D(v)| = d, |D(w)| = d' \right\} &= \\ &= \frac{\binom{m-d}{d'}}{\binom{m}{d}} = \prod_{i=0}^{d'-1} \left(1 - \frac{d}{m-i}\right) \leq \left(1 - \frac{d}{m}\right)^{d'} \leq \\ &\leq 1 - \frac{dd'}{m} + \binom{d'}{2} \left(\frac{d}{m}\right)^2 < 1 - \frac{dd'}{m} + \left(\frac{dd'}{m}\right)^2. \end{aligned}$$

Druga nierówność w lemacie wynika z pierwszej i faktu, że

$$\Pr_d = \sum_{d'} \Pr_{dd'} \Pr\{|D(w)| = d'\}.$$

□

3.4. Prawdopodobieństwo warunkowe

Wykorzystamy też prosty lemat dotyczący prawdopodobieństw warunkowych.

Lemat 3.8. Niech $(A_i)_{i \geq i_0}$ i $(B_i)_{i \geq i_0}$ będą ciągami zdarzeń takimi, że zachodzi $A_i \cap B_i = \emptyset$ dla wszystkich $i \geq i_0$. Dodatkowo niech

$$\Pr \{A_{i_0} \cup B_{i_0}\} \geq 1 - p_0$$

i

$$\Pr \left\{ A_i \cup B_i \mid \bigcap_{j=i_0}^{i-1} B_j \right\} \geq 1 - q,$$

dla $p_0, q \in [0; 1]$ i dla każdego $i > i_0$. Wtedy dla każdego $k \geq i_0$

$$\Pr \left\{ \bigcup_{i=i_0}^k A_i \cup \bigcap_{i=i_0}^k B_i \right\} \geq (1 - p_0)(1 - q)^{k-i_0}.$$

Dowód. Zauważmy, że

$$\bigcup_{i=i_0}^k A_i \cup \bigcap_{i=i_0}^k B_i \supseteq \bigcap_{i=i_0}^k B_i \cup \bigcup_{i=i_0+1}^k \left(A_i \cap \bigcap_{j=i_0}^{i-1} B_j \right) \cup A_{i_0}$$

i zdarzenia $\bigcap_{i=i_0}^k B_i$, $(A_i \cap \bigcap_{j=i_0}^{i-1} B_j)$ i A_{i_0} są parami rozłączne (gdyż A_i i B_i są rozłączne dla każdego $i \geq i_0$). Dlatego wystarczy pokazać, że

$$(3.6) \quad \Pr \left\{ \bigcap_{i=i_0}^k B_i \right\} + \sum_{i=i_0+1}^k \Pr \left\{ A_i \cap \bigcap_{j=i_0}^{i-1} B_j \right\} + \Pr \{A_{i_0}\} \geq \\ \geq (1 - p_0)(1 - q)^{k-i_0}$$

dla każdego $k > i_0$. Udowodnimy (3.6) stosując indukcję względem k .

Niech $k = i_0$, wtedy (3.6) wynika bezpośrednio z założeń lematu. Teraz załóżmy, że (3.6) jest spełnione dla $k \geq i_0$. Wtedy

$$\begin{aligned} & \Pr \left\{ \bigcap_{i=i_0}^{k+1} B_i \right\} + \sum_{i=i_0+1}^{k+1} \Pr \left\{ A_i \cap \bigcap_{j=i_0}^{i-1} B_j \right\} + \Pr \{A_{i_0}\} = \\ & = \Pr \left\{ \bigcap_{i=i_0}^{k+1} B_i \right\} + \Pr \left\{ A_{k+1} \cap \bigcap_{i=i_0}^k B_i \right\} + \\ & \quad + \sum_{i=i_0+1}^k \Pr \left\{ A_i \cap \bigcap_{j=i_0}^{i-1} B_j \right\} + \Pr \{A_{i_0}\} = \\ & = \Pr \left\{ A_{k+1} \cup B_{k+1} \mid \bigcap_{i=i_0}^k B_i \right\} \Pr \left\{ \bigcap_{i=i_0}^k B_i \right\} + \\ & \quad + \sum_{i=i_0+1}^k \Pr \left\{ A_i \cap \bigcap_{j=i_0}^{i-1} B_j \right\} + \Pr \{A_{i_0}\} \geq \\ & \geq \Pr \left\{ A_{k+1} \cup B_{k+1} \mid \bigcap_{i=i_0}^k B_i \right\} \cdot \\ & \quad \cdot \left(\Pr \left\{ \bigcap_{i=i_0}^k B_i \right\} + \sum_{i=i_0+1}^k \Pr \left\{ A_i \cap \bigcap_{j=i_0}^{i-1} B_j \right\} + \Pr \{A_{i_0}\} \right) \geq \\ & \geq (1 - p_0)(1 - q)^{k+1-i_0}. \end{aligned}$$

□

4. Równoważność modeli

Wyniki zawarte w tym rozdziale bazują na obserwacji, że dla bliskich sobie rozkładów $\mathcal{P}_{(m)}$ i $\mathcal{P}'_{(m)}$ grafy $G_s(\mathcal{V}, m, \mathcal{P}_{(m)})$ i $G_s(\mathcal{V}, m, \mathcal{P}'_{(m)})$ powinny mieć asymptotycznie podobne własności. Przedstawione rezultaty pozwalają między innymi w prosty sposób wyciągać wnioski dotyczące ogólnych rozkładów, z wyników dotyczących szczególnych rozkładów $\mathcal{P}_{(m)}$. Podobne wyniki związane z równoważnością modeli grafów losowych pojawiły się już w kontekście klasycznych modeli grafów losowych $G(n, p)$ i $G(n, M)$ (patrz [13, 43]) oraz losowych grafów skierowanych (patrz [33, 35, 37]).

4.1. Równoważność losowych grafów przecięć.

Dla danych dwóch rozkładów prawdopodobieństwa $\mathcal{P}_{(m)} = (P_0, \dots, P_m)$ i $\mathcal{P}'_{(m)} = (P'_0, \dots, P'_m)$ oznaczmy

$$(4.1) \quad w(\mathcal{P}'_{(m)} \prec \mathcal{P}_{(m)}) = \\ = \max\{\Pr\{Z'_n \leq Z_n\} : Z_n \text{ ma rozkład } \mathcal{P}_{(m)}, Z'_n \text{ ma rozkład } \mathcal{P}'_{(m)}\},$$

gdzie maksimum jest wzięte po wszystkich wektorach losowych (Z_n, Z'_n) z rozkładami brzegowymi $\mathcal{P}_{(m)}$ i $\mathcal{P}'_{(m)}$.

Zauważmy, że $w(\mathcal{P}'_{(m)} \prec \mathcal{P}_{(m)}) = 1$, gdy $\sum_{j \geq k} P'_j \leq \sum_{j \geq k} P_j$, dla każdego $k = 0, 1, \dots, m$. To znaczy $w(\mathcal{P}'_{(m)} \prec \mathcal{P}_{(m)}) = 1$, gdy rozkład $\mathcal{P}'_{(m)}$ jest stochastycznie zdominowany przez rozkład $\mathcal{P}_{(m)}$, co, tak jak w przypadku zmiennych losowych, oznaczać będziemy przez $\mathcal{P}'_{(m)} \prec \mathcal{P}_{(m)}$.

Lemat 4.1. *Dla dowolnej własności: rosnącej \mathcal{A} i malejącej \mathcal{B} oraz rozkładów prawdopodobieństwa $\mathcal{P}'_{(m)} \prec \mathcal{P}_{(m)}$, mamy*

$$(4.2) \quad \Pr\{G_s(\mathcal{V}, m, \mathcal{P}'_{(m)}) \in \mathcal{A}\} \leq \Pr\{G_s(\mathcal{V}, m, \mathcal{P}_{(m)}) \in \mathcal{A}\}, \\ \Pr\{G_s(\mathcal{V}, m, \mathcal{P}'_{(m)}) \in \mathcal{B}\} \geq \Pr\{G_s(\mathcal{V}, m, \mathcal{P}_{(m)}) \in \mathcal{B}\}.$$

Dodatkowo, dla dowolnych rozkładów $\mathcal{P}_{(m)}$ i $\mathcal{P}'_{(m)}$ mamy

$$(4.3) \quad \Pr\{G_s(\mathcal{V}, m, \mathcal{P}'_{(m)}) \in \mathcal{A}\} \\ \leq \Pr\{G_s(\mathcal{V}, m, \mathcal{P}_{(m)}) \in \mathcal{A}\} + 1 - w^n(\mathcal{P}'_{(m)} \prec \mathcal{P}_{(m)}), \\ \Pr\{G_s(\mathcal{V}, m, \mathcal{P}_{(m)}) \in \mathcal{B}\} \\ \leq \Pr\{G_s(\mathcal{V}, m, \mathcal{P}'_{(m)}) \in \mathcal{B}\} + 1 - w^n(\mathcal{P}'_{(m)} \prec \mathcal{P}_{(m)}),$$

gdzie $w(\mathcal{P}'_{(m)} \prec \mathcal{P}_{(m)})$ jest dana wzorem (4.1).

Dowód. Wystarczy udowodnić nierówności dla własności rosnących. Dowody dla własności malejących są analogiczne.

Aby udowodnić nierówność (4.2) skonstruujemy najpierw niezależne pary zbiorów losowych $(D(v_1), D(v_1)'), \dots, (D(v_n), D(v_n)'),$ których moce mają te same rozkłady oraz dla każdego i zachodzi $D(v_i)' \subseteq D(v_i)$. W dodatku rozkłady brzegowe tych par zmiennych będą takie, że moce zbiorów $|D(v_i)|$ i $|D(v_i)'|$ będą miały rozkłady $\mathcal{P}_{(m)}$ i $\mathcal{P}'_{(m)}$, odpowiednio. Zauważmy, że otrzymane w ten sposób losowe grafy $G_s(\mathcal{V}, m, \mathcal{P}_{(m)})$ i $G_s(\mathcal{V}, m, \mathcal{P}'_{(m)})$ spełniają zależność $E(G_s(\mathcal{V}, m, \mathcal{P}'_{(m)})) \subseteq E(G_s(\mathcal{V}, m, \mathcal{P}_{(m)}))$. Dlatego dla własności rosnącej \mathcal{A} , jeśli $G_s(\mathcal{V}, m, \mathcal{P}'_{(m)}) \in \mathcal{A}$ to $G_s(\mathcal{V}, m, \mathcal{P}_{(m)}) \in \mathcal{A}$. Stąd wynika (4.2).

W konstrukcji wyżej wspomnianych niezależnych par zmiennych losowych zastosujemy standardowy coupling. Przez coupling zbiorów losowych D i D' (jak także innych obiektów losowych) rozumiemy tutaj określenie wspólnej przestrzeni probabilistycznej na której zdefiniowany jest wektor losowy (D, D') o rozkładach brzegowych odpowiadających rozkładowi brzegowym zbiorów D i D' , odpowiednio. Korzystamy z używanej zwyczajowo w języku polskim angielskojęzycznej nazwy z braku dobrego polskiego odpowiednika.

Niech (Z_n, Z'_n) będzie losowym wektorem o rozkładach brzegowych, odpowiednio, $\mathcal{P}_{(m)}$ i $\mathcal{P}'_{(m)}$, takim, że $Z'_n \leq Z_n$. Mając daną wartość Z_n , wybieramy w sposób jednostajny podzbiór $D(v_1) \subseteq \mathcal{W}$ spośród wszystkich Z_n -elementowych podzbiorów zbioru \mathcal{W} . Mając dany zbiór $D(v_1)$, zdefiniujemy $D(v_1)' = D(v_1) \setminus S$, gdzie zbiór $S \subseteq D(v_1)$ jest losowym podzbiorem zbioru $D(v_1)$ o rozmiarze $Z_n - Z'_n$ wybranym w sposób jednostajny spośród wszystkich $(Z_n - Z'_n)$ -elementowych podzbiorów zbioru $D(v_1)$. Po powtórzeniu, niezależnie, tej konstrukcji n razy otrzymujemy rodzinę niezależnych par zmiennych losowych o wymaganych własnościach.

Dowód nierówności (4.3) przebiega podobnie. Niech (Z_n, Z'_n) będzie wektorem zmiennych losowych, dla którego osiągnięte zostaje maksimum z definicji $w(\mathcal{P}'_{(m)} \prec \mathcal{P}_{(m)})$. Dla danej wartości wektora (Z_n, Z'_n) takiej, że $Z'_n \leq Z_n$, możemy utworzyć parę podzbiorów $(D(v_i), D(v_i)')$ w ten sam sposób jak w dowodzie (4.2). Natomiast dla danej wartości wektora (Z_n, Z'_n) takiej, że $Z'_n > Z_n$, wybieramy $D(v_i)$ i $D(v_i)'$ niezależnie w sposób jednostajny spośród podzbiorów \mathcal{W} o Z_n i Z'_n elementach, odpowiednio. Teraz wystarczy zauważyć, że tak skonstruowany wektor $(|D(v_i)|, |D(v_i)'|)$ ma rozkłady brzegowe $\mathcal{P}_{(m)}$ i $\mathcal{P}'_{(m)}$. Dodatkowo, wiemy, że

$$\Pr\{D(v_i)' \subseteq D(v_i)\} \geq \Pr\{Z'_n \leq Z_n\} = w(\mathcal{P}'_{(m)} \prec \mathcal{P}_{(m)}).$$

Rozważmy teraz losowe grafy przecięć $G_s(\mathcal{V}, m, \mathcal{P}_{(m)})$ i $G_s(\mathcal{V}, m, \mathcal{P}'_{(m)})$ indukowane przez rodziny podzbiorów powstałe w wyniku n niezależnych eksperymentów konstruujących w powyżej opisany sposób pary $(D(v_i), D(v_i)')$. Dla tych grafów zachodzi nierówność $\Pr\{\mathcal{H}\} \geq w^n(\mathcal{P}'_{(m)} \prec \mathcal{P}_{(m)})$, gdzie $\mathcal{H} := \{E(G_s(\mathcal{V}, m, \mathcal{P}'_{(m)})) \subseteq E(G_s(\mathcal{V}, m, \mathcal{P}_{(m)}))\}$. Czyli ostatecznie

$$\begin{aligned} \Pr\{G_s(\mathcal{V}, m, \mathcal{P}'_{(m)}) \in \mathcal{A}\} &\leq \Pr\{\{G_s(\mathcal{V}, m, \mathcal{P}'_{(m)}) \in \mathcal{A}\} \cap \mathcal{H}\} + 1 - \Pr\{\mathcal{H}\} \\ &\leq \Pr\{\{G_s(\mathcal{V}, m, \mathcal{P}_{(m)}) \in \mathcal{A}\} \cap \mathcal{H}\} + 1 - \Pr\{\mathcal{H}\} \\ &\leq \Pr\{G_s(\mathcal{V}, m, \mathcal{P}_{(m)}) \in \mathcal{A}\} + 1 - \Pr\{\mathcal{H}\}. \end{aligned}$$

□

Wniosek 4.1. Niech $G_s(\mathcal{V}, m, \mathcal{P}_{(m)})$ i $G_s(\mathcal{V}, m, \mathcal{P}'_{(m)})$ będą ciągami grafów losowych oraz niech dla $w(\mathcal{P}'_{(m)} \prec \mathcal{P}_{(m)})$ danego wzorem (4.1)

$$1 - w(\mathcal{P}'_{(m)} \prec \mathcal{P}_{(m)}) = o(n^{-1}), \quad 1 - w(\mathcal{P}_{(m)} \prec \mathcal{P}'_{(m)}) = o(n^{-1}).$$

Wtedy dla ciągu własności monotonicznych $\mathcal{A}_n \in \mathcal{G}_n$ z (4.3) wynika, że

$$(4.4) \quad \Pr\{G_s(\mathcal{V}, m, \mathcal{P}_{(m)}) \in \mathcal{A}_n\} = \Pr\{G_s(\mathcal{V}, m, \mathcal{P}'_{(m)}) \in \mathcal{A}_n\} + o(1).$$

4.2. Równoważność modelu jednostajnego i dwumianowego

Poniższy lemat jest przydatnym narzędziem, dzięki któremu łatwo można z wyników uzyskanych dla grafu $G_s^{Bin}(\mathcal{V}, m, p)$ wywnioskować wyniki dotyczące grafu $G_s^{Jed}(\mathcal{V}, m, d)$ i na odwrót.

Lemat 4.2. Niech $s \geq 1$ oraz niech $\mathcal{A}_n \subseteq \mathcal{G}_n$ i $\mathcal{B}_n \subseteq \mathcal{G}_n$ będą, odpowiednio, ciągiem własności rosnących i malejących.

(i) Załóżmy, że $\ln n = o(d)$, $p_- = (d - \sqrt{3d \ln n})/m$ i $p_+ = (d + \sqrt{3d \ln n})/m$.

Wtedy

$$\begin{aligned} \Pr\{G_s^{Bin}(\mathcal{V}, m, p_-) \in \mathcal{A}_n\} - o(1) &\leq \Pr\{G_s^{Jed}(\mathcal{V}, m, d) \in \mathcal{A}_n\} \leq \\ &\leq \Pr\{G_s^{Bin}(\mathcal{V}, m, p_+) \in \mathcal{A}_n\} + o(1), \\ \Pr\{G_s^{Bin}(\mathcal{V}, m, p_+) \in \mathcal{B}_n\} - o(1) &\leq \Pr\{G_s^{Jed}(\mathcal{V}, m, d) \in \mathcal{B}_n\} \leq \\ &\leq \Pr\{G_s^{Bin}(\mathcal{V}, m, p_-) \in \mathcal{B}_n\} + o(1). \end{aligned}$$

(ii) Załóżmy, że $d_+ = mp + \sqrt{3mp \ln n}$, $d_- = mp - \sqrt{3mp \ln n}$ oraz $\ln n = o(mp)$. Wtedy

$$\begin{aligned} \Pr\{G_s^{Jed}(\mathcal{V}, m, d_-) \in \mathcal{A}_n\} - o(1) &\leq \Pr\{G_s^{Bin}(\mathcal{V}, m, p) \in \mathcal{A}_n\} \leq \\ &\leq \Pr\{G_s^{Jed}(\mathcal{V}, m, d_+) \in \mathcal{A}_n\} + o(1), \\ \Pr\{G_s^{Jed}(\mathcal{V}, m, d_+) \in \mathcal{B}_n\} - o(1) &\leq \Pr\{G_s^{Bin}(\mathcal{V}, m, p) \in \mathcal{B}_n\} \leq \\ &\leq \Pr\{G_s^{Jed}(\mathcal{V}, m, d_-) \in \mathcal{B}_n\} + o(1). \end{aligned}$$

Dowód. Pokażemy tylko pierwszą z powyższych nierówności. Dowody pozostałych są analogiczne. Niech Z_n^1 będzie zmienną losową o rozkładzie dwumianowym $\text{Bin}(m, p_-)$ oraz niech Z_n^2 będzie zmienną losową o rozkładzie zdegenerowanym \mathcal{P}_d skoncentrowanym w d (tzn. $P_d = 1$). Dla wektora losowego (Z_n^1, Z_n^2) mamy

$$\begin{aligned} 1 - w(\text{Bin}(m, p_-) \prec \mathcal{P}_d) &\leq 1 - \Pr\{Z_n^1 \leq Z_n^2\} = 1 - \Pr\{Z_n^1 \leq d\} = \\ &= \Pr\{Z_n^1 > d\} = \Pr\{Z_n^1 > \mathbb{E}Z_n^1 + \sqrt{3d \ln n}\} \leq \\ &\leq \exp\left(-\frac{3d \ln n}{2(d - \frac{2}{3}\sqrt{3d \ln n})}\right) = o\left(\frac{1}{n}\right). \end{aligned}$$

W przedostatnim kroku zastosowaliśmy nierówność Chernoffa (3.2). Nierówność wynika natychmiast z (4.3). \square

Następny lemat będzie dotyczył zbieżności według rozkładu zmiennych losowych liczących obiekty w grafach $G_s^{Bin}(\mathcal{V}, m, p)$ i $G_s^{Jed}(\mathcal{V}, m, d)$. Lemat jest prostym wnioskiem z lematu 4.2, lecz nie został ujęty w [12]. Wspominamy o nim, gdyż w rozprawie będziemy go wykorzystywali w przypadku zbieżności zmiennych losowych określających liczbę kopii podgrafów izomorficznych z kliką na h wierzchołkach oraz liczbę wierzchołków izolowanych.

Lemat 4.3. *Niech $s \geq 1$, X będzie zmienną losową dyskretną o wartościach całkowitych oraz $X_n = X_n(G)$ będzie funkcją liczącą obiekty w grafach z \mathcal{G}_n , taką, że dla dowolnej liczby całkowitej k ciąg $\mathcal{A}_{k,n} = \{G \in \mathcal{G}_n : X_n(G) \leq k\}$ jest ciągiem własności monotonicznych.*

(i) *Niech $\ln n = o(d)$, $p_- = (d - \sqrt{3d \ln n})/m$ i $p_+ = (d + \sqrt{3d \ln n})/m$. Jeśli $X_n(G_s^{Bin}(\mathcal{V}, m, p_-))$ i $X_n(G_s^{Bin}(\mathcal{V}, m, p_+))$ dążą według rozkładu do zmiennej losowej X , to $X_n(G_s^{Jed}(\mathcal{V}, m, d))$ dąży według rozkładu do zmiennej losowej X .*

(ii) *Niech $\ln n = o(mp)$, $d_- = mp - \sqrt{3mp \ln n}$ i $d_+ = mp + \sqrt{3mp \ln n}$. Jeśli $X_n(G_s^{Jed}(\mathcal{V}, m, d_-))$ i $X_n(G_s^{Jed}(\mathcal{V}, m, d_+))$ dążą według rozkładu do zmiennej losowej X , to $X_n(G_s^{Bin}(\mathcal{V}, m, p))$ dąży według rozkładu do zmiennej losowej X .*

Dowód. Pokażemy dowód (i). Dowód (ii) jest analogiczny. Niech k będzie dodatnią liczbą całkowitą. Ze zbieżności według rozkładu

$$\Pr\{G_s^{Bin}(\mathcal{V}, m, p_-) \in \mathcal{A}_{k,n}\} = \Pr\{X_n(G_s^{Bin}(\mathcal{V}, m, p_-)) \leq k\} \rightarrow \Pr\{X \leq k\}$$

oraz

$$\Pr\{G_s^{Bin}(\mathcal{V}, m, p_+) \in \mathcal{A}_{k,n}\} = \Pr\{X_n(G_s^{Bin}(\mathcal{V}, m, p_+)) \leq k\} \rightarrow \Pr\{X \leq k\},$$

czyli z lematu 4.2

$$\Pr\{X_n(G_s^{Jed}(\mathcal{V}, m, d)) \leq k\} = \Pr\{G_s^{Jed}(\mathcal{V}, m, d) \in \mathcal{A}_{k,n}\} \rightarrow \Pr\{X \leq k\}.$$

□

4.3. Proste wnioski dla $G^{Jed}(\mathcal{V}, m, d)$

Prace na temat losowego grafu przecięć $G^{Bin}(\mathcal{V}, m, p)$ pojawiły się zanim wzrosło zainteresowanie innymi modelami losowych grafów przecięć. W związku z tym struktura modelu $G^{Bin}(\mathcal{V}, m, p)$ została o wiele lepiej zbadana niż struktura innych modeli. Wystarczy wspomnieć tu prace [5, 6, 7, 20, 25, 40, 42, 52, 53, 58, 59]. Lemat 4.2 pozwala nam część wyników dotyczących grafu $G^{Jed}(\mathcal{V}, m, d)$ z tych już udowodnionych wywnioskować. Podajemy tu kilka przykładów. Dowody pomijamy, gdyż wystarczy w nich skorzystać z lematu 4.2.

4.3.1. Spójność

Poniższe twierdzenie dotyczące $G^{Bin}(\mathcal{V}, m, p)$ udowodnione zostało w [58].

Twierdzenie 4.1. *Niech $m = n^\alpha$, $\alpha > 1$, $p = \sqrt{(\ln n + \omega)/(nm)}$. Wtedy*

- (i) jeśli $\omega \rightarrow -\infty$, to $G^{\text{Bin}}(\mathcal{V}, m, p)$ jest niespójny asymptotycznie prawie na pewno
- (ii) jeśli $\omega \rightarrow \infty$, to $G^{\text{Bin}}(\mathcal{V}, m, p)$ jest spójny asymptotycznie prawie na pewno

Prostym wnioskiem jest analogiczne twierdzenie dotyczące $G^{\text{Jed}}(\mathcal{V}, m, d)$ dla $\alpha > 1$. Warto wspomnieć, że niezależnie poniższy rezultat uzyskali Backburn i Gerke w [8] bez odniesienia do lematu o równoważności.

Wniosek 4.2. Niech $m = n^\alpha$, $\alpha > 1$, $d = c\sqrt{m \ln n/n}$. Wtedy

- (i) jeśli $c < 0$, to $G^{\text{Jed}}(\mathcal{V}, m, d)$ jest niespójny asymptotycznie prawie na pewno
- (ii) jeśli $c > 0$, to $G^{\text{Jed}}(\mathcal{V}, m, d)$ jest spójny asymptotycznie prawie na pewno

4.3.2. Przejście fazowe

Poniższe twierdzenie zostało udowodnione w [5].

Twierdzenie 4.2. Niech $m = n^\alpha$, $\alpha > 1$, $p = \sqrt{c/nm}$. Wtedy

- (i) jeśli $c < 1$ to asymptotycznie prawie na pewno największa składowa spójności w $G^{\text{Bin}}(\mathcal{V}, m, p)$ ma $O(\ln n)$ wierzchołków;
- (ii) jeśli $c > 1$ to asymptotycznie prawie na pewno największa składowa spójności w $G^{\text{Bin}}(\mathcal{V}, m, p)$ ma $\Omega(n)$ wierzchołków.

Z lematu 4.2 łatwo wynika wniosek dla $G^{\text{Jed}}(\mathcal{V}, m, d)$.

Wniosek 4.3. Niech $m = n^\alpha$, $\alpha > 1$, $d = \sqrt{cm/n}$. Wtedy

- (i) jeśli $c < 1$ to asymptotycznie prawie na pewno największą składową spójności w $G^{\text{Jed}}(\mathcal{V}, m, d)$ ma $O(\ln n)$ wierzchołków;
- (ii) jeśli $c > 1$ to asymptotycznie prawie na pewno największą składową spójności w $G^{\text{Jed}}(\mathcal{V}, m, d)$ ma $\Omega(n)$ wierzchołków.

4.3.3. Krawędzie

W [58] udowodnione zostały proste zależności

Twierdzenie 4.3. Niech $m = n^\alpha$, $\alpha > 0$.

- (i) Niech $\tilde{p} = \tilde{p}(n) = \frac{1}{n\sqrt{m}}$. Wtedy
 - jeśli $p = o(\tilde{p})$, to asymptotycznie prawie na pewno $G^{\text{Bin}}(\mathcal{V}, m, p)$ nie ma krawędzi;
 - jeśli $\tilde{p} = o(p)$, to asymptotycznie prawie na pewno $G^{\text{Bin}}(\mathcal{V}, m, p)$ ma co najmniej jedną krawędź.
- (ii) Niech $p = p(n) = \sqrt{\frac{2 \ln n + \omega}{m}}$. Wtedy
 - jeśli $\omega \rightarrow -\infty$, to asymptotycznie prawie na pewno $G^{\text{Bin}}(\mathcal{V}, m, p)$ nie jest grafem pełnym;
 - jeśli $\omega \rightarrow +\infty$, to asymptotycznie prawie na pewno $G^{\text{Bin}}(\mathcal{V}, m, p)$ jest grafem pełnym.

Z lematu 4.2 i powyższego twierdzenia można wysnuć poniższy wniosek.

Wniosek 4.4. Niech $m = n^\alpha$.

- (i) Niech $\alpha > 2$ i $\tilde{d} = \tilde{d}(n) = \frac{\sqrt{m}}{n}$. Wtedy
- jeśli $d = o(\tilde{d})$, to asymptotycznie prawie na pewno $G^{Jed}(\mathcal{V}, m, d)$ nie ma krawędzi;
 - jeśli $\tilde{d} = o(d)$, to asymptotycznie prawie na pewno $G^{Jed}(\mathcal{V}, m, d)$ ma co najmniej jedną krawędź.
- (ii) Niech $\alpha > 0$ i $d = d(n) = c\sqrt{2m \ln n}$. Wtedy
- jeśli $c < 1$, to asymptotycznie prawie na pewno $G^{Jed}(\mathcal{V}, m, d)$ nie jest grafem pełnym;
 - jeśli $c > 1$, to asymptotycznie prawie na pewno $G^{Jed}(\mathcal{V}, m, d)$ jest grafem pełnym.

4.3.4. Równoważność $G^{Jed}(\mathcal{V}, m, d)$ i $G(n, \hat{p})$

W [25] został pokazany następujący wynik o równoważności modeli grafów $G^{Bin}(\mathcal{V}, m, p)$ i $G(n, \hat{p})$ (patrz wniosek 11 z [25]).

Twierdzenie 4.4. Niech $m = n^\alpha$ oraz $\alpha > 6$. Niech $p = p(n)$ spełnia nierówność

$$\frac{\omega}{n\sqrt{m}} \leq p \leq \sqrt{\frac{2 \ln n - \omega}{m}}$$

dla pewnej funkcji $\omega \rightarrow \infty$ (tzn. $G^{Bin}(\mathcal{V}, m, p)$ asymptotycznie prawie na pewno ma co najmniej jedną krawędź i nie jest grafem pełnym). Niech

$$\hat{p} = 1 - (1 - p_e)^m \quad \text{dla} \quad p_e = \frac{p^2}{(1 - p)^2 + np(1 - p) + \binom{n}{2}p^2}.$$

Wtedy dla dowolnego $a \in [0; 1]$ i dowolnej własności \mathcal{A}

$$\Pr\{G^{Bin}(\mathcal{V}, m, p) \in \mathcal{A}\} \rightarrow a \quad \text{wtedy i tylko wtedy, gdy} \quad \Pr\{G(n, \hat{p}) \in \mathcal{A}\} \rightarrow a.$$

Poniższy prosty wniosek z powyższego twierdzenia i lematu 4.2 pokazuje, że dla $\alpha > 6$ i odpowiednio zdefiniowanej wartości d modele $G(n, p)$ i $G^{Jed}(\mathcal{V}, m, d)$ są równoważne.

Wniosek 4.5. Niech $m = n^\alpha$ oraz $\alpha > 6$. Niech $\tilde{d} = \tilde{d}(n)$ spełnia nierówność

$$\frac{\omega\sqrt{m}}{n} \leq \tilde{d} \leq c\sqrt{2m \ln n}$$

dla pewnej funkcji $\omega \rightarrow \infty$ i stałej $c < 1$. Niech

$$\tilde{p} = 1 - \left(1 - \frac{\tilde{d}}{m}\right)^m.$$

Wtedy dla dowolnej własności monotonicznej \mathcal{A} , \tilde{p} jest funkcją progową własności \mathcal{A} w $G(n, p)$ wtedy i tylko wtedy gdy \tilde{d} jest funkcją progową tej własności w $G^{Jed}(\mathcal{V}, m, d)$.

5. Rozkład stopni wierzchołków

Rozkład stopni wierzchołków jest jedną z istotnych charakterystyk sieci. W tym rozdziale zostanie zaprezentowany wynik dotyczący stopni wierzchołków dla szerokiej klasy losowych grafów przecięć. W szczególności na jego podstawie można podać przykład rozkładu $\mathcal{P}_{(m)}$, dla którego podobnie jak dla sieci złożonych $G(\mathcal{V}, m, \mathcal{P}_{(m)})$ ma rozkład spełniający prawo potęgowe.

Pierwszym wynikiem dotyczącym rozkładu stopni wierzchołków w losowym grafie przecięć były wzory dokładne podane przez Godehardta i Jaworskiego w [26]. Wykorzystując te rezultaty, w tej samej pracy, Godehardt i Jaworski pokazali także, że dla $d = \sqrt{cm/n}$ zmienna losowa określająca stopień wierzchołka w $G^{Jed}(\mathcal{V}, m, d)$ dąży do rozkładu Poissona $Po(c)$. Kolejny wynik o zbieżności do rozkładu Poissona dotyczył grafu $G^{Bin}(\mathcal{V}, m, p)$ i został udowodniony przez Starka w [59]. Stark w tym artykule wykazał, że dla $p = \sqrt{c/nm}$ zmienna losowa określająca stopień danego wierzchołka w $G^{Bin}(\mathcal{V}, m, p)$ dąży według rozkładu do zmiennej losowej:

- (i) o rozkładzie skoncentrowanym w wartości 0, dla $\alpha < 1$;
- (ii) $X_1 + \dots + X_N$, gdzie N, X_1, X_2, \dots są niezależnymi zmiennymi losowymi o rozkładzie Poissona $Po(\sqrt{c})$, dla $\alpha = 1$;
- (iii) o rozkładzie Poissona $Po(c)$, dla $\alpha > 1$.

Wynik ten został uogólniony dla modelu $G^{Bin}(\mathcal{V}, m, W)$ przez Deijfena i Ketsa w [34]. Dodatkowo w [34] Deijfen i Kets podają przykład zmiennej losowej W , dla której stopnie wierzchołków w $G^{Bin}(\mathcal{V}, m, W)$ będą miały rozkład graniczny spełniający prawo potęgowe. Inne uogólnienie powyższego wyniku zostało zaproponowane przez Jaworskiego, Karońskiego i Starka w [34]. Autorzy pracy [34] poszukują rozkładów $\mathcal{P}_{(m)}$ dla których stopnie wierzchołków w $G(\mathcal{V}, m, \mathcal{P}_{(m)})$ mają graniczny rozkład Poissona. Ostatecznie problem zbieżności rozkładu stopnia w losowym grafie przecięć do rozkładu Poissona został rozwiązany przez Jaworskiego i Starka w [38].

Następnym rezultatem dotyczącym stopni wierzchołków w $G(\mathcal{V}, m, \mathcal{P}_{(m)})$ jest wynik udowodniony przez Bloznelisa w [11]. Twierdzenie zaproponowane w tym rozdziale jest wzmocnieniem twierdzenia z [11]. Zaproponowany w tym rozdziale dowód pozwala pominąć założenie $n = o(m)$ niezbędne w oryginalnym dowodzie z [11]. W kolejnym rozdziale udowodnimy, że ujęty w poniższym twierdzeniu przypadek $n \asymp m$ ma duże znaczenie w zastosowaniach ze względu na związany z nim współczynnik skupienia.

5.1. Twierdzenie o stopniach wierzchołków

W rozdziale tym będziemy wykorzystywać dodatkowe oznaczenia: przez $X_n = X_n(v) = \deg(v)$ będziemy oznaczać zmienną losową określającą stopień

wierzchołku $v \in \mathcal{V}$ w $G(\mathcal{V}, m, \mathcal{P}_{(m)})$, T_λ będzie oznaczać zmienną losową o rozkładzie Poissona z parametrem λ ($\text{Po}(\lambda)$) oraz przez $\mathbb{T}_X = \mathbb{E}(T_X)$ będziemy oznaczać zmienną losową o złożonym rozkładzie Poissona, tzn. zmienną losową o rozkładzie Poissona $\text{Po}(X)$, gdzie X jest zmienną losową o dodatnich wartościach rzeczywistych.

Z definicji złożonego rozkładu Poissona wynika, że dla zmiennej losowej Y o dodatnich wartościach rzeczywistych

$$(5.1) \quad \Pr\{\mathbb{T}_{Y\mathbb{E}Y} = k\} = \mathbb{E}\left(\frac{(Y\mathbb{E}Y)^k e^{-Y\mathbb{E}Y}}{k!}\right).$$

W [11] autor zauważa, że jeśli podstawimy za Y zmienną losową o ciągłym rozkładzie spełniającym prawo potęgowe, to zmienna losowa $\mathbb{T}_{Y\mathbb{E}Y}$ też spełnia prawo potęgowe. Dlatego poniższe twierdzenie pokazuje, że dla odpowiednio dobranego rozkładu $\mathcal{P}_{(m)}$ graf $G(\mathcal{V}, m, \mathcal{P}_{(m)})$ jest dobrym modelem teoretycznym sieci bezskalowych.

Twierdzenie 5.1. *Niech X_n będzie zmienną losową określającą stopień wierzchołka w $G(\mathcal{V}, m, \mathcal{P}_{(m)})$, Z_n będzie zmienną losową o rozkładzie $\mathcal{P}_{(m)}$ oraz $Y_n = \frac{Z_n}{r_n}$ dla $r_n = \sqrt{\frac{m}{n}}$. Jeśli*

- $Y_n \xrightarrow{D} Y$, Y_n dąży według rozkładu do pewnej zmiennej losowej Y ;
- $\mathbb{E}Y_n \rightarrow \mathbb{E}Y$ przy $n \rightarrow \infty$;
- $\mathbb{E}Y_n < \infty$ i $\mathbb{E}Y < \infty$,

to X_n dąży według rozkładu do zmiennej losowej o złożonym rozkładzie Poissona $\mathbb{T}_{Y\mathbb{E}Y}$.

5.2. Wyniki pomocnicze

Dowód zaczniemy od przedstawienia kilku prostych faktów. Poniższy fakt jest prostą własnością zmiennych losowych zbieżnych według rozkładu (patrz np. zadanie 6 do rozdziału 8.2 z [31]).

Fakt 5.1. *Jeśli $\bar{Y}_n \xrightarrow{D} Y$ oraz $\mathbb{E}\bar{Y}_n \rightarrow \mathbb{E}Y$, to $\bar{Y}_n \mathbb{E}\bar{Y}_n \xrightarrow{D} Y\mathbb{E}Y$.*

Dla dowolnych dwóch zmiennych losowych X i Y definiujemy odległość $d_{TV}(X, Y)$ jak w podrozdziale 3.2.3. Kolejne lematy dotyczą rozkładu Poissona. Dowód pierwszego z nich można znaleźć w [3] (patrz wzór (1.23) str.8).

Lemat 5.1. *Jeśli X i Y są zmiennymi losowymi o rozkładzie dwumiamowym $\text{Bin}(N, p)$ i Poissona $\text{Po}(Np)$, odpowiednio, to $d_{TV}(X, Y) \leq p$.*

Kolejny lemat wynika z rozważań zawartych w [3] (patrz rozdział 1).

Lemat 5.2. *Jeśli X i Y są zmiennymi losowymi o rozkładzie Poissona $\text{Po}(\lambda)$ i $\text{Po}(\lambda + \varepsilon)$, odpowiednio, to $d_{TV}(X, Y) \leq \varepsilon$.*

Dowód. Z [3] (patrz wzór (1.9) str.5) dla dowolnej funkcji ograniczonej $g : \mathbb{N} \rightarrow \mathbb{R}$:

$$\mathbb{E}((\lambda + \varepsilon)g(Y + 1) - Yg(Y)) = 0$$

Dla dowolnego $A \subseteq \mathbb{N}$ zdefiniujemy $g_A : \mathbb{N} \rightarrow \mathbb{R}$ jak w [3] (patrz (1.10) str.6), czyli

$$\begin{cases} g_A(0) = 0; \\ \lambda g_A(j+1) - j g_A(j) = \mathbb{I}_A(j) - \Pr\{X \in A\}; \quad \text{dla } j \geq 0, \end{cases}$$

gdzie \mathbb{I}_A jest funkcją charakterystyczną zbioru A . Wtedy z [3] (patrz lemat 1.1.1)

$$\sup_{j \in \mathbb{N}} \sup_{A \subseteq \mathbb{N}} |g_A(j)| \leq 1.$$

Z definicji g_A jest ograniczona oraz

$$\begin{aligned} |\Pr\{Y \in A\} - \Pr\{X \in A\}| &= \\ &= |\mathbb{E}\mathbb{I}_A(Y) - \Pr\{X \in A\}| = \\ &= |\mathbb{E}(\mathbb{I}_A(Y) - \Pr\{X \in A\})| = \\ &= |\mathbb{E}(\lambda g_A(Y+1) - Y g_A(Y))| = \\ &= |\mathbb{E}((\lambda + \varepsilon)g_A(Y+1) - Y g_A(Y)) - \mathbb{E}(\varepsilon g_A(Y+1))| = \\ &= \varepsilon |\mathbb{E}(g_A(Y+1))| \leq \varepsilon. \end{aligned}$$

□

Następny lemat jest pierwszym większym krokiem w kierunku udowodnienia twierdzenia 5.1.

Lemat 5.3. Niech \bar{Z}_n będzie zmienną losową o rozkładzie $\overline{\mathcal{P}}_{(m)}$, $r_n = \sqrt{\frac{m}{n}}$, $\bar{X}_n = \bar{X}_n(v)$ będzie zmienną losową określającą stopień wierzchołka $v \in \mathcal{V}$ w $G(\mathcal{V}, m, \overline{\mathcal{P}}_{(m)})$, wtedy

$$d_{TV}(\bar{X}_n, \mathbb{T}_{\bar{Z}_n \mathbb{E}\bar{Z}_n / r_n^2}) \leq \varepsilon(n),$$

gdzie $\varepsilon(n) = \frac{2(\mathbb{E}(\bar{Z}_n))^2}{m} + \frac{(\mathbb{E}(\bar{Z}_n^2))^2}{m^2}$.

Dowód. Niech $\bar{X}_n^d = \bar{X}_n^d(v)$ będzie zmienną losową określającą stopień wierzchołka $v \in \mathcal{V}$ w $G(\mathcal{V}, m, \overline{\mathcal{P}}_{(m)})$ pod warunkiem, że $|D_n(v)| = d$. Z niezależności wyborów zbiorów własności wynika, że zmienna losowa \bar{X}_n^d ma rozkład dwumianowy $\text{Bin}(n-1, \text{Pr}_d)$ (gdzie Pr_d jest określone jak w (3.5)). Podstawmy $\lambda_d = d\mathbb{E}\bar{Z}_n/r_n^2$, wtedy z lematu 3.7

$$\lambda_d - \frac{d\mathbb{E}(\bar{Z}_n)}{m} - \frac{d^2\mathbb{E}(\bar{Z}_n^2)}{m^2} \leq (n-1)\text{Pr}_d \leq \lambda_d.$$

Zatem z lematu 5.1 i lematu 5.2 mamy:

$$\begin{aligned} d_{TV}(\bar{X}_n^d, T_{(n-1)\text{Pr}_d}) &\leq \text{Pr}_d \leq \frac{d\mathbb{E}(\bar{Z}_n)}{m} \\ d_{TV}(T_{(n-1)\text{Pr}_d}, T_{\lambda_d}) &\leq \frac{d\mathbb{E}(\bar{Z}_n)}{m} + \frac{d^2\mathbb{E}(\bar{Z}_n^2)}{m^2} \end{aligned}$$

i ostatecznie

$$d_{TV}(\bar{X}_n^d, T_{\lambda_d}) \leq \varepsilon(n, d),$$

gdzie $\varepsilon(n, d) = \frac{2d\mathbb{E}(\bar{Z}_n)}{m} + \frac{d^2\mathbb{E}(\bar{Z}_n^2)}{m^2}$. Dlatego dla każdego $A \subseteq \mathbb{N}$:

$$(5.2) \quad \Pr\{T_{\lambda_d} \in A\} - \varepsilon(n, d) \leq \Pr\{\bar{X}_n^d \in A\} \leq \Pr\{T_{\lambda_d} \in A\} + \varepsilon(n, d).$$

Zauważmy, że

$$\begin{aligned} \sum_{d=0}^m \varepsilon(n, d) \Pr\{\bar{Z}_n = d\} &= \varepsilon(n), \\ \sum_{d=0}^m \Pr\{T_{\lambda_d} \in A\} \Pr\{\bar{Z}_n = d\} &= \Pr\{\mathbb{T}_{\bar{Z}_n \mathbb{E}\bar{Z}_n/r_n^2} \in A\} \quad \text{oraz} \\ \sum_{d=0}^m \Pr\{\bar{X}_n^d \in A\} \Pr\{\bar{Z}_n = d\} &= \Pr\{\bar{X}_n \in A\}. \end{aligned}$$

Stąd z (5.2) wynika, że

$$\Pr\{\mathbb{T}_{\bar{Z}_n \mathbb{E}\bar{Z}_n/r_n^2} \in A\} - \varepsilon(n) \leq \Pr\{\bar{X}_n \in A\} \leq \Pr\{\mathbb{T}_{\bar{Z}_n \mathbb{E}\bar{Z}_n/r_n^2} \in A\} + \varepsilon(n),$$

gdzie $\varepsilon(n) = \frac{2(\mathbb{E}(\bar{Z}_n))^2}{m} + \frac{(\mathbb{E}(\bar{Z}_n^2))^2}{m^2}$. Dlatego:

$$d_{TV}(\bar{X}_n, \mathbb{T}_{\bar{Z}_n \mathbb{E}\bar{Z}_n/r_n^2}) \leq \varepsilon(n).$$

□

Fakt 5.2. *Jeśli*

- (i) $\bar{Y}_n \xrightarrow{D} Y$, Y_n dąży według rozkładu do zmiennej losowej Y ;
- (ii) $\mathbb{E}\bar{Y}_n \rightarrow \mathbb{E}Y$;
- (iii) $\mathbb{E}\bar{Y}_n < \infty$ i $\mathbb{E}Y < \infty$;

wtedy dla dowolnego $\omega(n) \rightarrow \infty$

$$(5.3) \quad \mathbb{E}\bar{Y}_n \mathbb{I}_{\bar{Y}_n \geq \omega(n)} = o(1).$$

Jeśli dodatkowo

$$\Pr\{\bar{Y}_n \leq n\} = 1,$$

to

$$(5.4) \quad \frac{2(\mathbb{E}\bar{Y}_n)^2}{n} + \frac{(\mathbb{E}\bar{Y}_n^2)^2}{n^2} = o(1).$$

Dowód. Część dowodu (5.3) jest podobna do dowodu jednostajnej ograniczoności z [11]. Niech $\omega(n) \rightarrow \infty$, $\varepsilon > 0$ i Δ będzie takie, że $\Pr\{Y = \Delta\} = 0$ i $\mathbb{E}Y \mathbb{I}_{Y \geq \Delta} \leq \varepsilon/3$ (Skoro $\mathbb{E}Y < \infty$, to istnieje takie Δ). Zatem ze zbieżności według rozkładu

$$\mathbb{E}Y \mathbb{I}_{Y < \Delta} - \mathbb{E}\bar{Y}_n \mathbb{I}_{\bar{Y}_n < \Delta} = o(1)$$

i istnieje N takie, że dla $n > N$ mamy

$$\begin{aligned} \omega(n) &> \Delta, \\ \mathbb{E}Y \mathbb{I}_{Y < \Delta} - \mathbb{E}\bar{Y}_n \mathbb{I}_{\bar{Y}_n < \Delta} &< \varepsilon/3 \\ &\text{i} \\ \mathbb{E}\bar{Y}_n - \mathbb{E}Y &< \varepsilon/3. \end{aligned}$$

Zatem dla $n > N$

$$\begin{aligned} \mathbb{E}\bar{Y}_n \mathbb{I}_{\bar{Y}_n \geq \omega(n)} &\leq \\ &\leq \mathbb{E}\bar{Y}_n \mathbb{I}_{\bar{Y}_n \geq \Delta} = \mathbb{E}\bar{Y}_n - \mathbb{E}Y + \mathbb{E}Y \mathbb{I}_{Y < \Delta} - \mathbb{E}\bar{Y}_n \mathbb{I}_{\bar{Y}_n < \Delta} + \mathbb{E}Y \mathbb{I}_{Y \geq \Delta} < \varepsilon \end{aligned}$$

Dlatego

$$\mathbb{E}\bar{Y}_n \mathbb{I}_{\bar{Y}_n \geq \omega(n)} = o(1).$$

Ponieważ $\Pr\{\bar{Y}_n \leq n\} = 1$, to dla każdego ciągu $\omega(n)$ takiego, że $\omega(n) = o(n)$ i $\omega(n) \rightarrow \infty$, z (5.3) i własności wartości oczekiwanej

$$\begin{aligned} \mathbb{E}\bar{Y}_n^2 &= \mathbb{E}\bar{Y}_n^2 \mathbb{I}_{\bar{Y}_n < \omega(n)} + \mathbb{E}\bar{Y}_n^2 \mathbb{I}_{\bar{Y}_n \geq \omega(n)} \\ &\leq \omega(n) \mathbb{E}\bar{Y}_n + n \mathbb{E}\bar{Y}_n \mathbb{I}_{\bar{Y}_n \geq \omega(n)} = o(n) \end{aligned}$$

i

$$\frac{2(\mathbb{E}\bar{Y}_n)^2}{n} + \frac{(\mathbb{E}\bar{Y}_n^2)^2}{n^2} = \frac{2(\mathbb{E}Y + o(1))^2}{n} + \frac{o(n^2)}{n^2} = o(1).$$

□

5.3. Dowód twierdzenia o stopniach wierzchołków

Dowód twierdzenia 5.1. Przypomnijmy, że przez $X_n = X_n(v)$ oznaczamy zmienną losową określającą stopień wierzchołka $v \in \mathcal{V}$ w grafie $G(\mathcal{V}, m, \mathcal{P}_{(m)})$. Niech \mathbb{A} będzie zdarzeniem, że $|D(w)| \leq \sqrt{nm}$ dla wszystkich wierzchołków $w \in \mathcal{V}$. Wtedy z równania (5.3) oraz lematu 3.2

$$\begin{aligned} \Pr\{\mathbb{A}^c\} &= \Pr\{\exists w \in \mathcal{V} |D(w)| > \sqrt{nm}\} \leq \\ &\leq n \Pr\{Z_n > \sqrt{nm}\} = n \Pr\{Y_n > n\} \leq n \frac{\mathbb{E}Y_n \mathbb{I}_{Y_n \geq n}}{n} = o(1). \end{aligned}$$

Oznaczmy przez $G(\mathcal{V}, m, \overline{\mathcal{P}}_{(m)})$ graf $G(\mathcal{V}, m, \mathcal{P}_{(m)})$ przy założeniu, że zaszło zdarzenie \mathbb{A} . Niech $\bar{X}_n = \bar{X}_n(v)$ będzie stopniem wierzchołka v w grafie $G(\mathcal{V}, m, \overline{\mathcal{P}}_{(m)})$ oraz $\bar{Z}_n = |D(v)|$ zmienną losową $|D(v)|$ pod warunkiem zajścia zdarzenia \mathbb{A} . Oczywiście wtedy \bar{Z}_n i $\bar{Y}_n = \bar{Z}_n/r_n$ są, odpowiednio, zmiennymi losowymi Z_n i Y_n obcięzonymi, odpowiednio, do $(-\infty; \sqrt{nm}]$ i $(-\infty; n]$. To znaczy dla dowolnego zbioru borelowskiego $A \subseteq \mathbb{R}$

$$\begin{aligned} \Pr\{\bar{Z}_n \in A\} &= \frac{\Pr\{Z_n \in A \cap (-\infty; \sqrt{nm}]\}}{\Pr\{Z_n \in (-\infty; \sqrt{nm}]\}} \\ \text{oraz } \Pr\{\bar{Y}_n \in A\} &= \frac{\Pr\{Y_n \in A \cap (-\infty; n]\}}{\Pr\{Y_n \in (-\infty; n]\}}. \end{aligned}$$

Zauważmy, że $G(\mathcal{V}, m, \overline{\mathcal{P}}_{(m)})$ jest losowym grafem przecięć generowanym przez rozkład $\overline{\mathcal{P}}_{(m)}$ będący obcięciem rozkładu $\mathcal{P}_{(m)}$ do zbioru $(-\infty; \sqrt{nm}]$. W dodatku, skoro $Y_n \xrightarrow{D} Y$, to $\overline{Y}_n \xrightarrow{D} Y$ (gdyż dla dowolnego $x \in \mathbb{R}$ takiego, że $\Pr\{Y = x\} = 0$ istnieje N , takie że $x \leq N$, wtedy dla $n \geq N$

$$\Pr\{\overline{Y}_n \leq x\} = \Pr\{Y_n \leq x\} / \Pr\{Y_n \leq n\} \rightarrow \Pr\{Y \leq x\}$$

oraz

$$\mathbb{E}\overline{Y}_n \rightarrow \mathbb{E}Y \quad (\mathbb{E}\overline{Y}_n = (\mathbb{E}Y_n - \mathbb{E}Y_n \mathbb{I}_{Y_n > n}) / \Pr\{Y_n \leq n\}).$$

Czyli z lematu 5.3

$$d_{TV}(\overline{X}_n, \mathbb{T}_{\overline{Z}_n \mathbb{E}\overline{Z}_n / r_n^2}) \leq \frac{2(\mathbb{E}(\overline{Z}_n))^2}{m} + \frac{(\mathbb{E}(\overline{Z}_n^2))^2}{m^2}.$$

Skoro $\overline{Z}_n = r_n \overline{Y}_n$ oraz $r_n^2 = m/n$, to z faktu 5.2

$$d_{TV}(\overline{X}_n, \mathbb{T}_{\overline{Y}_n \mathbb{E}\overline{Y}_n}) \leq \frac{2(\mathbb{E}(\overline{Y}_n))^2}{n} + \frac{(\mathbb{E}(\overline{Y}_n^2))^2}{n^2} = o(1).$$

Dlatego dla każdego $k \in \mathbb{N}$

$$(5.5) \quad |\Pr\{\overline{X}_n = k\} - \Pr\{\mathbb{T}_{\overline{Y}_n \mathbb{E}\overline{Y}_n} = k\}| = o(1).$$

Z faktu 5.1 $\overline{Y}_n \mathbb{E}\overline{Y}_n \xrightarrow{D} Y \mathbb{E}Y$. Podstawmy $f_k(x) = \Pr\{T_x = k\} = \frac{x^k}{k!} e^{-x}$. Funkcja $f_k(x)$ jest ciągła i ograniczona, dlatego z własności zbieżności według rozkładu

$$\mathbb{E}f_k(\overline{Y}_n \mathbb{E}\overline{Y}_n) \rightarrow \mathbb{E}f_k(Y \mathbb{E}Y).$$

To znaczy z definicji rozkładu złożonego Poissona (5.1)

$$(5.6) \quad \Pr\{\mathbb{T}_{\overline{Y}_n \mathbb{E}\overline{Y}_n} = k\} \rightarrow \Pr\{\mathbb{T}_{Y \mathbb{E}Y} = k\}.$$

Dlatego z (5.5) oraz (5.6) dla dowolnego $k \in \mathbb{N}$ mamy

$$\Pr\{\overline{X}_n = k\} \rightarrow \Pr\{\mathbb{T}_{Y \mathbb{E}Y} = k\}.$$

Czyli ostatecznie

$$\begin{aligned} \Pr\{X_n = k\} &= \Pr\{\overline{X}_n = k\} \Pr\{\mathbb{A}\} + \Pr\{X_n = k | \mathbb{A}^c\} \Pr\{\mathbb{A}^c\} \rightarrow \\ &\rightarrow \Pr\{\mathbb{T}_{Y \mathbb{E}Y} = k\}, \end{aligned}$$

co dla zmiennych losowych o wartościach w \mathbb{N} jest równoważne z

$$X_n \xrightarrow{D} \mathbb{T}_{Y \mathbb{E}Y}.$$

□

6. Współczynnik skupienia

Współczynnik skupienia jest parametrem, który mierzy „lokalne” zależności między połączeniami w sieci. Im wyższy współczynnik tym większe takie zależności. Zakłada się, że w dobrym modelu sieci złożonej, gdy liczba wierzchołków w modelu dąży do nieskończoności, współczynnik skupienia dąży do stałej.

W kontekście losowych grafów przecięć lokalny współczynnik skupienia badali Deijfen i Kets w [19]. Pokazali oni przykład zmiennej losowej W , dla której stopnie wierzchołków w $G^{Bin}(\mathcal{V}, m, W)$ mają rozkład spełniający prawo potęgowe. Udowodnili także, że współczynnik skupienia w grafie dla tej wyznaczonej zmiennej losowej W jest asymptotycznie niezależny od n przy $m \asymp n$. W tym rozdziale pokażemy analogiczny wynik dotyczący współczynnika skupienia dla ogólniejszego modelu grafu $G(\mathcal{V}, m, \mathcal{P}_{(m)})$. Poniższe twierdzenie można zastosować dla grafu spełniającego założenia twierdzenia 5.1 oraz twierdzenia z [11].

6.1. Twierdzenie o lokalnym współczynniku skupienia

Twierdzenie 6.1. *Niech $\text{Clus}_{loc}(v)$ oznacza lokalny współczynnik skupienia wierzchołka $v \in \mathcal{V}$ w losowym grafie przecięć $G(\mathcal{V}, m, \mathcal{P}_{(m)})$. Niech Z_n będzie zmienną losową o rozkładzie $\mathcal{P}_{(m)}$ oraz $Y_n = \frac{Z_n}{r_n}$ dla $r_n = \sqrt{m/n}$. Jeśli $r_n = o(\sqrt[3]{m})$, $\Pr\{Z_n \geq 2\} = 1$ i $\mathbb{E}\frac{Z_n}{r_n} = O(1)$, wtedy dla dowolnego $v \in \mathcal{V}$*

$$\text{Clus}_{loc}(v) = \mathbb{E} \frac{1}{Z_n} + o(1).$$

Dowód. Oznaczmy przez $G = G(\mathcal{V}, m, \mathcal{P}_{(m)})$ oraz $E(G) = E(G(\mathcal{V}, m, \mathcal{P}_{(m)}))$. Niech $|D(v)| = d_v \geq 2$, $|D(w)| = d_w \geq 2$ oraz $|D(u)| = d_u \geq 2$. Niech także $d_v, d_w, d_u \leq \Delta = o(\sqrt[3]{m})$. Przypomnijmy, że

$$\begin{aligned} \text{Clus}_{loc}(v) &= \\ &= \Pr\{wu \in E(G) | vw, vu \in E(G)\} = \frac{\Pr\{wu, vw, vu \in E(G)\}}{\Pr\{vw \in E(G)\} \Pr\{vu \in E(G)\}}. \end{aligned}$$

Wprowadźmy dodatkowe oznaczenia

$$\Pr_{vwu}\{\cdot\} = \Pr\left\{ \cdot \mid |D(v)| = d_v, |D(w)| = d_w, |D(u)| = d_u \right\}$$

oraz

$$\begin{aligned}
P_{\cap k} &= \Pr_{vwu} \{|D(v) \cap D(w)| = k\}, \\
P_{1|\cap k} &= \\
&= \Pr_{vwu} \{wu, vw, vu \in E(G), D(v) \cap D(w) \cap D(u) \neq \emptyset \mid |D(v) \cap D(w)| = k\}, \\
P_{2|\cap k} &= \\
&= \Pr_{vwu} \{wu, vw, vu \in E(G), D(v) \cap D(w) \cap D(u) = \emptyset \mid |D(v) \cap D(w)| = k\}.
\end{aligned}$$

Wtedy

$$(6.1) \quad \Pr_{vwu} \{wu, vw, vu \in E(G)\} = \sum_{k=1}^{\min\{d_v, d_w\}} P_{1|\cap k} P_{\cap k} + \sum_{k=1}^{\min\{d_v, d_w\}} P_{2|\cap k} P_{\cap k}.$$

Dla $d_v + d_w \leq m/2$, $k \geq 1$ oraz $k < \min\{d_v, d_w\}$

$$\begin{aligned}
P_{\cap k+1} &= \frac{\binom{d_v}{k+1} \binom{m-d_v}{d_w-k-1}}{\binom{m}{d_w}} = \\
&= \frac{(d_v)_{k+1} (d_w)_{k+1}}{(k+1)! (m)_{k+1}} \frac{(m-d_v)_{d_w-k-1}}{(m-k-1)_{d_w-k-1}} = \\
&= \frac{(d_v)_k (d_w)_k}{k! (m)_k} \cdot \frac{(m-d_v)_{d_w-k}}{(m-k)_{d_w-k}} \cdot \\
&\quad \cdot \frac{(d_v-k)(d_w-k)}{(k+1)(m-k)} \cdot \frac{(m-k)}{(m-d_v-d_w+k+1)} \leq \\
&\leq P_{\cap k} \frac{d_v d_w}{m} \frac{2}{k+1} \leq P_{\cap k} \frac{d_v d_w}{m}.
\end{aligned}$$

Dlatego dla dostatecznie dużych n

$$\begin{aligned}
P_{\cap 1} &\leq \frac{d_v \binom{m-1}{d_w-1}}{\binom{m}{d_w}} = \frac{d_v d_w}{m}; \\
P_{\cap k} &\leq P_{\cap 1} \left(\frac{d_v d_w}{m} \right)^{k-1}; \\
P_{1|\cap k} &\leq \frac{k \binom{m-1}{d_u-1}}{\binom{m}{d_u}} = \frac{k d_u}{m} \leq \frac{\Delta d_u}{m}; \\
P_{2|\cap k} &\leq \frac{(d_v-k)(d_w-k) \binom{m-2}{d_u-2}}{\binom{m}{d_u}} \leq \frac{d_v d_w d_u (d_u-1)}{m(m-1)} \leq \frac{d_v d_w}{m} \cdot \frac{d_u^2}{m}.
\end{aligned}$$

Stąd dla n takiego, że $m/(m - d_v d_w) \leq 2$ mamy

$$\begin{aligned} \sum_{k=2}^{\max\{d_v, d_w\}} P_{1|\cap k} P_{\cap k} &\leq P_{\cap 1} \sum_{k=2}^{\max\{d_v, d_w\}} \frac{\Delta d_u}{m} \left(\frac{d_v d_w}{m} \right)^{k-1} \leq \\ &\leq P_{\cap 1} \frac{\Delta d_u}{m} \cdot \frac{d_v d_w}{m} \cdot \frac{m}{m - d_v d_w} \leq \frac{d_v d_w d_u}{m^2} \cdot \frac{2\Delta^3}{m} \\ \sum_{k=1}^{\max\{d_v, d_w\}} P_{2|\cap k} P_{\cap k} &\leq P_{\cap 1} \sum_{k=1}^{\max\{d_v, d_w\}} \frac{d_u^2}{m} \left(\frac{d_v d_w}{m} \right)^k \leq \\ &\leq P_{\cap 1} \frac{d_u^2}{m} \cdot \frac{d_v d_w}{m} \cdot \frac{m}{m - d_v d_w} \leq \frac{d_v d_w d_u}{m^2} \cdot \frac{2\Delta^3}{m} \end{aligned}$$

oraz

$$\begin{aligned} \frac{d_v d_w d_u}{m^2} \left(1 - \frac{2\Delta^2}{m} \right) &\leq \frac{d_v \binom{m-1}{d_w-1}}{\binom{m}{d_w}} \cdot \frac{1 \cdot \binom{m-d_v-d_w+1}{d_u-1}}{\binom{m}{d_u}} \leq \\ &\leq P_{\cap 1} P_{1|\cap 1} \leq \frac{d_v \binom{m-1}{d_w-1}}{\binom{m}{d_w}} \cdot \frac{1 \cdot \binom{m-1}{d_u-1}}{\binom{m}{d_u}} = \frac{d_v d_w d_u}{m^2}. \end{aligned}$$

Ostatecznie z (6.1) i powyższych oszacowań

$$\frac{d_v d_w d_u}{m^2} \left(1 - \frac{2\Delta^2}{m} \right) \leq \Pr_{vwu} \{wu, vw, vu \in E(G)\} \leq \frac{d_v d_w d_u}{m^2} \left(1 + \frac{4\Delta^3}{m} \right).$$

W dodatku z lematu 3.7

$$\begin{aligned} \frac{d_v d_w}{m} \left(1 - \frac{\Delta^2}{m} \right) &\leq \Pr_{vwu} \{vw \in E(G)\} \leq \frac{d_v d_w}{m}, \\ \frac{d_v d_u}{m} \left(1 - \frac{\Delta^2}{m} \right) &\leq \Pr_{vwu} \{vu \in E(G)\} \leq \frac{d_v d_u}{m}. \end{aligned}$$

Stąd

$$\Pr_{vwu} \{wu \in E(G) | vw, vu \in E(G)\} = \frac{1}{d_v} \left(1 + O\left(\frac{\Delta^3}{m}\right) \right),$$

gdzie $O(\cdot)$ jest jednostajnie ograniczone względem $d_v, d_w, d_u \leq \Delta$.
Oznaczmy

$$\xi := \Pr\{Z_n > \Delta\}.$$

Wtedy

$$\Pr\{|D(v)| > \Delta \text{ lub } |D(w)| > \Delta \text{ lub } |D(u)| > \Delta\} \leq 3\xi.$$

oraz

$$\begin{aligned}
\text{Clus}_{loc}(v) &\leq \sum_{d_v, d_w, d_u \leq \Delta} \sum_{vwu} \Pr\{wu \in E(G) | vw, vu \in E(G)\} \\
&\quad \cdot \Pr\{|D(v)| = d_v, |D(w)| = d_w, |D(u)| = d_u\} + 3\xi \\
&= \left(1 + O\left(\frac{\Delta^3}{m}\right)\right) \cdot \sum_{d_v \leq \Delta} \frac{1}{d_v} \Pr\{|D(v)| = d_v\} \\
&\quad \cdot \sum_{d_w \leq \Delta} \Pr\{|D(w)| = d_w\} \cdot \sum_{d_u \leq \Delta} \Pr\{|D(u)| = d_u\} + 3\xi \\
&\leq \left(1 + O\left(\frac{\Delta^3}{m}\right)\right) \mathbb{E} \frac{1}{Z_n} + 3\xi
\end{aligned}$$

i

$$\begin{aligned}
\text{Clus}_{loc}(v) &\geq \sum_{d_v, d_w, d_u \leq \Delta} \sum_{vwu} \Pr\{wu \in E(G) | vw, vu \in E(G)\} \\
&\quad \cdot \Pr\{|D(v)| = d_v, |D(w)| = d_w, |D(u)| = d_u\} \\
&= \left(1 + O\left(\frac{\Delta^3}{m}\right)\right) \cdot \sum_{d_v \leq \Delta} \frac{1}{d_v} \Pr\{|D(v)| = d_v\} \\
&\quad \cdot \sum_{d_w \leq \Delta} \Pr\{|D(w)| = d_w\} \cdot \sum_{d_u \leq \Delta} \Pr\{|D(u)| = d_u\} \\
&= \left(1 + O\left(\frac{\Delta^3}{m}\right)\right) (1 - \xi)^2 \left(\mathbb{E} \frac{1}{Z_n} - \sum_{d_v > \Delta} \frac{1}{d_v} \Pr\{|D(v)| = d_v\} \right) \\
&\geq \left(1 + O\left(\frac{\Delta^3}{m}\right)\right) (1 - \xi)^2 \left(\mathbb{E} \frac{1}{Z_n} - \frac{\xi}{\Delta} \right).
\end{aligned}$$

Niech $\omega(n) \rightarrow \infty$, $\omega(n) = o(\sqrt[3]{m}/r_n)$ i $\Delta = r_n \omega(n)$. Wtedy

$$\frac{\Delta^3}{m} = \frac{r_n^3 \omega(n)^3}{m} = o(1)$$

i z lematu 3.1

$$\xi = \Pr\{Z_n > \Delta\} = \Pr\left\{\frac{Z_n}{r_n} > \omega(n)\right\} \leq \frac{\mathbb{E} \frac{Z_n}{r_n}}{\omega(n)} = o(1).$$

Stąd natychmiast wynika teza twierdzenia. \square

Przyjmijmy teraz, że spełnione są założenia twierdzenia 5.1. Wówczas prawdziwe są również założenia twierdzenia 6.1 o ile $r_n = \sqrt{m/n} = o(\sqrt[3]{m})$. Zatem dla $m = o(n^3)$ i $n = o(m)$ mamy

$$\mathbb{E} \frac{1}{Z_n} = \sqrt{\frac{n}{m}} \mathbb{E} \frac{1}{Y_n} = o(1),$$

czyli w tym przypadku

$$\text{Clus}_{loc}(v) = o(1).$$

Natomiast, jeśli $m = cn$ dla pewnej stałej $c > 0$, założenie, że $Z_n \geq 2$ z prawdopodobieństwem równym 1, jest równoważne założeniu, że z prawdopodobieństwem równym 1 zachodzi $Y_n = Z_n/\sqrt{c} \geq 2/\sqrt{c}$. Stąd ze zbieżności według rozkładu i powyższego twierdzenia

$$\text{Clus}_{loc}(v) = \mathbb{E} \frac{1}{Z_n} = \sqrt{c} \mathbb{E} \frac{1}{Y_n} = \sqrt{c} \mathbb{E} f(Y_n) \rightarrow \sqrt{c} \mathbb{E} f(Y) = \sqrt{c} \mathbb{E} \frac{1}{Y},$$

dla ciągłej i ograniczonej funkcji

$$f(x) = \begin{cases} \frac{\sqrt{c}}{2}, & \text{dla } x \leq \frac{2}{\sqrt{c}}; \\ \frac{1}{x}, & \text{dla } x \geq \frac{2}{\sqrt{c}}. \end{cases}$$

Otrzymujemy więc, że

$$\text{Clus}_{loc}(v) \sim \sqrt{c} \mathbb{E} \frac{1}{Y}$$

i lokalny współczynnik skupienia przy założeniach z twierdzenia 5.1 dąży do stałej różnej od zera tylko wtedy gdy $m \asymp n$.

7. Kliki

W poprzednim rozdziale zaprezentowane zostało twierdzenie dotyczące lokalnego współczynnika skupienia w grafie $G(\mathcal{V}, m, \mathcal{P}_{(m)})$. Wynikało z niego, że przy dodatkowych założeniach graf ten jest dobrym modelem sieci złożonych gdy $n \asymp m$. Wcześniej, w podrozdziale 1.4.5 wspomniane były także inne definicje współczynnika skupienia. Ich przykładem jest określenie globalnego współczynnika skupienia.

Pierwszym krokiem w kierunku zbadania globalnego współczynnika skupienia danego modelu grafowego jest uzyskanie wyników dotyczących jego liczby trójkątów. Niezależnie od definicji współczynnika skupienia, jego wartość jest nieodłącznie związana z liczbą klik w porównaniu do liczby ścieżek. W tym rozdziale będziemy zajmować się liczbą kopii klik w różnych modelach grafów przecięć. Szczególnie będziemy zainteresowani liczbą klik w $G(\mathcal{V}, m, \mathcal{P}_{(m)})$ w momencie pojawienia się pierwszych z nich (tzn. dla rozkładów $\mathcal{P}_{(m)}$ będących funkcjami progowymi dla własności zawierania danej kliki). Główny wynik będzie dotyczył losowego grafu przecięć $G^{Bin}(\mathcal{V}, m, p)$, ale wnioski z niego płynące będą już dotyczyły interesujących nas z punktu widzenia modelowania sieci grafów $G^{Jed}(\mathcal{V}, m, d)$.

Wykorzystamy w naszych rozważaniach wyniki dotyczące funkcji progowych w losowym grafie przecięć przedstawione w pracy [40]. Są one częściowo powiązane z wynikami dotyczącymi grafów k -dzielnych zaprezentowanymi w pracy [39].

Część wyników zaprezentowanych w tym rozdziale została zawarta w złożonej do recenzji pracy [57] wspólnej z D. Starkiem.

7.1. Funkcja progowa dla własności posiadania kliki jako podgrafu

Klasycznym założeniem dla grafu $G^{Bin}(\mathcal{V}, m, p)$ (przyjętym na przykład w [40, 58]) ułatwiającym analizę tego losowego grafu przecięć jest

$$(7.1) \quad m = \lfloor n^\alpha \rfloor$$

dla pewnej stałej $\alpha > 0$. W [40] została wyznaczona funkcja progowa w grafie $G^{Bin}(\mathcal{V}, m, p)$ dla własności posiadania kliki K_h jako podgrafu ($h \geq 3$ jest pewną stałą całkowitą dodatnią). Zostało udowodnione, że dla $\tau(K_h)$ takiego, że

$$\tau(K_h) = \begin{cases} n^{-1}m^{-1/h} & \text{dla } \alpha \leq 2h/(h-1); \\ n^{-1/(h-1)}m^{-1/2} & \text{dla } \alpha \geq 2h/(h-1) \end{cases}$$

zachodzi

$$\lim_{n \rightarrow \infty} \Pr\{K_h \subseteq G^{Bin}(\mathcal{V}, m, p)\} = \begin{cases} 1 & \text{jeśli } p/\tau(K_h) \rightarrow 0, \\ 0 & \text{jeśli } p/\tau(K_h) \rightarrow \infty, \end{cases}$$

gdzie „ $K_h \subseteq G^{Bin}(\mathcal{V}, m, p)$ ” oznacza zdarzenie, że graf $G^{Bin}(\mathcal{V}, m, p)$ zawiera jako podgraf kopię grafu K_h .

Z lematu 4.2 o równoważności modeli $G^{Bin}(\mathcal{V}, m, p)$ i $G^{Jed}(\mathcal{V}, m, d)$ można wywnioskować, że w $G^{Jed}(\mathcal{V}, m, d)$ dla $\alpha > \frac{h}{h-1}$ funkcja progowa własności istnienia kopii klik K_h w $G^{Jed}(\mathcal{V}, m, d)$ jest postaci

$$\tau_d(K_h) = \begin{cases} n^{-1} m^{\frac{h-1}{h}} & \text{dla } h/(h-1) < \alpha \leq 2h/(h-1); \\ n^{-1/(h-1)} m^{1/2} & \text{dla } \alpha \geq 2h/(h-1). \end{cases}$$

W dalszej części pokażemy, że gdy p (odpowiednio d) jest funkcją progową, to liczba klik K_h w $G^{Bin}(\mathcal{V}, m, p)$ (odpowiednio w $G^{Jed}(\mathcal{V}, m, d)$) dąży do zmiennej losowej o rozkładzie Poissona.

7.2. Zbieżność liczby klik do rozkładu Poissona

Przypomnijmy, że odległość $d_{TV}(X, Y)$ między zmiennymi losowymi X i Y została określona w podrozdziale 3.2.3. Dla wygody przez $d_{TV}(X, \text{Po}(\lambda))$ będziemy oznaczać odległość między zmienną losową X a zmienną losową Y o rozkładzie Poissona z parametrem λ .

Twierdzenie 7.1. *Niech $c > 0$ będzie stałą i $h \geq 3$ będzie ustaloną liczbą całkowitą oraz*

$$p \sim \begin{cases} cn^{-1} m^{-\frac{1}{h}} & \text{dla } 0 < \alpha < 2h/(h-1); \\ cn^{-\frac{h+1}{h-1}} & \text{dla } \alpha = 2h/(h-1); \\ cn^{-\frac{1}{h-1}} m^{-\frac{1}{2}} & \text{dla } \alpha > 2h/(h-1). \end{cases}$$

Niech X_n będzie zmienną losową liczącą podgrafy izomorficzne z K_h w grafie $G^{Bin}(\mathcal{V}, m, p)$.

(i) *Jeśli $0 < \alpha < 2h/(h-1)$, to $\mathbb{E}X_n \sim c^h/h!$ oraz*

$$d_{TV}(X_n, \text{Po}(\mathbb{E}X_n)) = O\left(n^{-\frac{\alpha}{h}}\right);$$

(ii) *Jeśli $\alpha = 2h/(h-1)$ to $\mathbb{E}X_n \sim (c^h + c^{h(h-1)})/h!$ oraz*

$$d_{TV}(X_n, \text{Po}(\mathbb{E}X_n)) = O\left(n^{-\frac{2}{h-1}}\right);$$

(iii) *Jeśli $\alpha > 2h/(h-1)$ to $\mathbb{E}X_n \sim c^{h(h-1)}/h!$ oraz*

$$d_{TV}(X_n, \text{Po}(\mathbb{E}X_n)) = O\left(n^{\left(h - \frac{\alpha(h-1)}{2} - \frac{2}{h-1}\right)} + n^{-1}\right).$$

Bezpośrednio z powyższego twierdzenia, uwag zawartych w podrozdziale 3.2.3 i lematu 4.3 wynika wniosek dla modelu jednostajnego.

Wniosek 7.1. Niech $c > 0$ będzie stałą i $h \geq 3$ będzie ustaloną liczbą całkowitą oraz

$$d \sim \begin{cases} cn^{-1}m^{\frac{h-1}{h}} & \text{dla } h/(h-1) < \alpha < 2h/(h-1); \\ cn & \text{dla } \alpha = 2h/(h-1); \\ cn^{-\frac{1}{h-1}}m^{\frac{1}{2}} & \text{dla } \alpha > 2h/(h-1). \end{cases}$$

Jeżeli X_n jest zmienną losową liczącą podgrafy izomorficzne z K_h w grafie $G^{Jed}(\mathcal{V}, m, d)$, to X_n dąży według rozkładu do zmiennej losowej o rozkładzie Poissona z parametrem λ , gdzie

$$\lambda = \begin{cases} c^h/h! & \text{dla } h/(h-1) < \alpha < 2h/(h-1); \\ (c^h + c^{h(h-1)})/h! & \text{dla } \alpha = 2h/(h-1); \\ c^{h(h-1)}/h! & \text{dla } \alpha > 2h/(h-1). \end{cases}$$

7.3. Dowód twierdzenia o liczbie klik

Dowód będzie przebiegał w kilku fazach. Będziemy w nim wykorzystywać metodę Steina przedstawioną w podrozdziale 3.2.3.

Dla wygody będziemy zakładać, że

$$(7.2) \quad \alpha = \frac{2\delta h}{h-1}$$

dla pewnej stałej $\delta > 0$. Wtedy dla stałej $c > 0$ określonej w założeniach twierdzenia,

$$(7.3) \quad p(n) \sim \begin{cases} cn^{-1}m^{-\frac{1}{h}} & \text{dla } 0 < \delta < 1; \\ cn^{-\frac{h+1}{h-1}} & \text{dla } \delta = 1; \\ cn^{-\frac{1}{h-1}}m^{-\frac{1}{2}} & \text{dla } \delta > 1. \end{cases}$$

Łatwo sprawdzić, że dla m i p zdefiniowanych jak w (7.1), (7.2) i (7.3)

$$(7.4) \quad mp^2 = o(1)$$

dla każdego $\delta > 0$ i

$$(7.5) \quad mp \rightarrow \infty$$

dla $\delta > \frac{1}{2}$.

7.3.1. Zastosowanie metody Steina

Jak już wspomnieliśmy wyżej zastosujemy metodę Steina przedstawioną w podrozdziale 3.2.3. Załóżmy, że $\Gamma = \{K_{\mathcal{V}'} : \mathcal{V}' \subseteq \mathcal{V}, |\mathcal{V}'| = h\}$ jest zbiorem indeksów indykatorowych zmiennych losowych. Przyjmijmy również, że dla każdego elementu ze zbioru indeksów $\beta \in \Gamma$ zachodzi

$$I_\beta = \mathbb{I}(\beta \subseteq G^{Bin}(\mathcal{V}, m, p)),$$

gdzie

$$\mathbb{I}(\beta \subseteq G^{Bin}(\mathcal{V}, m, p)) = \begin{cases} 1 & , \text{ gdy } \beta \subseteq G^{Bin}(\mathcal{V}, m, p); \\ 0 & , \text{ gdy } \beta \not\subseteq G^{Bin}(\mathcal{V}, m, p). \end{cases}$$

Przyjmując takie oznaczenia, zmienna losowa $W = \sum_{\beta \in \Gamma} I_\beta$ określa liczbę kopii klik na h wierzchołkach w $G^{Bin}(\mathcal{V}, m, p)$. Także, z definicji z podrozdziału 3.2.3 $\pi_\beta = \mathbb{E}I_\beta$, czyli

$$\pi_\beta = \Pr\{\beta \subseteq G^{Bin}(\mathcal{V}, m, p)\}.$$

W dodatku

$$(7.6) \quad \begin{aligned} \mathbb{E}(I_\beta I_\gamma) &= \Pr\{\beta \subseteq G^{Bin}(\mathcal{V}, m, p), \gamma \subseteq G^{Bin}(\mathcal{V}, m, p)\} \\ &= \Pr\{\beta \cup \gamma \subseteq G^{Bin}(\mathcal{V}, m, p)\}, \end{aligned}$$

gdzie $\beta \cup \gamma$ jest grafem o zbiorze wierzchołków $V(\beta \cup \gamma) = V(\beta) \cup V(\gamma)$ i zbiorze krawędzi $E(\beta \cup \gamma) = E(\beta) \cup E(\gamma)$.

Zdefiniujmy teraz graf zależności L o zbiorze wierzchołków Γ i zbiorze krawędzi takim, że jeśli $\beta = K_{\mathcal{V}'}$ i $\gamma = K_{\mathcal{V}''}$, to $\{\beta, \gamma\} \in E(L)$ wtedy i tylko wtedy gdy $\mathcal{V}' \cap \mathcal{V}'' \neq \emptyset$. Podgrafy o rozłącznych zbiorach wierzchołków powstają niezależnie w $G^{Bin}(\mathcal{V}, m, p)$, więc jest to dobrze określony graf zależności. Z (3.4) dla $\lambda = \mathbb{E}W$ wynika, że

$$(7.7) \quad \begin{aligned} d_{TV}(W, \text{Po}(\lambda)) &\leq \\ &\leq \min(\lambda^{-1}, 1) \left(\sum_{\beta \in V(L)} \pi_\beta^2 + \sum_{\beta\gamma \in E(L)} (\mathbb{E}(I_\beta I_\gamma) + \pi_\beta \pi_\gamma) \right). \end{aligned}$$

W kolejnych podrozdziałach, w celu oszacowania normy całkowitego wahania, zajmiemy się oszacowaniem wielkości π_β i $\mathbb{E}(I_\beta I_\gamma)$.

7.3.2. Pokrycia klikowe

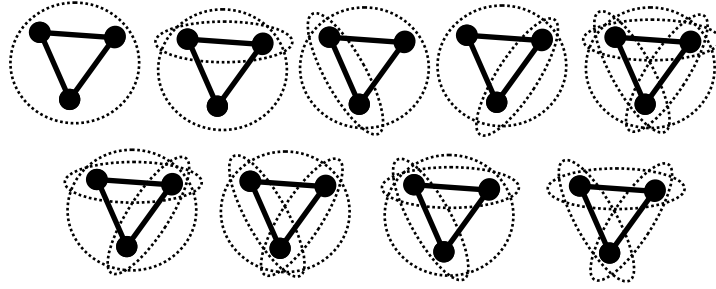
Aby oszacować π_β oraz $\mathbb{E}(I_\beta I_\gamma)$, wykorzystamy pojęcie pokryć klikowych wprowadzone w [40]. Pojęcie to zmodyfikujemy nieznacznie dla naszych potrzeb.

Definicja 7.1. Niech H będzie dowolnym podgrafem grafu pełnego $K_{\mathcal{V}}$. Zbiór \mathcal{C} podzbiorów zbioru $V(H)$ nazywamy **pokryciem klikowym grafu H** , jeśli

- (i) $|C| \geq 2$ dla każdego $C \in \mathcal{C}$ oraz
- (ii) dla każdej krawędzi $v_1 v_2 \in E(H)$ istnieje $C \in \mathcal{C}$ takie, że $v_1, v_2 \in C$.

Często, jeśli będziemy wiedzieć jaki podgraf rozpatrujemy lub gdy będziemy mówić o pokryciu dowolnego podgrafu, będziemy mówić po prostu o pokryciu klikowym \mathcal{C} . Zauważmy, że z powyższej definicji wynika, że kliki indukowane w $K_{\mathcal{V}}$ przez zbiory z \mathcal{C} pokrywają wszystkie krawędzie podgrafu H . Także niektóre krawędzie podgrafu mogą być zawarte w więcej niż jednej klicie indukowanej przez zbiory z \mathcal{C} .

Przykład 7.1. Na rysunku 7.1 pokazane zostały wszystkie pokrycia klikowe trójkąta (jest ich dziewięć).



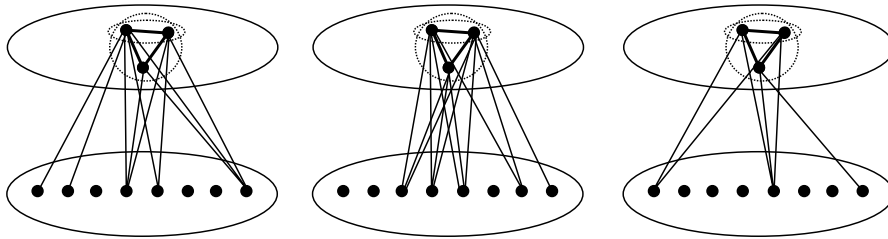
Rysunek 7.1. Pokrycia klikowe trójkąta

W losowym grafie przecięć zbiór wierzchołków, które mają pewną własność $w \in \mathcal{W}$, tzn. zbiór $\{v : w \in D(v)\}$, tworzy klikę w grafie $G^{Bin}(\mathcal{V}, m, p)$ oraz zbiór krawędzi w $G^{Bin}(\mathcal{V}, m, p)$ jest sumą m zbiorów krawędzi klik generowanych przez elementy z \mathcal{W} .

Definicja 7.2. Mówimy, że $H \subseteq G^{Bin}(\mathcal{V}, m, p)$ jest **indukowany przez pokrycie klikowe** $\mathcal{C} = \{C_1, \dots, C_t\}$, jeżeli istnieje rodzina rozłącznych niepustych podzbiorów $\{W_1, \dots, W_t\}$ zbioru \mathcal{W} , takich, że

1. dla każdego $1 \leq i \leq t$, każdy element z W_i jest własnością wszystkich wierzchołków z C_i (tzn. należy do $D(v)$ dla każdego $v \in C_i$) oraz żaden element z W_i nie jest własnością żadnego innego wierzchołka z $V(H)$;
2. każda z własności $w \in \mathcal{W} \setminus \bigcup_{i=1}^t W_i$ jest własnością co najwyżej jednego wierzchołka z $V(H)$.

Zauważmy, że jest wiele sposobów na to, żeby dany podgraf był indukowany przez dane pokrycie klikowe. Rysunek 7.2 pokazuje trzy sposoby na to jak trójkąt może być indukowany przez pokrycie klikowe składające się z dokładnie jednego zbioru o mocy 2 i jednego zbioru o mocy 3.



Rysunek 7.2. Przykład, jak trójkąt może być indukowany przez pokrycie klikowe składające się z dokładnie jednego zbioru o mocy 2 i jednego zbioru o mocy 3.

Z definicji wynika, że w danym losowym grafie $G^{Bin}(\mathcal{V}, m, p)$ podgraf H jest indukowany przez dokładnie jedno pokrycie klikowe H oraz każdy graf ma skończoną liczbę pokryć klikowych. Jest tak dlatego, że każdy podzbiór wierzchołków może wystąpić w pokryciu klikowym co najwyżej raz a podzbiorów skończonego zbioru wierzchołków jest skończona liczba. Dlatego, jeśli

oznaczymy przez $\pi(H, \mathcal{C})$ prawdopodobieństwo, że podgraf H jest indukowany przez pokrycie klikowe $\mathcal{C} = \{C_1, \dots, C_t\}$, wtedy

$$(7.8) \quad \pi(H) = \sum_{\text{pokrycia klikowe } \mathcal{C} \text{ podgrafu } H} \pi(H, \mathcal{C}).$$

Zatem, skoro jest skończona liczba pokryć klikowych danego podgrafu, asymptotycznie wartość $\pi(H)$ zależy od pokryć klikowych \mathcal{C} podgrafu H dla których $\pi(H, \mathcal{C})$ dominuje asymptotycznie. Aby oszacować $\pi(H, \mathcal{C})$, wykorzystamy wzór

$$(7.9) \quad \pi(H, \mathcal{C}) \sim \prod_{i=1}^t mp^{|C_i|}.$$

Wynika on bezpośrednio z oszacowań na $\pi(H, \mathcal{C})$ z [40] (patrz strona 138), gdyż zakładamy tutaj, że w pokryciu klikowym zdefiniowanym jak powyżej $|C_i| \geq 2$ oraz $mp^2 = o(1)$.

W dalszych rozważaniach wyróżnimy dwa specyficzne rodzaje pokryć.

Definicja 7.3. *Pokrycie klikowe nazywamy **pokryciem krawędziami**, jeśli składa się tylko ze zbiorów dwuelementowych. Natomiast pokrycie klikowe składające się z dokładnie jednego zbioru będziemy nazywać **pokryciem jednym zbiorem**.*

7.3.3. Modyfikacje pokryć klikowych

Dla konkretnego podgrafu H (później będziemy rozpatrywać tylko podgrafy typu β lub $\beta \cup \gamma$, gdzie $\beta, \gamma \in \Gamma$) będziemy zainteresowani znalezieniem takich pokryć klikowych \mathcal{C} dla których $\pi(H, \mathcal{C})$ jest największe rzędowo spośród wszystkich pokryć grafu H . W tym celu będziemy badać zbieżność wyrażenia $\frac{\pi(H, \mathcal{C})}{\pi(H, \mathcal{C}')}$ dla różnych pokryć klikowych \mathcal{C} i \mathcal{C}' podgrafu H .

Dane pokrycie \mathcal{C} będziemy porównywać z innymi pokryciami uzyskanymi z \mathcal{C} poprzez zamianę niektórych zbiorów.

Definicja 7.4. *Niech \mathcal{C} będzie pokryciem podgrafu H . Mówimy, że \mathcal{C}' jest **uzyskana z \mathcal{C} przez scalenie dwóch zbiorów** jeśli dla pewnych zbiorów $C_1, C_2 \in \mathcal{C}$, pokrycie \mathcal{C}' jest uzyskane z \mathcal{C} poprzez zamianę zbiorów C_1 i C_2 na zbiór $C_1 \cup C_2$. Mówimy, że \mathcal{C}' jest **uzyskana z \mathcal{C} poprzez pokrycie krawędziami zbioru** jeśli dla pewnego $C \in \mathcal{C}$, \mathcal{C}' jest uzyskane z \mathcal{C} przez zastąpienie zbioru C przez rodzinę zbiorów dwuelementowych odpowiadających krawędziom grafu $H[C]$.*

Z (7.1), (7.2) i (7.3), wynika, że istnieje liczba rzeczywista a taka, że $p(n) \asymp n^{-a}$, gdzie

$$a = \begin{cases} \frac{h-1+2\delta}{h-1} & \text{dla } \delta \leq 1; \\ \frac{1+\delta h}{h-1} & \text{dla } \delta \geq 1. \end{cases}$$

Określmy wartość k_0 jako

$$k_0 = \frac{\alpha}{2a - \alpha}.$$

Z (7.2), wynika, że

$$k_0 = \begin{cases} \frac{\delta h}{h - \delta h - 1 + 2\delta} & \text{dla } \delta \leq 1; \\ \delta h & \text{dla } \delta \geq 1. \end{cases}$$

Dodatkowo mamy

$$(7.10) \quad \begin{aligned} 0 < k_0 < h & \text{ dla } \delta < 1; \\ k_0 = h & \text{ dla } \delta = 1; \\ k_0 > h & \text{ dla } \delta > 1. \end{aligned}$$

W lematkach poniżej otrzymamy, że zamiana zbiorów o mocy mniejszej od k_0 na pokrycia krawędziami oraz scalanie zbiorów o mocy większej od k_0 zwiększa wartość $\pi(H, \mathcal{C})$.

Lemat 7.1. *Niech H będzie ustalonym grafem, m i p będą spełniały (7.1), (7.2) i (7.3) oraz niech \mathcal{C} będzie dowolnym pokryciem klikowym H . Jeśli \mathcal{C}' jest pokryciem klikowym H otrzymanym z \mathcal{C} poprzez zastąpienie $C \in \mathcal{C}$ takiego, że $|C| = k > 2$, poprzez pokrycie krawędziowe podgrafu $H[C]$, wtedy jeśli $k_0 \geq 3$, to*

$$\frac{\pi(H, \mathcal{C})}{\pi(H, \mathcal{C}')} = o(1) \quad \text{dla } 2 < k < k_0;$$

$$\frac{\pi(H, \mathcal{C})}{\pi(H, \mathcal{C}')} = O(1) \quad \text{dla } k = k_0.$$

Jeśli dodatkowo $H[C]$ jest kliką, to

$$\frac{\pi(H, \mathcal{C})}{\pi(H, \mathcal{C}')} = o(1) \quad \text{dla } 2 < k < k_0;$$

$$\frac{\pi(H, \mathcal{C})}{\pi(H, \mathcal{C}')} \asymp 1 \quad \text{dla } k = k_0;$$

$$\frac{\pi(H, \mathcal{C}')}{\pi(H, \mathcal{C})} = o(1) \quad \text{dla } k > \max\{k_0, 2\}.$$

Dowód. Niech H będzie dowolnym ustalonym grafem, m i p spełniają (7.1), (7.2) i (7.3), oraz \mathcal{C} będzie dowolnym pokryciem klikowym podgrafu H . Porównamy prawdopodobieństwa $\pi(H, \mathcal{C})$ i $\pi(H, \mathcal{C}')$, gdzie \mathcal{C}' jest otrzymany z \mathcal{C} poprzez zastąpienie $C \in \mathcal{C}$, $|C| = k$, pokryciem krawędziami podgrafu $H[C]$. Niech e będzie liczbą krawędzi w $H[C]$. Wtedy,

$$mp^k \asymp n^{\alpha - ka}$$

i

$$(mp^2)^{\binom{k}{2}} \asymp n^{\frac{k(k-1)(\alpha-2a)}{2}}.$$

Zatem z (7.9)

$$\frac{\pi(H, \mathcal{C})}{\pi(H, \mathcal{C}')} \asymp \frac{mp^k}{(mp^2)^e} \leq \frac{mp^k}{(mp^2)^{\binom{k}{2}}} = n^{\frac{f(k)}{2}},$$

gdzie

$$\begin{aligned} f(k) &= 2\alpha - 2ka - k(k-1)(\alpha - 2a) = \\ &= (2a - \alpha)k^2 - (4a - \alpha)k + 2\alpha = \\ &= (2a - \alpha)(k-2)(k-k_0). \end{aligned}$$

Stąd

$$f(k) = 0 \Leftrightarrow k = 2 \text{ lub } k = k_0$$

i skoro $2a - \alpha > 0$

$$f(k) < 0 \text{ dla } k \in (2, k_0) \text{ przy } k_0 > 2$$

i

$$f(k) > 0 \text{ dla } k > \max\{k_0, 2\}.$$

W dodatku, jeśli $e = \binom{k}{2}$, to

$$\frac{\pi(H, \mathcal{C})}{\pi(H, \mathcal{C}')} \asymp n^{\frac{f(k)}{2}}.$$

□

Pokrycie klikowe \mathcal{C} nazywamy **prostym pokryciem klikowym** jeśli dla dowolnych różnych $C_1, C_2 \in \mathcal{C}$ mamy $|C_1 \cap C_2| \leq 1$.

Lemat 7.2. *Niech H będzie dowolnym ustalonym grafem, m i p spełniają (7.1), (7.2) i (7.3) oraz \mathcal{C} będzie prostym pokryciem klikowym grafu H . Załóżmy także, że $|C_1 \cap C_2| = 1$ oraz że w H dla $C_1, C_2 \in \mathcal{C}$ istnieje dokładnie k krawędzi $\{e_1, \dots, e_k\}$ z jednym wierzchołkiem w C_1 a drugim w C_2 . Jeśli \mathcal{C}' jest pokryciem klikowym grafu H otrzymanym z \mathcal{C} poprzez zastąpienie C_1 i C_2 zbiorem $C_1 \cup C_2$ oraz scaleniem z nim wszystkich zbiorów zawierających co najmniej jedną krawędź z $\{e_1, \dots, e_k\}$ ze zbiorem $C_1 \cup C_2$, wtedy,*

$$\frac{\pi(H, \mathcal{C})}{\pi(H, \mathcal{C}')} = o(1) \quad \text{dla } k > \frac{1}{2}k_0 - \frac{1}{2}$$

i

$$\frac{\pi(H, \mathcal{C})}{\pi(H, \mathcal{C}')} = O(1) \quad \text{dla } k = \frac{1}{2}k_0 - \frac{1}{2}.$$

Dowód. Niech H będzie dowolnym ustalonym grafem, $p \asymp n^{-a}$ i $m = n^\alpha$ oraz niech $k_1 + 1 \geq 2$ i $k_2 + 1 \geq 2$ będą, odpowiednio, mocami zbioru C_1 i C_2 z pokrycia klikowego \mathcal{C} . Załóżmy także, że $\{e_1, \dots, e_k\}$ jest zbiorem wszystkich krawędzi o dokładnie jednym wierzchołku w C_1 i dokładnie jednym wierzchołku w C_2 . Skoro \mathcal{C} jest prostym pokryciem klikowym, to dla każdego e_i jest dokładnie jeden zbiór $C_{e_i} \in \mathcal{C}$ zawierający e_i . Także, dla $i \neq j$, $C_{e_i} \neq C_{e_j}$ (w przeciwnym przypadku, $|C_{e_i} \cap C_1| \geq 2$ lub $|C_{e_i} \cap C_2| \geq 2$). Z treści lematu wynika, że \mathcal{C}' jest otrzymany z \mathcal{C} poprzez zamianę zbiorów $C_1, C_2, C_{e_1}, \dots, C_{e_k}$ przez jeden zbiór $C = C_1 \cup C_2 \cup C_{e_1} \cup \dots \cup C_{e_k}$. Z założenia, że \mathcal{C} jest proste mamy, że $|C_{e_i} \cap (C_1 \cup C_2)| = 2$ dla każdego $1 \leq i \leq k$, więc

$$\begin{aligned} \left| C_1 \cup C_2 \cup \bigcup_{i=1}^k C_{e_i} \right| &= \left| C_1 \cup C_2 \cup \bigcup_{i=1}^k (C_{e_i} \setminus (C_1 \cup C_2)) \right| \\ &\leq |C_1 \cup C_2| + \sum_{i=1}^k (|C_{e_i}| - 2) \\ &= k_1 + k_2 + 1 + \sum_{i=1}^k (|C_{e_i}| - 2) \end{aligned}$$

i

$$\frac{1}{p^{|C_1 \cup C_2 \cup \bigcup_{i=1}^k C_{e_i}|}} \leq \frac{1}{p^{k_1+k_2+1} \prod_{i=1}^k p^{|C_{e_i}|-2}}.$$

Zatem

$$\begin{aligned} \frac{\pi(H, \mathcal{C})}{\pi(H, \mathcal{C}')} &\sim \frac{mp^{k_1+1} mp^{k_2+1} \prod_{i=1}^k mp^{|C_{e_i}|}}{mp^{|C_1 \cup C_2 \cup \bigcup_{i=1}^k C_{e_i}|}} = O(1) \frac{m^2 p^{k_1+k_2+2}}{mp^{k_1+k_2+1}} \prod_{i=1}^k \frac{mp^{|C_{e_i}|}}{p^{|C_{e_i}|-2}} = \\ &= O(1) mp (mp^2)^k = O(1) n^{\alpha-a+k(\alpha-2a)} = O(1) n^{g(k)}, \end{aligned}$$

gdzie $g(k) = \alpha - a + k(\alpha - 2a)$. Skoro $2a - \alpha > 0$,

$$\begin{aligned} g(k) \leq 0 &\Leftrightarrow \alpha - a + k(\alpha - 2a) \leq 0 \\ &\Leftrightarrow \alpha - a \leq k(2a - \alpha) \\ &\Leftrightarrow k \geq \frac{\alpha - a}{2a - \alpha} = \frac{\frac{1}{2}\alpha - a + \frac{1}{2}\alpha}{2a - \alpha} = \frac{1}{2}k_0 - \frac{1}{2}. \end{aligned}$$

□

7.3.4. Optymalne pokrycia klikowe

Pokrycie klikowe \mathcal{C} grafu H będziemy nazywać **optymalnym pokryciem klikowym** jeśli $\frac{\pi(H, \mathcal{C}')}{\pi(H, \mathcal{C})} = O(1)$ dla każdego innego pokrycia klikowego \mathcal{C}' grafu H . Zauważmy, że skoro jest skończona liczba pokryć klikowych, to optymalne pokrycie istnieje oraz

$$(7.11) \quad \pi(H) \sim \sum_{\mathcal{C} \text{ optymalne pokrycie klikowe}} \pi(H, \mathcal{C}).$$

Z następnego lematu będzie wynikać, że wszystkie optymalne pokrycia klikowe są proste.

Lemat 7.3. *Niech $mp^2 = o(1)$ oraz \mathcal{C} będzie optymalnym pokryciem klikowym grafu H . Wtedy w \mathcal{C} nie ma dwóch takich zbiorów $C_1, C_2 \in \mathcal{C}$, że $|C_1 \cap C_2| \geq 2$.*

Dowód. Załóżmy nie wprost, że w pokryciu \mathcal{C} istnieją zbiory C_1 i C_2 takie, że $|C_1 \cap C_2| \geq 2$. Stwórzmy nowe pokrycie klikowe \mathcal{C}' z \mathcal{C} zamieniając zbiory C_1 i C_2 przez $C = C_1 \cup C_2$. Wtedy $\pi(H, \mathcal{C}) \sim \pi(H, \mathcal{C}') \frac{mp^{|C_1|} mp^{|C_2|}}{mp^{|C|}} = \pi(H, \mathcal{C}') mp^{|C_1 \cap C_2|} = o(\pi(H, \mathcal{C}'))$, czyli \mathcal{C} nie jest optymalnym pokryciem klikowym. □

Będziemy też wykorzystywać poniższy lemat.

Lemat 7.4.

Niech $mp^2 = o(1)$ oraz \mathcal{C} będzie optymalnym pokryciem klikowym grafu H . Wtedy nie istnieje zbiór $C \in \mathcal{C}$ taki, że $C = C_1 \cup C_2$, $C_1 \cap C_2 = \emptyset$, $C_1, C_2 \neq \emptyset$ i nie ma żadnych krawędzi między C_1 a C_2 w H .

Dowód. Załóżmy, że w \mathcal{C} jest zbiór C taki, że $C = C_1 \cup C_2$, $C_1 \cap C_2 = \emptyset$, $C_1, C_2 \neq \emptyset$ i nie ma żadnej krawędzi między C_1 i C_2 w H .

Jeśli $|C_1| = 1$ i $|C_2| = 1$, to $\mathcal{C}' = \mathcal{C} \setminus \{C\}$ jest pokryciem klikowym H i $\pi(H, \mathcal{C}) = \pi(H, \mathcal{C}') mp^2 = o(\pi(H, \mathcal{C}'))$

Niech $|C_1| = 1$ i $|C_2| \geq 2$. Jeśli \mathcal{C}' jest uzyskana z \mathcal{C} przez zastąpienie C przez C_2 , wtedy \mathcal{C}' jest pokryciem klikowym i $\pi(H, \mathcal{C}) = \pi(H, \mathcal{C}') \frac{mp^{|C_1|}}{mp^{|C_2|}} = o(\pi(H, \mathcal{C}'))$

Niech $|C_1| \geq 2$ i $|C_2| \geq 2$. Jeśli \mathcal{C}' jest otrzymany z \mathcal{C} przez zastąpienie C przez C_1 i C_2 , wtedy $\pi(H, \mathcal{C}) = \pi(H, \mathcal{C}') \frac{mp^{|C_1|}}{m^2 p^{|C|}} = o(\pi(H, \mathcal{C}'))$. \square

Teraz wykorzystamy rezultaty z poprzedniego podrozdziału, aby znaleźć pokrycia optymalne grafów pełnych. Następnie w lemacie 7.6 określimy jak wyglądają optymalne pokrycia klikowe sumy dwóch klik. Powinniśmy wspomnieć tutaj, że jeśli bardzo dokładnie przeanalizuje się dowód wniosku 6 z [40], można wywnioskować z niego lemat 7.5. Przedstawimy tutaj inny sposób udowodnienia tego wyniku z dwóch powodów. Po pierwsze, metody w nim zastosowane będą niezbędne do pokazania kolejnego lematu. Po drugie jest on naturalną konsekwencją przedstawionych wcześniej lematów.

Lemat 7.5. *Niech β będzie kliką na h wierzchołkach. Dla m spełniającego (7.1) i (7.2) oraz p spełniającego (7.3), jedynymi optymalnymi pokryciami klikowymi β są:*

- (i) *pokrycie jednym zbiorem dla $\delta < 1$;*
- (ii) *pokrycie krawędziami oraz pokrycie jednym zbiorem dla $\delta = 1$;*
- (iii) *pokrycie krawędziami dla $\delta > 1$.*

Dowód. Niech \mathcal{C} będzie prostym pokryciem klikowym.

(i) Załóżmy, że $\delta < 1$. Wiemy z (7.10), że $0 < k_0 < h$. Jeśli pokrycie \mathcal{C} składa się ze zbiorów o rozmiarze co najwyżej k_0 wtedy z lematu 7.1 zachodzi $\pi(\beta, \mathcal{C}) = O(\pi(\beta, \mathcal{C}'))$ i $\pi(\beta, \mathcal{C}') = o(\pi(\beta, \mathcal{C}''))$, gdzie \mathcal{C}' jest pokryciem krawędziami oraz \mathcal{C}'' jest pokryciem przez jeden zbiór. Jeśli \mathcal{C} zawiera zbiór C , dla którego $|C| - 1 \geq \max\{k_0, 2\} - 1 > \frac{1}{2}k_0 - \frac{1}{2}$, wtedy jeśli scalimy zbiór C z innymi zbiorami z \mathcal{C} jeden za drugim, wtedy otrzymamy pokrycie jednym zbiorem \mathcal{C}'' oraz z lematu 7.2 mamy $\pi(\beta, \mathcal{C}) = o(\pi(\beta, \mathcal{C}''))$. Dlatego, \mathcal{C}'' jest jedynym optymalnym pokryciem klikowym.

(ii) Załóżmy, że $\delta = 1$. W tym przypadku $k_0 = h$. Wiemy, że jeśli w dowolnym pokryciu klikowym \mathcal{C} grafu β zastąpimy wszystkie zbiory o co najwyżej $h - 1$ wierzchołkach przez pokrycia krawędziami otrzymamy pokrycie klikowe \mathcal{C}' , dla którego $\pi(\beta, \mathcal{C}) = o(\pi(\beta, \mathcal{C}'))$. Dlatego z lematu 7.1 pokrycie krawędziami i pokrycie jednym zbiorem na h wierzchołkach są jedynymi optymalnymi pokryciami klikowymi grafu β .

(iii) Załóżmy, że $\delta > 1$. Wtedy $k_0 > h$ zatem z lematu 7.1, dla dowolnego \mathcal{C} różnego od pokrycia krawędziami, $\pi(\beta, \mathcal{C}) = o(\pi(\beta, \mathcal{C}'))$ gdzie \mathcal{C}' jest pokryciem krawędziami. \square

Niech teraz β i γ będą klikami przecinającymi się na co najmniej jednym wierzchołku. Będziemy chcieli znaleźć co najmniej jedno optymalne pokrycie klikowe grafu $\beta \cup \gamma$ aby oszacować rząd wielkości wyrażenia (7.6) występującego w (7.7).

Będziemy mówić, że pokrycie klikowe \mathcal{C}' grafu H jest **co najmniej tak optymalne** jak pokrycie klikowe \mathcal{C} grafu H , jeśli $\pi(H, \mathcal{C}) = O(\pi(H, \mathcal{C}'))$. Wtedy możemy zamienić \mathcal{C} na \mathcal{C}' **bez utraty optymalności**.

Lemat 7.6. Niech β i γ będą pełnymi grafami na h wierzchołkach takimi, że $V(\beta) \cap V(\gamma) \neq \emptyset$, m spełnia (7.1) i (7.2) oraz p spełnia (7.3). Wtedy pewne optymalne pokrycie klikowe $\beta \cup \gamma$:

(i) jest pokryciem klikowym zawierającym co najmniej jeden zbiór C , dla którego $C \supseteq V(\beta)$ lub $C \supseteq V(\gamma)$ a poza tym tylko zbiory dwuelementowe dla $\delta \leq 1$.

(ii) jest albo pokryciem krawędziami albo pokryciem jednym zbiorem dla $\delta > 1$.

Dowód. Niech \mathcal{C} będzie dowolnym pokryciem klikowym grafu $\beta \cup \gamma$. Załóżmy dodatkowo, że \mathcal{C} jest proste oraz każdy zbiór z \mathcal{C} jest albo całkowicie zawarty w jednym ze zbiorów $V(\beta)$, $V(\gamma)$ lub ma co najmniej jeden element w $V(\beta) \cap V(\gamma)$. W innym przypadku z lematów 7.3 i 7.4 wynika, że pokrycie \mathcal{C} można zamienić na bardziej optymalne pokrycie spełniające te warunki. Udowodnimy, że bez utraty optymalności możemy zastąpić pokrycie \mathcal{C} pokryciem opisanym w (i) lub pokryciem opisanym w (ii).

Jeśli wszystkie zbiory w pokryciu \mathcal{C} mają moc co najwyżej k_0 , wtedy z lematu 7.1 pokrycie krawędziami grafu $\beta \cup \gamma$ jest co najmniej tak optymalne jak \mathcal{C} .

Jeśli w \mathcal{C} jest co najmniej jeden zbiór o mocy $|C| \geq \lfloor k_0 \rfloor + 1$, wtedy albo ten zbiór jest cały zawarty w $V(\beta)$ lub $V(\gamma)$ albo ma co najmniej jeden wierzchołek w $V(\beta) \cap V(\gamma)$. W drugim przypadku co najmniej $\frac{|C|-1}{2} + 1$ wierzchołków zbioru C jest w jednym ze zbiorów $V(\beta)$ i $V(\gamma)$. Dlatego bez utraty ogólności możemy założyć, że $|C \cap V(\gamma)| \geq \frac{\lfloor k_0 \rfloor}{2} + 1 \geq \frac{k_0 - 1}{2} + 1$ tzn. $|C \cap V(\gamma)| - 1 \geq \frac{k_0}{2} - \frac{1}{2}$. Dla dowolnego wierzchołka $v \in V(\gamma)$ spoza C , istnieje zbiór C_1 zawierający v oraz jeden wierzchołek z C . W dodatku, skoro $|C \cap V(\gamma)| - 1 \geq \frac{k_0}{2} - \frac{1}{2}$, jest co najmniej $\frac{k_0}{2} - \frac{1}{2}$ krawędzi w $\beta \cup \gamma$ łączących $v \in C_1$ z wierzchołkami z C . Z lematu 7.2 bez utraty optymalności możemy scalić C_1 , C oraz wszystkie zbiory pokrycia zawierające co najmniej jedną krawędź łączącą wierzchołki z C_1 i C . Bez utraty optymalności pokrycia można powtarzać tę procedurę aż do momentu, gdy w pokryciu klikowym mamy jeden duży zbiór zawierający wszystkie wierzchołki z $V(\gamma)$. Tak uzyskane pokrycie klikowe jest co najmniej tak optymalne jak \mathcal{C} . Dla wszystkich zbiorów z \mathcal{C} o rozmiarze większym niż k_0 , jeśli nie zostały scalone wcześniej, można powtórzyć tę procedurę tak długo jak można, ostatecznie uzyskując pokrycie klikowe co najmniej tak optymalne jak pierwotne. Tak utworzone pokrycie klikowe zawiera co najmniej jeden zbiór zawierający $V(\beta)$ lub $V(\gamma)$. Ostatecznie, wszystkie pozostałe zbiory o rozmiarze co najwyżej k_0 , jeśli takie nadal istnieją w zmodyfikowanym pokryciu i nie mogą już być scalone, zgodnie z lematem 7.1 mogą bez utraty optymalności być zamienione na odpowiednie pokrycia krawędziami.

Z powyższych rozważań dowolne pokrycie klikowe można zastąpić bez utraty optymalności na pokrycie krawędziami lub pokrycie zawierające co najmniej jeden zbiór zawierający $V(\beta)$ lub $V(\gamma)$ i zbiory dwuelementowe poza tym. Wystarczy zatem rozważyć tylko poniższe przypadki.

(i) Załóżmy teraz, że $\delta \leq 1$ i \mathcal{C} jest pokryciem krawędziami. W tym przypadku, jeśli skorzystamy z drugiej części lematu 7.1, możemy udowodnić,

że pokrycie składające się z jednego zbioru $C = V(\gamma)$ oraz wszystkich innych zbiorów 2-elementowych jest co najmniej tak optymalne jak \mathcal{C} .

(ii) Załóżmy teraz, że $\delta > 1$ i \mathcal{C} jest pokryciem zawierającym co najmniej jeden zbiór zawierający $V(\beta)$ lub $V(\gamma)$ i zbiory dwuelementowe poza tym. Z lematu 7.1 proste pokrycie klikowe dwoma zbiorami C_1 i C_2 zawierającymi, odpowiednio, $V(\beta)$ i $V(\gamma)$, bez utraty optymalności może być zastąpione przez pokrycie krawędziami o ile $|C_1| < k_0$ i $|C_2| < k_0$. Dodatkowo, skoro $|V(\beta \cup \gamma)| \leq 2h - 1$, $k_0 > h$ i rozważamy tylko proste pokrycia klikowe, możemy bez utraty ogólności założyć, że \mathcal{C} zawiera jeden duży zbiór C taki, że $C \supseteq V(\gamma)$ a poza tym tylko zbiory dwuelementowe. Niech $x = |C \cap V(\beta)|$ oraz

$$(7.12) \quad s = h - |V(\beta) \cap V(\gamma)|.$$

Jeśli $x = 1$, to $s = h - 1$ oraz $C = V(\gamma)$. W tym przypadku z lematu 7.1 wnioskujemy, że pokrycie krawędziami jest co najmniej tak optymalne jak \mathcal{C} . Załóżmy teraz, że $x \geq 2$ (tzn. $\max\{h - s, 2\} \leq x \leq h$). Skoro pokrycie krawędziami jest optymalne dla klikki na h wierzchołkach, $\pi_\gamma \sim (mp^2)^{\binom{h}{2}}$ i

$$\pi(\beta \cup \gamma, \mathcal{C}) \sim (mp^2)^{\binom{h}{2} - \binom{x}{2}} mp^{s+x} \sim \pi_\gamma \frac{mp^x}{(mp^2)^{\binom{x}{2}}} p^s.$$

Dla $s = h - 1$, skoro $k_0 > h$, z dowodu lematu 7.1 możemy wywnioskować, że $mp^x / (mp^2)^{\binom{x}{2}}$ jest największe rzędowo dla $x = 2$. Przypomnijmy, że z (7.5) $mp \rightarrow \infty$ dla $\delta > 1$. W związku z tym z lematu 7.1 wynika, że dla $x = 2$

$$\begin{aligned} \pi(\beta \cup \gamma, \mathcal{C}) &\sim (mp^2)^{\binom{h}{2}} p^{h-1} = O\left((mp^2)^{\binom{h}{2}} \cdot mp \cdot p^{h-1}\right) \\ &= O\left((mp^2)^{2\binom{h}{2}} \frac{mp^h}{(mp^2)^{\binom{h}{2}}}\right) \\ &= O\left((mp^2)^{2\binom{h}{2}}\right) = O((\pi(\beta \cup \gamma), \mathcal{C}')), \end{aligned}$$

gdzie \mathcal{C}' jest pokryciem krawędziami. Dlatego pokrycie krawędziami jest co najmniej tak optymalne jak \mathcal{C} . Dla $1 \leq s \leq h - 2$, mamy $h - s \leq x \leq h$ oraz z dowodu lematu 7.1 wnioskujemy, że $mp^x / (mp^2)^{\binom{x}{2}}$ jest największe rzędowo dla $x = h - s$ lub $x = h$. Jeśli $x = h - s$, wtedy $C = V(\gamma)$ oraz z lematu 7.1 pokrycie krawędziami jest co najmniej tak optymalne jak \mathcal{C} . Jeśli $x = h$, to \mathcal{C} jest pokryciem jednym zbiorem. \square

7.3.5. Prawdopodobieństwo istnienie klik i ich sum jako podgrafów

W tym podrozdziale zajmiemy się oszacowaniem wielkości π_β oraz $\mathbb{E}(I_\beta I_\gamma)$ występujących w (7.7).

Lemat 7.7. *Załóżmy, że m spełnia (7.1) i (7.2) oraz p spełnia (7.3). Wtedy, dla każdego $\beta \in \Gamma$,*

$$(7.13) \quad \pi_\beta \sim \begin{cases} \frac{c^h}{n^h} & \text{dla } \delta < 1; \\ \frac{c^h + c^{h(h-1)}}{n^h} & \text{dla } \delta = 1; \\ \frac{c^{h(h-1)}}{n^h} & \text{dla } \delta > 1. \end{cases}$$

Zdefiniujmy λ jako

$$(7.14) \quad \lambda_n := \mathbb{E}X_n = \binom{n}{h} \pi_\beta.$$

Wtedy,

$$(7.15) \quad \lim_{n \rightarrow \infty} \lambda_n = \begin{cases} \frac{c^h}{h!} & \text{dla } \delta < 1; \\ \frac{c^h + c^{h(h-1)}}{h!} & \text{dla } \delta = 1; \\ \frac{c^{h(h-1)}}{h!} & \text{dla } \delta > 1. \end{cases}$$

Dowód. Niech β będzie kliką na h wierzchołkach. Z lematu 7.5, (7.9) i (7.11) wynika, że:

- dla $\delta < 1$ i \mathcal{C} pokrycia jednym zbiorem

$$\pi_\beta \sim \pi(\beta, \mathcal{C}) \sim mp^h \sim \frac{c^h}{n^h};$$

- dla $\delta = 1$ i \mathcal{C} pokrycia jednym zbiorem oraz \mathcal{C}' pokrycia krawędziami grafu β

$$\pi_\beta \sim \pi(\beta, \mathcal{C}) + \pi(\beta, \mathcal{C}') \sim mp^h + (mp^2)^{\binom{h}{2}} \sim \frac{c^h + c^{h(h-1)}}{n^h};$$

- dla $\delta > 1$
i \mathcal{C} pokrycia krawędziami

$$\pi_\beta \sim \pi(\beta, \mathcal{C}) \sim (mp^2)^{\binom{h}{2}} \sim \frac{c^{h(h-1)}}{n^h}.$$

Granice w (7.15) wynikają bezpośrednio z (7.14) oraz (7.13). \square

Lemat 7.8.

Załóżmy, że m spełnia (7.1) oraz (7.2) i że p spełnia (7.3). Zdefiniujmy s jak w (7.12). Wtedy, dla każdego $\beta, \gamma \in \Gamma$,

$$\Pr\{\beta \subseteq G^{Bin}(\mathcal{V}, m, p), \gamma \subseteq G^{Bin}(\mathcal{V}, m, p)\} = \begin{cases} O(\pi^2 + \pi p^{h-1}) & \text{jeśli } \delta \leq 1 \text{ i } s = h - 1; \\ O(\pi p^s) & \text{jeśli } \delta \leq 1 \text{ i } s < h - 1; \\ O(\pi^2 + mp^{2h-1}) & \text{jeśli } \delta > 1 \text{ i } s = h - 1; \\ O\left(mp^{h+s} + \pi^2 n^{\frac{(h-s)(h-s-1)}{h-1}}\right) & \text{jeśli } \delta > 1 \text{ i } s < h - 1, \end{cases}$$

gdzie $\pi = \pi_\zeta$ dla dowolnego $\zeta \in \Gamma$.

Dowód. Niech \mathcal{C} będzie pokryciem klikowym $\beta \cup \gamma$ opisanym w lemacie 7.6 oraz niech C będzie największą kliką w \mathcal{C} .

Załóżmy najpierw, że $\delta \leq 1$:

PRZYPADEK 1. $s = h - 1$ i $|C \cap V(\beta)| \leq 1$, co implikuje $|V(\beta) \cap V(\gamma)| = 1$.

Skoro C nie pokrywa żadnej krawędzi z β , zatem $\mathcal{C} \setminus \{C\}$ jest pokryciem grafu β . Czyli

$$\pi(\beta \cup \gamma, \mathcal{C}) = O(1)\pi^2.$$

PRZYPADEK 2. $s < h - 1$ lub $|C \cap V(\beta)| \geq 2$. Bez utraty ogólności załóżmy, że $V(\gamma) \subseteq C$. Oczywiście $\mathcal{C}[\beta] = \{C' \cap V(\beta) : C' \in \mathcal{C}\}$ jest pokryciem klikowym grafu β oraz

$$\pi(\beta \cup \gamma, \mathcal{C}) = \pi(\beta, \mathcal{C}[\beta])p^s = O(1)\pi p^s.$$

Założmy dalej, że $\delta > 1$.

PRZYPADEK 1. $|C| = 2$, czyli \mathcal{C} jest pokryciem krawędziami. Dla $s = h - 1$

$$\pi(\beta \cup \gamma, \mathcal{C}) = O(1)\pi^2.$$

Dla $s < h - 1$ wykorzystując (7.3) otrzymujemy.

$$\pi(\beta \cup \gamma, \mathcal{C}) \sim \pi^2(mp^2)^{-\binom{h-s}{2}} \asymp \pi^2 n^{\frac{(h-s)(h-s-1)}{h-1}}.$$

PRZYPADEK 2. $|C| = h + s$. Ponieważ jest to pokrycie jednym zbiorem, mamy

$$\pi(\beta \cup \gamma, \mathcal{C}) \sim mp^{h+s}.$$

□

7.3.6. Dowód twierdzenia 7.1

Dowód twierdzenia 7.1. Wykorzystamy (7.7) z grafem zależności L zdefiniowanym w podrozdziale 7.3.1. Zauważmy, że

$$\sum_{\beta \in V(L)} \pi_\beta^2 \leq \sum_{\beta \gamma \in E(L)} \pi_\beta \pi_\gamma,$$

dlatego musimy tylko oszacować poniższe sumy:

$$\sum_{\beta \gamma \in E(L)} \pi_\beta \pi_\gamma$$

i

$$(7.16) \quad \sum_{\beta \in V(L)} \sum_{s=1}^{h-1} \sum_{\substack{\gamma \in V(L) \\ |V(\beta) \cap V(\gamma)|=h-s}} \Pr\{\beta \subseteq G^{Bin}(\mathcal{V}, m, p), \gamma \subseteq G^{Bin}(\mathcal{V}, m, p)\}.$$

Zauważmy, że dla danego $\beta \in V(L)$,

$$|\{\gamma \in V(L) : |V(\beta) \cap V(\gamma)| = h - s\}| = O(n^s),$$

$$|\{\gamma \in V(L) : |V(\beta) \cap V(\gamma)| \neq 0\}| = O(n^{h-1})$$

i

$$|E(L)| = |V(L)| \cdot |\{\gamma \in V(L) : |V(\beta) \cap V(\gamma)| \neq 0\}| = O(n^{2h-1}).$$

Zatem, skoro $n^h \pi = O(1)$, więc

$$\sum_{\beta, \gamma \in E(L)} \pi_\beta \pi_\gamma = O(n^{2h-1} \pi^2) = O(n^{-1}).$$

Teraz wykorzystamy lemat 7.8 do oszacowania (7.16). Dla $\delta \leq 1$, zachodzi $np \asymp m^{-\frac{1}{h}} = n^{-\frac{2\delta}{h-1}}$. Jeśli $\delta \leq 1$ i $s = h - 1$, to

$$\begin{aligned} \sum_{\beta \in V(L)} \sum_{\substack{\gamma \in V(L) \\ |V(\beta) \cap V(\gamma)| = h-s}} \Pr\{\beta \subseteq G^{Bin}(\mathcal{V}, m, p), \gamma \subseteq G^{Bin}(\mathcal{V}, m, p)\} \\ = O(n^{2h-1} \pi^2 + \pi n^{2h-1} p^{h-1}) \\ = O(n^{-1} + (np)^{h-1}) \\ = O(n^{-1} + n^{-2\delta}). \end{aligned}$$

Jeśli $\delta \leq 1$ i $1 \leq s < h - 1$, to

$$\begin{aligned} \sum_{\beta \in V(L)} \sum_{\substack{\gamma \in V(L) \\ |V(\beta) \cap V(\gamma)| = h-s}} \Pr\{\beta \subseteq G^{Bin}(\mathcal{V}, m, p), \gamma \subseteq G^{Bin}(\mathcal{V}, m, p)\} \\ = O(n^h \pi n^s p^s) \\ = O(n^{-\frac{2\delta s}{h-1}}) \\ = O(n^{-\frac{2\delta}{h-1}}). \end{aligned}$$

Dla $\delta > 1$ można sprawdzić, że $np \asymp n^{\frac{h-2}{h-1}} m^{-\frac{1}{2}} = n^{\frac{(1-\delta)h-2}{h-1}} = o(1)$. Jeśli $\delta > 1$ i $s = h - 1$, to

$$\begin{aligned} \sum_{\beta \in V(L)} \sum_{\substack{\gamma \in V(L) \\ |V(\beta) \cap V(\gamma)| = h-s}} \Pr\{\beta \subseteq G^{Bin}(\mathcal{V}, m, p), \gamma \subseteq G^{Bin}(\mathcal{V}, m, p)\} = \\ = O(n^{2h-1} \pi^2 + n^{2h-1} m p^{2h-1}) = \\ = O(n^{-1} + n^{h+1} m p^{h+1}) = \\ = O(n^{-1} + (np)^{h+1} m) = \\ = O\left(n^{-1} + n^{\frac{(h+1)((1-\delta)h-2)+2\delta h}{h-1}}\right) = \\ = O\left(n^{-1} + n^{((1-\delta)h - \frac{2}{h-1})}\right). \end{aligned}$$

Jeśli $\delta > 1$ i $1 \leq s < h - 1$, to

$$\begin{aligned}
\sum_{\beta \in V(L)} \sum_{\substack{\gamma \in V(L) \\ |V(\beta) \cap V(\gamma)| = h-s}} \Pr\{\beta \subseteq G^{Bin}(\mathcal{V}, m, p), \gamma \subseteq G^{Bin}(\mathcal{V}, m, p)\} &= \\
&= O\left(n^{h+s} m p^{h+s} + n^{2h} \pi^2 n^{s-h} n^{\frac{(h-s)(h-s-1)}{h-1}}\right) = \\
&= O\left(n^{h+1} m p^{h+1} + n^{\frac{(h-s)(h-s-1) - (h-s)(h-1)}{h-1}}\right) = \\
&= O\left(n^{((1-\delta)h - \frac{2}{h-1})} + n^{\frac{-(h-s)s}{h-1}}\right) = \\
&= O\left(n^{((1-\delta)h - \frac{2}{h-1})} + n^{-1}\right).
\end{aligned}$$

Twierdzenie 7.1 wynika z (7.7), powyższych oszacowań i z (7.15). \square

8. Średnica

Poniższy rozdział jest poświęcony określeniu długości średnicy największej składowej spójności w losowym grafie $G^{Jed}(\mathcal{V}, m, d)$. Będzie nas interesować składowa zawierająca $\Omega(n)$ wierzchołków, gdyż to ona stanowi część sieci sensorowej modelowanej przez $G^{Jed}(\mathcal{V}, m, d)$, w której znajdują się sensory komunikujące się nawzajem. W sieci sensorowej odpowiadającej grafowi $G^{Jed}(\mathcal{V}, m, d)$ długość średnicy największej składowej jest maksymalną liczbą „skoków”, jakie musi zrobić wysyłana informacja, żeby dotrzeć do każdego sensora połączonego w sieci. Na razie nie będziemy się skupiać na problemie istnienia składowej o $\Omega(n)$ wierzchołkach. Odpowiedź na pytanie, czy asymptotycznie prawie na pewno taka składowa istnieje znajduje się w rozdziale 10. Poniższe rozważania można potraktować jako wstęp do dowodu istnienia tej składowej i będą one przydatne w rozdziale 10 przy określaniu funkcji progowej dla spójności i przejścia fazowego w $G^{Jed}(\mathcal{V}, m, d)$. Wynik ten można uogólnić dla grafu $G(\mathcal{V}, m, \mathcal{P}_{(m)})$ o rozkładach $\mathcal{P}_{(m)}$, w których wartości P_i są niezerowe tylko dla skończonej liczby indeksów i . W tym celu można zastosować metody zaproponowane przez Bloznelisa w [9, 10, 11]. Jednakże bardziej ogólny wynik, na przykład wynik dla rozkładów zaproponowanych w rozdziale 5, jest trudny do pokazania, gdyż dla grafu niespójnego posiadanie średnicy o długości co najmniej k nie jest własnością monotoniczną. Wspominają o tym Bollobás, Janson i Riordan w [14] w uwadze 14.10.

8.1. Twierdzenie o długości średnicy w $G^{Jed}(\mathcal{V}, m, d)$

Twierdzenie 8.1. *Niech $c > 0$ będzie stałą, $d = d(n) \geq 2$ ciągiem liczb całkowitych oraz G będzie składową spójności grafu $G^{Jed}(\mathcal{V}, m, d)$ zawierającą $\Omega(n)$ wierzchołków.*

(i) *Jeśli*

$$\frac{d^2 n}{m \ln n} \sim c,$$

to asymptotycznie prawie na pewno

$$\text{diam}(G) \sim \frac{\ln n}{\ln \ln n}.$$

(ii) *Jeśli $c > 1$ i*

$$\frac{d(d-1)n}{m} \sim c,$$

to asymptotycznie prawie na pewno

$$\text{diam}(G) \asymp \ln n.$$

W dowodach twierdzeń będziemy korzystali z własności drzewa BFS, które powstaje w wyniku procedury przeszukiwania grafu wszerz (patrz np. [18]). W trakcie procedury działającej na danym grafie, rozpoczętej w danym wierzchołku tego grafu (zwanym korzeniem drzewa) powstaje poddrzewo badanego grafu, w którym zbiór wierzchołków w odległości i od korzenia pokrywa się ze zbiorem wierzchołków w odległości i od korzenia w rozpatrywanym grafie. Dowód twierdzenia jest oparty na kilku spostrzeżeniach dotyczących drzewa BFS powstałego w wyniku procedury działającej na $G^{Jed}(\mathcal{V}, m, d)$ rozpoczętej w wierzchołku $v \in \mathcal{V}$. Dokładniej rzecz ujmując dowód bazuje na oszacowaniach wielkości zbiorów wierzchołków w odległości i od v . Najpierw znajdziemy oszacowanie dolne i górne na prawdopodobieństwo dołączenia nowego wierzchołka do drzewa BFS, przy założeniu, że nie zawiera ono dużej liczby wierzchołków. Następnie pokażemy, że dla interesujących nas przypadków grafu $G^{Jed}(\mathcal{V}, m, d)$, proces BFS asymptotycznie prawie na pewno skończy się w niewielkiej liczbie kroków lub można jego kolejne kroki przybliżyć pewnymi procesami gałązkowymi. W końcu, wykorzystując przybliżenie procesami gałązkowymi, udowodnimy, że asymptotycznie prawie na pewno, dla dowolnych dwóch wierzchołków procesy BFS rozpoczynające się w tych wierzchołkach kończą się szybko lub ich drzewa BFS łączą się.

Idea dowodu bazuje na pracy [16] Chung i Lu, która dotyczyła średnicy losowego grafu $G(n, p)$. W przypadku grafu $G^{Jed}(\mathcal{V}, m, d)$ dowód jest bardziej skomplikowany ze względu na zależności między powstającymi krawędziami. Konstruując drzewo BFS musimy, nie tylko kontrolować liczbę wierzchołków dołączonych do drzewa, lecz także liczbę własności jakie te wierzchołki posiadają.

8.2. Drzewo BFS w grafie $G^{Jed}(\mathcal{V}, m, d)$

W podrozdziale tym będziemy rozważać własności drzewa BFS o korzeniu w danym wierzchołku $v \in \mathcal{V}$ w grafie $G^{Jed}(\mathcal{V}, m, d)$. Mając na uwadze strukturę losowego grafu przecięć, w celu ułatwienia dalszej analizy problemu wprowadzimy bardzo niestandardową definicję drzewa BFS, która jednak jest równoważna definicji klasycznej. Mianowicie skonstruowane drzewo BFS będzie podgrafem pewnego nieskończonego, skierowanego, oznaczonego drzewa T .

Dla ustalonego uporządkowania zbioru \mathcal{V} , drzewo T będzie miało następujące własności:

- (i) korzeniem drzewa T jest kopia wierzchołka v ;
- (ii) zbiór następników każdego wierzchołka jest kopią zbioru \mathcal{V} ;
- (iii) każdy wierzchołek (poza korzeniem) ma stopień wejścia 1 a korzeń ma stopień wejścia 0;
- (iv) indeksowanie wierzchołków jest zgodne z ich uporządkowaniem w \mathcal{V} (tzn. dla dowolnego wierzchołka $w \in \mathcal{V}$ i dowolnych dwóch jego następników u_1 i u_2 , u_1 ma mniejszy indeks niż u_2 wtedy i tylko wtedy gdy jest przed u_2 w porządku ustalonym w \mathcal{V}).

Procedurę BFS w $G^{Jed}(\mathcal{V}, m, d)$ opiszemy jako konstrukcję losowego poddrzewa drzewa T . To losowe poddrzewo będzie izomorficzne z klasycznym drzewem BFS. Procedura będzie podzielona na kroki a każdy krok będzie się składał z $|\mathcal{V}|$ podkroków. Najpierw dodajemy do poddrzewa korzeń (kopię wierzchołka $v = v_1$). W pierwszym podkroku pierwszego kroku bierzemy pierwszy (zgodnie z porządkiem ustalonym w \mathcal{V}) następnik korzenia w T i dodajemy go do poddrzewa wtedy i tylko wtedy gdy jest on sąsiadem wierzchołka v w grafie $G^{Jed}(\mathcal{V}, m, d)$. W następnych podkrokach pierwszego kroku, kolejno dla każdego następnika korzenia, postępujemy tak samo. Po $|\mathcal{V}|$ podkrokach dodajemy v do zbioru rozpatrzonych wierzchołków. W następnym kroku bierzemy wierzchołek u dodany do drzewa jako drugi (oznaczymy kopię wierzchołka u w \mathcal{V} przez $v_2, v_2 \in \mathcal{V}$). W $|\mathcal{V}|$ podkrokach tego kroku dodajemy jeden po drugim do poddrzewa następniki wierzchołka u w T wtedy i tylko wtedy gdy są kopiami sąsiadów wierzchołka v_2 w $G^{Jed}(\mathcal{V}, m, d)$ oraz ich kopie nie zostały dodane do poddrzewa wcześniej. Na zakończenie tego kroku dodajemy v_2 do zbioru wierzchołków rozpatrzonych. W kolejnym kroku bierzemy kolejny wierzchołek dodany do poddrzewa i postępujemy analogicznie do przypadku wierzchołka u . Procedura kończy się w momencie, gdy zbiór wierzchołków drzewa jest równoliczny ze zbiorem wierzchołków rozpatrzonych.

Założmy teraz, że będziemy przeprowadzać procedurę BFS do momentu, gdy zbiór wierzchołków drzewa jest równoliczny ze zbiorem wierzchołków rozpatrzonych (tzn. nie będzie już można dodać nowego wierzchołka do drzewa) lub drzewo BFS będzie miało $n_{max} = n/(\ln n)^{4/3}$ wierzchołków. Oszacujemy teraz prawdopodobieństwo dodania nowego wierzchołka do drzewa BFS.

Po t ($t \geq 0$) krokach i t' podkrokach kroku $t + 1$ zbiór wierzchołków rozpatrzonych jest mocy t . Założmy, że w tym momencie drzewo BFS zawiera jeszcze kopie s innych, nierozpatrzonych, wierzchołków. Oznaczmy zbiór wszystkich wierzchołków $G^{Jed}(\mathcal{V}, m, d)$, których kopie zostały dodane do drzewa BFS do tego momentu przez $V_{BFS} = \{v_1, \dots, v_t, v_{t+1}, \dots, v_{t+s}\}$ (indeksy wierzchołków są zgodne z porządkiem dodawania ich kopii do drzewa BFS). Z poprzednich założeń $t+s \leq n_{max}$, gdyż w przeciwnym przypadku zakończylibyśmy już procedurę. Ustalmy teraz oznaczenia:

$$(8.1) \quad \begin{aligned} B &= \emptyset, & |B| &= 0, & \text{dla } t &= 0; \\ B &= D(\{v_1, \dots, v_t\}), & |B| &\leq dt, & \text{dla } t &> 0; \\ A &= D(\{v_{t+1}, \dots, v_{t+s}\}), & |A| &\leq ds, & \text{dla } t &\geq 0, \end{aligned}$$

gdzie $D(S)$ oznacza zbiór wszystkich własności wierzchołków ze zbioru S , tzn.:

$$(8.2) \quad D(S) = \bigcup_{v \in S} D(v).$$

Oznaczmy także dla $t \geq 0$

$$(8.3) \quad D^*(v_{t+1}) = D(v_{t+1}) \setminus B = D(v_{t+1}) \setminus D(\{v_1, \dots, v_t\}).$$

Zauważmy, że zgodnie z powyższą definicją, jeśli w tym momencie do drzewa BFS zostanie dodana kopia wierzchołka $v' = v_{t+s+1}$, to

$$(8.4) \quad D^*(v') = D(v') \setminus D(\{v_1, \dots, v_{t+s}\}).$$

Kopia następnego (zgodnie z porządkiem ustalonym w \mathcal{V}) wierzchołka ze zbioru $\mathcal{V} \setminus V_{BFS}$ dodana jest do drzewa BFS z prawdopodobieństwem

$$1 - \binom{m-|B|-|D^*(v_{t+1})|}{d} / \binom{m-|B|}{d}.$$

Z oszacowania

$$(8.5) \quad \begin{aligned} \frac{\binom{x-y}{z}}{\binom{x}{z}} &= \frac{(x-y)_z}{(x)_z} = \prod_{i=0}^{z-1} \left(1 - \frac{y}{x-i}\right) \geq 1 - \sum_{i=0}^{z-1} \frac{y}{x-i} = \\ &= 1 - \sum_{i=0}^{z-1} \left(\frac{y}{x-i} - \frac{y}{x}\right) - \frac{zy}{x} = 1 - \frac{zy}{x} - \sum_{i=0}^{z-1} \frac{yi}{x(x-i)} \geq \\ &\geq 1 - \frac{zy}{x} - \frac{y}{x(x-z)} \cdot \frac{z(z-1)}{2} \geq 1 - \frac{zy}{x} - \frac{z^2y}{2x(x-z)} \end{aligned}$$

i faktu, że $|D^*(v_1)| \leq d$, $|D^*(v_{t+1})| \leq d-1$ (dla $t > 0$) i $|B| \leq dn_{max}$ uzyskujemy

$$\begin{aligned} 1 - \frac{\binom{m-|B|-|D^*(v_{t+1})|}{d}}{\binom{m-|B|}{d}} &\leq \frac{d|D^*(v_{t+1})|}{m-|B|} \left(1 + \frac{d}{2(m-|B|-d)}\right) \leq \\ &\leq \begin{cases} \frac{d^2}{m} \left(1 + \frac{dn_{max}}{m-dn_{max}} + \frac{d}{2(m-dn_{max}-d)} + \frac{d^2n_{max}}{2(m-dn_{max})(m-dn_{max}-d)}\right) & \text{dla } t = 0; \\ \frac{d(d-1)}{m} \left(1 + \frac{dn_{max}}{m-dn_{max}} + \frac{d}{2(m-dn_{max}-d)} + \frac{d^2n_{max}}{2(m-dn_{max})(m-dn_{max}-d)}\right) & \text{dla } t > 0. \end{cases} \end{aligned}$$

Oznaczmy wartości oszacowania górnego na prawdopodobieństwo dodania wierzchołka do drzewa BFS przez

$$(8.6) \quad \begin{aligned} P'_+ &= \frac{d^2}{m} \left(1 + \frac{dn_{max}}{m-dn_{max}} + \frac{d}{2(m-dn_{max}-d)} + \frac{d^2n_{max}}{2(m-dn_{max})(m-dn_{max}-d)}\right); \\ P_+ &= \frac{d(d-1)}{m} \left(1 + \frac{dn_{max}}{m-dn_{max}} + \frac{d}{2(m-dn_{max}-d)} + \frac{d^2n_{max}}{2(m-dn_{max})(m-dn_{max}-d)}\right). \end{aligned}$$

Dodatkowo zauważmy, że podczas każdego kroku i podkroku procedury zachodzi $|\mathcal{V} \setminus V_{BFS}| \leq n$.

Naszym kolejnym celem jest uzyskanie oszacowania dolnego P_- (odpowiednio P'_-) na prawdopodobieństwo dodania wierzchołka do drzewa BFS, dla którego to oszacowania zachodziłoby $P_- \sim P_+$ (odpowiednio $P'_- \sim P'_+$). Uzyskanie takiego oszacowania dolnego nie jest możliwe w ogólnym przypadku. Problem stanowią wierzchołki, dla których $|D^*(v_{t+1})| < d-1$. Dlatego ograniczymy się do rozpatrywania podgrafu drzewa BFS składającego się z korzenia i z wierzchołków, dla kopii których $|D^*(v_{t+1})| = d-1$. Okazuje się, że liczba wierzchołków poza tym podgrafem będzie asymptotycznie mała w porównaniu do liczby wierzchołków dla których $|D^*(v_{t+1})| = d-1$. O wierzchołkach dla których $|D^*(v_{t+1})| = d-1$ będziemy mówić, że dodają one do zbioru $D(V_{BFS})$ $d-1$ nowych własności.

Jak w przypadku oszacowania górnego zakładamy, że jesteśmy po t krokach ($t \geq 0$) i t' podkrokach kroku $t+1$, $V_{BFS} = \{v_1, \dots, v_t, v_{t+1}, \dots, v_{t+s}\}$ oraz dokładnie t pierwszych wierzchołków zbioru V_{BFS} jest wierzchołkami już rozpatrzonymi. Załóżmy także, że wierzchołek v_{t+1} dodał $d-1$ własności do zbioru $D(V_{BFS})$ lub jest korzeniem. Teraz sprawdzamy, czy kolejny (zgodnie

z uporządkowaniem \mathcal{V}) wierzchołek v' został dodany do V_{BFS} i jego kopia została dodana do rozpatrywanego podgrafu (tzn. czy jest sąsiadem v_{t+1} w $G^{Jed}(\mathcal{V}, m, d)$, którego kopia nie została jeszcze dodana do drzewa BFS i czy

$$|D^*(v')| = |D(v') \setminus (D(\{v_1, \dots, v_{s+t}\}))| = d - 1.$$

Skoro $|D^*(v_1)| = d$ oraz, dla $t > 0$, $|D^*(v_{t+1})| = d - 1$ (zakładaliśmy, że kopia v_{t+1} została dodana do rozpatrywanego podgrafu drzewa BFS), więc prawdopodobieństwo, że w tym podkroku kopia v' jest dodana do podgrafu drzewa BFS wynosi $\binom{|D^*(v_{t+1})|}{1} \binom{m-|B|-|A|}{d-1} / \binom{m-|B|}{d}$. Zatem z faktu, że $|A| + |B| \leq dn_{max}$ i z oszacowania (8.5) wynika, że

$$\begin{aligned} & \frac{\binom{|D^*(v_{t+1})|}{1} \binom{m-|B|-|A|}{d-1}}{\binom{m-|B|}{d}} = \frac{|D^*(v_{t+1})|d}{(m-|B|-d+1)} \frac{\binom{m-|B|-|A|}{d-1}}{\binom{m-|B|}{d-1}} \geq \\ & \geq \frac{|D^*(v_{t+1})|d}{m} \left(1 - \frac{(d-1)|A|}{m-|B|} - \frac{(d-1)^2|A|}{2(m-|B|)(m-|B|-d+1)} \right) \geq \\ & \geq \begin{cases} \frac{d^2}{m} \left(1 - \frac{d^2 n_{max}}{m-dn_{max}} - \frac{d^3 n_{max}}{2(m-dn_{max})(m-dn_{max}-d)} \right) & \text{dla } t = 0; \\ \frac{(d-1)d}{m} \left(1 - \frac{d^2 n_{max}}{m-dn_{max}} - \frac{d^3 n_{max}}{2(m-dn_{max})(m-dn_{max}-d)} \right) & \text{dla } t > 0. \end{cases} \end{aligned}$$

Czyli możemy oszacować z dołu prawdopodobieństwo dodania wierzchołka do podgrafu drzewa BFS przez

$$(8.7) \quad \begin{aligned} P'_- &= \frac{d^2}{m} \left(1 - \frac{d^2 n_{max}}{m-dn_{max}} - \frac{d^3 n_{max}}{2(m-dn_{max})(m-dn_{max}-d)} \right); \\ P_- &= \frac{(d-1)d}{m} \left(1 - \frac{d^2 n_{max}}{m-dn_{max}} - \frac{d^3 n_{max}}{2(m-dn_{max})(m-dn_{max}-d)} \right). \end{aligned}$$

Dodajmy także, że w każdym podkroku rozpatrywanej procedury BFS z dodatkowym momentem stopu mamy $|\mathcal{V} \setminus V_{BFS}| \geq n - n_{max}$.

8.3. Procesy gałązkowe

Liczba wierzchołków dodanych do V_{BFS} w procedurze BFS rozważanej w poprzednim podrozdziale, w dalszej części rozprawy będzie przybliżona liczbą potomków w procesie gałązkowym. Skupimy się teraz na tych procesach i udowodnimy kilka ich własności. Większość rezultatów umieszczona w tym podrozdziale jest analogiczna do rezultatów otrzymanych w [16]. W związku z tym, że tutaj będą one sformułowane w języku procesów gałązkowych a nie procedury BFS, niektóre szczegóły dowodów, jak i sformułowania lematów będą się różnić od tych zawartych w [16].

W rozdziale tym wszystkie granice będą przy $N \rightarrow \infty$. Asymptotyczna notacja $O(\cdot)$ i $o(\cdot)$ będzie także względem N .

W rozważaniach naszych będziemy zainteresowani uogólnionym procesem gałązkowym \mathcal{B} . W procesie tym liczba dzieci i -tego osobnika procesu będzie zmienną losową X_i . Dokładniej, będziemy się zajmować procesami, które oznaczymy przez $\mathcal{B}(P', P, N)$, w których liczba dzieci pierwszego osobnika procesu (osobnika tego nazwiemy v) ma rozkład dwumianowym $\text{Bin}(N, P')$ a liczba dzieci każdego następnego osobnika (potomka) procesu ma rozkład

dwumianowy $\text{Bin}(N, P)$ ($P' \geq P$) oraz liczby dzieci różnych osobników (potomków) nie zależą od siebie. W celu ułatwienia dalszej analizy opiszemy ten proces w podobny sposób jak proces BFS. To znaczy, przebieg procesu gałązkowego będziemy opisywać jako tworzenie poddrzewa pewnego nieskończonego, skierowanego, oznaczonego drzewa T , którego struktura będzie analogiczna do struktury drzewa T opisanego w poprzednim podrozdziale. A mianowicie T będzie miało następujące własności:

- (i) korzeniem drzewa T jest wierzchołek (osobnik) v ;
- (ii) zbiór następników każdego wierzchołka (osobnika) jest mocy N ;
- (iii) każdy wierzchołek (poza korzeniem) ma stopień wejścia 1 a korzeń ma stopień wejścia 0;

Proces gałązkowy $\mathcal{B}(P', P, N)$ można opisać w następujący sposób. Najpierw bierzemy wierzchołek v i niezależnie, jeden za drugim, zgodnie z indeksowaniem wierzchołków, dodajemy z prawdopodobieństwem P' do poddrzewa każdy z jego N następników. Następniki dodane do drzewa są to dzieci v w procesie. Stanowią one pierwsze pokolenie. Teraz bierzemy pierwsze (zgodnie z indeksowaniem wierzchołków T) dziecko wierzchołka v i dodajemy z prawdopodobieństwem P każdy z jego N następników z T do poddrzewa. Potem postępujemy tak samo z drugim dzieckiem wierzchołka v , a potem kolejno z każdym osobnikiem zgodnie z kolejnością dodawania do poddrzewa.

Krokiem procesu będziemy nazywać dodanie wszystkich dzieci jednego wierzchołka do poddrzewa. Podkrokiem kroku będziemy nazywać dodanie lub nie następnika wierzchołka rozpatrywanego w danym kroku. Potomków, których dzieci już zostały dodane do poddrzewa będziemy nazywać rozpatrzonymi. Będziemy mówić, że proces wymiera, gdy nie ma już nierozpatrzonych potomków.

Zbiór wierzchołków (potomków) w odległości i od wierzchołka v będziemy nazywać i -tym pokoleniem procesu. Liczbę potomków w i -tym pokoleniu procesu $\mathcal{B}(P', P, N)$ oznaczmy przez g_i (tzn. $g_0 = 1$, g_1 ma rozkład dwumianowy $\text{Bin}(N, P')$, itd.).

Udowodnimy teraz kilka lematów dotyczących liczebności kolejnych pokoleń. Niektóre stałe w dowodach mogą być wybrane inaczej, ale podajemy je dla przykładu. We wszystkich dowodach będziemy wykorzystywać poniższe oznaczenie

$$(8.8) \quad \Pr_*\{\cdot\} = \Pr\{\cdot | g_{i-1} = g\}.$$

Lemat 8.1. *Niech $NP = C_N \ln N$, $C_N \rightarrow C$, $i_0 \geq 1$, b_{i_0} będzie dużą stałą. Niech g_i będzie rozmiarem i -tego pokolenia w procesie gałązkowym $\mathcal{B}(P', P, N)$. Jeśli*

$$g_{i_0} \geq b_{i_0},$$

to dla dostatecznie dużych N z prawdopodobieństwem $1 - o\left(\frac{1}{N}\right)$

$$g_i \geq b(C_N \ln N)^{i-i_0} \text{ dla każdego } i \geq i_0,$$

gdzie $b = b_{i_0} - \sqrt{\frac{4b_{i_0}}{C}}$.

Dowód. Niech b_{i_0} będzie dużą stałą. Dla $i \geq i_0 + 1$ zdefiniujmy b_i za pomocą rekurencji

$$(8.9) \quad b_i = b_{i-1} - \frac{\sqrt{3(i-i_0)b_{i-1}}}{\sqrt{C_N} \sqrt{(C_N \ln N)^{i-1-i_0}}}.$$

Niech $b = b_{i_0} - \sqrt{\frac{4b_{i_0}}{C}}$. Wtedy $b_i \leq b_{i_0}$ dla każdego $i \geq i_0$. Ponieważ dla każdego $0 < x < 1$ zachodzi $\sum_{k=0}^{\infty} (k+1)x^k = \frac{1}{(1-x)^2}$ oraz $\frac{C_N \ln N}{\sqrt{C_N}(\sqrt{(C_N \ln N)-1})^2} \rightarrow \frac{1}{\sqrt{C}}$, więc dla każdego $i \geq i_0 + 1$ i dostatecznie dużego N

$$\begin{aligned} b_i &= b_{i-1} - \frac{\sqrt{3(i-i_0)b_{i-1}}}{\sqrt{C_N} \sqrt{(C_N \ln N)^{i-1-i_0}}} = \\ &= b_{i_0} - \sum_{k=i_0}^{i-1} \frac{\sqrt{3(k+1-i_0)b_k}}{\sqrt{C_N} \sqrt{(C_N \ln N)^{k-i_0}}} \geq \\ &\geq b_{i_0} - \sqrt{\frac{3b_{i_0}}{C_N}} \sum_{k=0}^{i-1-i_0} \frac{k+1}{\sqrt{(C_N \ln N)^k}} \geq \\ &\geq b_{i_0} - \sqrt{\frac{3b_{i_0}}{C_N}} \frac{C_N \ln N}{(\sqrt{(C_N \ln N)-1})^2} \geq b_{i_0} - \sqrt{\frac{4b_{i_0}}{C}} = b. \end{aligned}$$

Założmy teraz, że dla danego $i \geq i_0 + 1$, $g_{i-1} = g \geq b_{i-1}(C_N \ln N)^{i-i_0-1}$. Wtedy

$$\mathbb{E} = \mathbb{E}(g_i | g_{i-1} = g) \geq b_{i-1}(C_N \ln N)^{i-i_0}.$$

Z rekurencji (8.9) mamy

$$\frac{(b_{i-1} - b_i)^2}{b_{i-1}} (C_N \ln N)^{i-i_0} = 3(i-i_0) \ln N.$$

Dlatego z nierówności Chernoffa (3.1)

$$\begin{aligned} \Pr_* \left\{ g_i \leq b_i (C_N \ln N)^{i-i_0} \right\} &\leq \Pr_* \left\{ g_i \leq \frac{b_i}{b_{i-1}} \mathbb{E} \right\} = \\ &= \Pr_* \left\{ g_i \leq \mathbb{E} - \frac{(b_{i-1} - b_i)}{b_{i-1}} \mathbb{E} \right\} \leq \\ &\leq \exp \left(-\frac{(b_{i-1} - b_i)^2}{2b_{i-1}^2} \mathbb{E} \right) \leq \\ &\leq \exp \left(-\frac{(b_{i-1} - b_i)^2}{2b_{i-1}} (C_N \ln N)^{i-i_0} \right) \leq \\ &\leq \exp \left(-\frac{3}{2} (i-i_0) \ln N \right) = \left(\frac{1}{N^{\frac{3}{2}}} \right)^{i-i_0}, \end{aligned}$$

gdzie \Pr_* jest zdefiniowane przez (8.8). Zdefiniujmy A_i jako zdarzenie, że $g_i \geq b(C_N \ln N)^{i-i_0}$ i A'_i jako zdarzenie, że $g_i \geq b_i(C_N \ln N)^{i-i_0}$. Z poprzednich obliczeń, jeśli zsumujemy po wszystkich $g \geq b_{i-1}(C_N \ln N)^{i-i_0-1}$ i wy-

korzystamy własność, że rozmiar i -tego pokolenia zależy tylko od rozmiaru $(i - 1)$ -wszego pokolenia, mamy

$$\Pr \left\{ A'_i \left| \bigcap_{j=i_0}^{i-1} A'_j \right. \right\} = \Pr \left\{ A'_i \left| A'_{i-1} \right. \right\} \geq 1 - \left(\frac{1}{N^{\frac{3}{2}}} \right)^{i-i_0} \text{ for } i \geq i_0 + 1.$$

Stąd

$$\begin{aligned} \Pr \left\{ \bigcap_{i=i_0+1}^{\infty} A_i \left| A'_{i_0} \right. \right\} &\geq \Pr \left\{ \bigcap_{i=i_0+1}^{\infty} A'_i \left| A'_{i_0} \right. \right\} = \\ &= \prod_{i=i_0+1}^{\infty} \Pr \left\{ A'_i \left| \bigcap_{j=i_0}^{i-1} A'_j \right. \right\} \geq \\ &\geq \prod_{i=i_0+1}^{\infty} \left(1 - \left(\frac{1}{N^{\frac{3}{2}}} \right)^{i-i_0} \right) \geq \\ &\geq 1 - \sum_{i=i_0+1}^{\infty} \left(\frac{1}{N^{\frac{3}{2}}} \right)^{i-i_0} = \\ &= 1 - \frac{1}{N^{\frac{3}{2}}} \frac{N^{\frac{3}{2}}}{N^{\frac{3}{2}} - 1} = 1 - o\left(\frac{1}{N}\right). \end{aligned}$$

□

Lemat 8.2. Niech $i_0 \geq 1$ i g_i będzie rozmiarem i -tego pokolenia procesu gałęzowego $\mathcal{B}(P', P, N)$. W dodatku, niech

- (i) $NP = C_N \rightarrow C > 1$,
- (ii) b_i będzie zdefiniowane przez rekurencję:

$$\begin{aligned} b_{i_0} &= \frac{4C}{(\sqrt{C} - 1)^2}; \\ b_i &= b_{i-1} - \frac{\sqrt{3b_{i-1}}}{\sqrt{C_N^{i-i_0}}}, \text{ dla każdego } i \geq i_0 + 1; \end{aligned}$$

- (iii) $g_{i_0} \geq b_{i_0}$.

Wtedy dla dostatecznie dużych N

$$b_i \geq \frac{C}{2(\sqrt{C} - 1)^2} \text{ dla każdego } i \geq i_0,$$

i

$$\Pr \left\{ g_i \leq b_i C_N^{i-i_0} \ln N \mid g_{i-1} \geq b_{i-1} C_N^{i-i_0-1} \ln N \right\} \leq \frac{1}{N^{\frac{3}{2}}} \text{ dla każdego } i > i_0.$$

Dowód. Niech $b_{i_0} = \frac{4C}{(\sqrt{C}-1)^2}$ oraz, dla $i > i_0$, zdefiniujmy b_i za pomocą rekurencji $b_i = b_{i-1} - \frac{\sqrt{3b_{i-1}}}{\sqrt{C_N^{i-i_0}}}$. Mamy wtedy $b_i \leq b_{i_0}$ dla każdego $i \geq i_0$,

$\frac{1}{1-\sqrt{C_N}} \rightarrow \frac{1}{1-\sqrt{C}}$, i $4 - 2\sqrt{3} > \frac{1}{2}$, dlatego dla dostatecznie dużych N dla każdego $i \geq i_0 + 1$

$$\begin{aligned} b_i &= b_{i-1} - \frac{\sqrt{3b_{i-1}}}{\sqrt{C_N}^{i-i_0}} \geq \\ &\geq b_{i_0} - \sqrt{3b_{i_0}} \sum_{j=0}^{i-i_0-1} \frac{1}{\sqrt{C_N}^{j+1}} \geq \\ &\geq b_{i_0} - \sqrt{3b_{i_0}} \frac{1}{\sqrt{C_N} - 1} = \\ &= \frac{4C}{(\sqrt{C} - 1)^2} - \frac{2\sqrt{3}\sqrt{C}}{(\sqrt{C} - 1)(\sqrt{C_N} - 1)} \geq \frac{C}{2(\sqrt{C} - 1)^2}. \end{aligned}$$

Jeśli założymy, że dla danego $i \geq i_0 + 1$, $g_{i-1} = g \geq b_{i-1}C_N^{i-i_0-1} \ln N$, wtedy

$$\mathbb{E} = \mathbb{E}(g_i | g_{i-1} = g) \geq b_{i-1}C_N^{i-i_0} \ln N.$$

Z nierówności Chernoffa (3.1) wynika, więc że

$$\begin{aligned} \Pr_* \left\{ g_i \leq b_i C_N^{i-i_0} \ln N \right\} &\leq \Pr_* \left\{ g_i \leq \frac{b_i}{b_{i-1}} \mathbb{E} \right\} \leq \\ &\leq \exp \left(-\frac{(b_{i-1} - b_i)^2}{2b_{i-1}} C_N^{i-i_0} \ln N \right) \leq \\ &\leq \exp \left(-\frac{3}{2} \ln N \right) = \frac{1}{N^{\frac{3}{2}}}. \end{aligned}$$

Po zsumowaniu po wszystkich g takich, że $g \geq b_{i-1}C_N^{i-i_0-1} \ln N$, otrzymujemy tezę lematu. \square

Lemat 8.3. Niech $NP = C_N \ln N$, $C_N \rightarrow C$, $NP' = C'_N \ln N$, $C'_N \rightarrow C'$ i g_i będzie rozmiarem i -tego pokolenia procesu gałęzkiego $\mathcal{B}(P', P, N)$. Wtedy z prawdopodobieństwem $1 - o(\frac{1}{N})$ dla dostatecznie dużych N

$$g_i \leq b \frac{C'_N}{C_N} (C_N \ln N)^i \text{ for all } i \geq 1,$$

gdzie $b = 4 + \frac{2}{C'}$.

Dowód. Niech $b_0 = 1$ i $b_1 = \frac{\sqrt{12C'_N+1+2C'_N+1}}{2C'_N}$. Wiemy, że $g_0 = 1$ oraz że $\mathbb{E}g_1 = C'_N \ln N$, zatem z nierówności Chernoffa (3.2)

$$\begin{aligned} \Pr\{g_1 \geq b_1 C'_N \ln N\} &= \Pr\{g_1 \geq C'_N \ln N + (b_1 - 1)C'_N \ln N\} \leq \\ &\leq \exp \left(-\frac{3(b_1 - 1)^2}{2(2 + b_1)} C'_N \ln N \right) = \frac{1}{N^{\frac{3}{2}}}. \end{aligned}$$

Dla $i \geq 2$ zdefiniujemy rekurencyjnie ciąg b_i

$$b_i = b_{i-1} + \frac{2\sqrt{b_{i-1}}}{\sqrt{C'_N}} \frac{\sqrt{i-1}}{\sqrt{C_N \ln N}^{i-1}}.$$

Wówczas mamy

$$(b_i - b_{i-1})^2 = \frac{4b_{i-1}}{C'_N} \frac{i-1}{(C_N \ln N)^{i-1}}$$

oraz

$$\frac{b_i - b_{i-1}}{3b_{i-1}} = \frac{2}{3\sqrt{b_{i-1}C'_N}} \frac{\sqrt{i-1}}{\sqrt{C_N \ln N}^{i-1}} < \frac{2}{3\sqrt{C'_N C_N \ln N}} = o(1),$$

dla dostatecznie dużych N , gdzie $o(1)$ w tym przypadku oznacza jednostajną zbieżność względem wszystkich $i \geq 2$. Ostatnia nierówność wynika z tego, że $b_{i-1} > 1$ i $\frac{\sqrt{i-1}}{\sqrt{C_N \ln N}^{i-1}}$ jest malejąca dla $i \geq 2$ i $C_N \ln N > 2$.

Założmy teraz, że dla danego $i \geq 2$, $g_{i-1} = g \leq b_{i-1} \frac{C'_N}{C_N} (C_N \ln N)^{i-1}$. Wtedy

$$\mathbb{E} = \mathbb{E}(g_i | g_{i-1} = g) \leq b_{i-1} \frac{C'_N}{C_N} (C_N \ln N)^i.$$

Z powyższego i nierówności Chernoffa (3.2) wynika, że

$$\begin{aligned} \Pr_* \left\{ g_i \geq b_i \frac{C'_N}{C_N} (C_N \ln N)^i \right\} &= \\ &= \Pr_* \left\{ g_i \geq b_{i-1} \frac{C'_N}{C_N} (C_N \ln N)^i + (b_i - b_{i-1}) \frac{C'_N}{C_N} (C_N \ln N)^i \right\} \leq \\ &\leq \Pr_* \left\{ g_i \geq \mathbb{E} + (b_i - b_{i-1}) \frac{C'_N}{C_N} (C_N \ln N)^i \right\} \leq \\ &\leq \exp \left(- \left(1 - \frac{(b_i - b_{i-1}) \frac{C'_N}{C_N} (C_N \ln N)^i}{3\mathbb{E}} \right) \frac{(b_i - b_{i-1})^2 \frac{(C'_N)^2}{C_N^2} (C_N \ln N)^{2i}}{2\mathbb{E}} \right) \leq \\ &\leq \exp \left(- \left(1 - \frac{b_i - b_{i-1}}{3b_{i-1}} \right) \frac{(b_i - b_{i-1})^2 \frac{C'_N}{C_N} (C_N \ln N)^i}{2b_{i-1}} \right) \leq \\ &\leq \exp \left(- \left(1 - \frac{2}{3\sqrt{C_N^2 \ln N}} \right) 2(i-1) \ln N \right) \leq \left(\frac{1}{N^{\frac{3}{2}}} \right)^{i-1}, \end{aligned}$$

dla dostatecznie dużych N , jednostajnie względem wszystkich i .

Udowodnimy teraz indukcyjnie, że, dla dostatecznie dużych N , $b_i \leq b$ dla każdego i , gdzie $b = 4 + \frac{2}{C'}$. Dla dostatecznie dużych N zachodzi $b_0 < b$ i $b_1 < 4 + \frac{1}{C'_N} < b$ (to wynika z nierówności $(\sqrt{12C'_N + 1} - 1)^2 > 0$ i $C'_N \rightarrow C'$). Założmy teraz, że $b_j < b$ dla każdego $1 \leq j < i$. Skoro $b_1 < 4 + \frac{1}{C'_N}$, $C'_N \rightarrow C'$

i $\frac{2\sqrt{b}}{\sqrt{C'_N C_N \ln N}} \frac{C_N \ln N}{(\sqrt{C_N \ln N} - 1)^2} = o(1)$ otrzymujemy

$$\begin{aligned}
b_i &= b_{i-1} + \frac{2\sqrt{b_{i-1}}}{\sqrt{C'_N}} \frac{\sqrt{i-1}}{\sqrt{C_N \ln N^{i-1}}} = b_1 + \sum_{j=1}^{i-1} \frac{2\sqrt{b_j}}{\sqrt{C'_N}} \frac{\sqrt{j}}{\sqrt{C_N \ln N^j}} \leq \\
&\leq b_1 + \frac{2\sqrt{b}}{\sqrt{C'_N C_N \ln N}} \sum_{j=1}^{i-1} \frac{j}{\sqrt{C_N \ln N^{j-1}}} \leq \\
&\leq 4 + \frac{1}{C'_N} + \frac{2\sqrt{b}}{\sqrt{C'_N C_N \ln N}} \frac{C_N \ln N}{(\sqrt{C_N \ln N} - 1)^2} < \\
&< 4 + \frac{2}{C'}
\end{aligned}$$

dla dostatecznie dużych N .

Jak w poprzednich dowodach, definiujemy dla $i \geq 0$, A_i jako zdarzenie, że $g_i \leq b(C_N \ln N)^i$ oraz A'_i jako zdarzenie, że $g_i \leq b_i(C_N \ln N)^i$. Z poprzednich wyników, jeśli zsumujemy po wszystkich $g \leq b_{i-1}(C_N \ln N)^{i-1}$ i wykorzystamy fakt, że rozmiar i -tego pokolenia w procesie gałęzowym zależy tylko od rozmiaru $(i-1)$ -wszego pokolenia, to dla $i \geq 2$ mamy

$$\Pr \left\{ A'_i \mid \bigcap_{j=1}^{i-1} A'_j \right\} = \Pr \left\{ A'_i \mid A'_{i-1} \right\} \geq 1 - \left(\frac{1}{N^{\frac{3}{2}}} \right)^{i-1}$$

i

$$\Pr \{ A'_1 \} \geq 1 - \frac{1}{N^{\frac{3}{2}}}.$$

Zatem

$$\begin{aligned}
\Pr \left\{ \bigcap_{i=1}^{\infty} A_i \right\} &\geq \Pr \left\{ \bigcap_{i=1}^{\infty} A'_i \right\} = \left(\prod_{i=2}^{\infty} \Pr \left\{ A'_i \mid \bigcap_{j=1}^{i-1} A'_j \right\} \right) \Pr \{ A'_1 \} \geq \\
&\geq \left(\prod_{i=2}^{\infty} \left(1 - \left(\frac{1}{N^{\frac{3}{2}}} \right)^{i-1} \right) \right) \left(1 - \frac{1}{N^{\frac{3}{2}}} \right) \geq \\
&\geq 1 - \frac{1}{N^{\frac{3}{2}}} - \sum_{i=2}^{\infty} \left(\frac{1}{N^{\frac{3}{2}}} \right)^{i-1} = \\
&= 1 - \frac{1}{N^{\frac{3}{2}}} - \frac{1}{N^{\frac{3}{2}}} \frac{N^{\frac{3}{2}}}{N^{\frac{3}{2}} - 1} = 1 - o\left(\frac{1}{N}\right).
\end{aligned}$$

□

Lemat 8.4. Niech $NP = C_N$, $C_N \rightarrow C > 1$, $NP' = C'_N$, $C'_N \rightarrow C' > 1$ i g_i będzie rozmiarem i -tego pokolenia procesu gałęzowego $\mathcal{B}(P', P, N)$. Wtedy z prawdopodobieństwem $1 - o(\frac{1}{N})$ dla dostatecznie dużych N

$$g_i \leq i^2 C'_N C_N^{i-1} \ln N \text{ dla każdego } i \geq 1.$$

Dowód. Niech $g_0 = 1$ i $b_i = i^2$ dla każdego $i \geq 1$. Wtedy $\mathbb{E}g_1 = C'_N$ i z nierówności Chernoffa (3.2)

$$\begin{aligned} \Pr(g_1 \geq C'_N \ln N) &= \Pr(g_1 \geq C'_N + C'_N(\ln N - 1)) \leq \\ &\leq \exp\left(-\frac{3}{2} \frac{(\ln N - 1)^2}{(\ln N + 2)} C'_N\right) = \\ &= \exp\left(-\frac{3}{2} \frac{(\ln N - 1)^2}{\ln N (\ln N + 2)} C'_N \ln N\right) = o\left(\frac{1}{N}\right). \end{aligned}$$

Założmy teraz, że dla $i \geq 2$ zachodzi $g_{i-1} = g \leq b_{i-1} C'_N C_N^{i-2} \ln N$, wtedy $E = \mathbb{E}(g_i | g_{i-1} = g) \leq b_{i-1} C'_N C_N^{i-1} \ln N$. Z nierówności Chernoffa (3.2) i faktu, że dla $i \geq 2$

$$\frac{3(b_i - b_{i-1})^2}{2(b_i + 2b_{i-1})} = \frac{3(4i^2 - 4i + 1)}{2(3i^2 - 4i + 2)} > \frac{3}{2},$$

otrzymujemy, że

$$\begin{aligned} \Pr_* \{g_i \geq b_i C'_N C_N^{i-1} \ln N\} &= \\ &= \Pr_* \{g_i \geq b_{i-1} C'_N C_N^{i-1} \ln N + (b_i - b_{i-1}) C'_N C_N^{i-1} \ln N\} \leq \\ &\leq \Pr_* \{g_i \geq E + (b_i - b_{i-1}) C'_N C_N^{i-1} \ln N\} \leq \\ &\leq \exp\left(-\frac{3}{2} \frac{(b_i - b_{i-1})^2 (C'_N)^2 C_N^{2i-2} \ln^2 N}{(3E + (b_i - b_{i-1}) C'_N C_N^{i-1} \ln N)}\right) \leq \\ &\leq \exp\left(-\frac{3}{2} \frac{(b_i - b_{i-1})^2}{(b_i + 2b_{i-1})} C'_N C_N^{i-1} \ln N\right) \leq \left(\frac{1}{N^{\frac{3}{2}}}\right)^{C'_N C_N^{i-1}}. \end{aligned}$$

Zdefiniujmy dla $i \geq 1$, A_i jako zdarzenie, że $g_i \leq b_i C'_N C_N^{i-1} \ln N$. Z poprzednich wyników, jeśli zsumujemy po wszystkich $g \leq b_{i-1} C'_N C_N^{i-2} \ln N$ korzystając znowu z faktu, że rozmiar i -tego pokolenia zależy tylko od rozmiaru poprzedniego pokolenia, to dla $i \geq 2$ mamy

$$\Pr \left\{ A_i \mid \bigcap_{j=1}^{i-1} A_j \right\} = \Pr \{A_i | A_{i-1}\} \geq 1 - \left(\frac{1}{N^{\frac{3}{2}}}\right)^{C'_N C_N^{i-1}}$$

i

$$\Pr \{A_1\} = 1 - o\left(\frac{1}{N}\right).$$

Zatem, podobnie jak w lemacie 8.3

$$\begin{aligned} \Pr \left\{ \bigcap_{i=1}^{\infty} A_i \right\} &\geq 1 - o\left(\frac{1}{N}\right) - \sum_{i=2}^{\infty} \left(\frac{1}{N^{\frac{3}{2}}}\right)^{C'_N C_N^{i-1}} = \\ &= 1 - o\left(\frac{1}{N}\right) - \frac{1}{N^{\frac{3}{2}}} \sum_{i=2}^{\infty} \left(\frac{1}{N^{\frac{3}{2}}}\right)^{C'_N C_N^{i-1}-1} = 1 - o\left(\frac{1}{N}\right), \end{aligned}$$

ponieważ $C'_N, C_N > 1$ i dla dostatecznie dużych N szereg $\sum_{i=2}^{\infty} \left(\frac{1}{N^{\frac{3}{2}}}\right)^{C'_N C_N^{i-1}-1}$ jest zbieżny. \square

Lemat 8.5. Niech $NP = C_N \rightarrow C > 1$, b_N będzie takie, że $b_N/\ln N \rightarrow \infty$ i g_k będzie rozmiarem k -tego pokolenia w procesie gałęzkowym $\mathcal{B}(P', P, N)$. wtedy z prawdopodobieństwem $1 - o(\frac{1}{N})$ dla dostatecznie dużych N , jeśli dla danego $k \geq 1$

$$g_{k-1} \leq b_N,$$

to

$$g_k \leq 2b_N.$$

Dowód. Jeśli $g_{k-1} = g \leq b_N$, to $E = \mathbb{E}(g_k | g_{k-1} = g) \leq C_N b_N$. zatem z nierówności Chernoffa (3.2)

$$\begin{aligned} \Pr_* \{g_k \geq 2C_N b_N\} &\leq \Pr_* \{g_k \geq E + C_N b_N\} \leq \\ &\leq \exp\left(-\frac{3}{8}C_N b_N\right) = o\left(\frac{1}{N}\right). \end{aligned}$$

Po zsumowaniu po wszystkich $g \leq b_N$ otrzymujemy natychmiast tezę lematu. \square

8.4. Przybliżenie BFS procesami gałęzkowymi

W naszych rozważaniach skupiamy się na przypadku gdy $d^2 n/m = O(\ln n)$ (tzn. $d^2 n_{max} = O(m n_{max} \ln n/n) = O(m/(\ln n)^{\frac{1}{3}}) = o(m)$). Zatem z (8.6), (8.7), wynika, że dopóki drzewo BFS ma nie więcej niż n_{max} wierzchołków, procedura BFS lub jej niektóre kroki mogą być przybliżone dwoma procesami gałęzkowymi $\mathcal{B}_+(P'_+, P_+, N_+)$ i $\mathcal{B}_-(P'_-, P_-, N_-)$ z rozkładami liczby potomków $\text{Bin}(N_+, P'_+)$, $\text{Bin}(N_+, P_+)$ i $\text{Bin}(N_-, P'_-)$, $\text{Bin}(N_-, P_-)$, odpowiednio, gdzie

$$(8.10) \quad \begin{aligned} N_+ &= n, \\ P'_+ &= \frac{d^2}{m} \left(1 + O\left(\frac{dn_{max}}{m}\right)\right), \quad P_+ = \frac{d(d-1)}{m} \left(1 + O\left(\frac{dn_{max}}{m}\right)\right), \end{aligned}$$

oraz

$$(8.11) \quad \begin{aligned} N_- &= n - n_{max}, \\ P'_- &= \frac{d^2}{m} \left(1 - O\left(\frac{d^2 n_{max}}{m}\right)\right), \quad P_- = \frac{d(d-1)}{m} \left(1 - O\left(\frac{d^2 n_{max}}{m}\right)\right). \end{aligned}$$

Inaczej mówiąc liczbę wierzchołków dodanych w kolejnych krokach BFS ograniczyć można z dołu przez liczbę potomków procesu \mathcal{B}_- oraz od góry przez liczbę potomków procesu \mathcal{B}_+ .

Wprowadźmy nowe oznaczenia. Niech g_i^+ i g_i^- będą wielkościami i -tego pokolenia, odpowiednio, procesów \mathcal{B}_+ i \mathcal{B}_- . Przypomnijmy, że $d(u, v) = k$ oznacza, że długość najkrótszej ścieżki z $u \in \mathcal{V}$ do $v \in \mathcal{V}$ jest równa k . Oznaczmy dla $v \in \mathcal{V}$ i $k \geq 0$

$$\begin{aligned} \Gamma_0(v) &= \{v\}, \\ \Gamma_k(v) &:= \{u \in \mathcal{V} : d(v, u) = k\}, \quad \text{dla } k \geq 1, \end{aligned}$$

oraz

$$\mathcal{N}_k(v) := \bigcup_{i=0}^k \Gamma_i(v).$$

Z rozważań w podrozdziale 8.2, jak długo $|\mathcal{N}_k(v)| \leq n_{max}$, tak długo możemy ograniczyć z dołu i z góry proces BFS w taki sposób, że

$$(8.12) \quad \begin{aligned} g_k^- &\leq |\Gamma_k(v)| \leq g_k^+; \\ \sum_{i=0}^k g_i^- &\leq |\mathcal{N}_k(v)| \leq \sum_{i=0}^k g_i^+; \end{aligned}$$

z prawdopodobieństwem równym jeden.

Przypomnijmy, że ograniczenie z dołu przez proces gałązkowy \mathcal{B}_- wynika z oszacowania na prawdopodobieństwo dodania wierzchołka do podgrafu drzewa BFS zawierającego tylko kopie wierzchołków dodających $d - 1$ elementów do $D(V_{BFS})$. Ponieważ będziemy też zainteresowani przybliżeniem tylko niektórych kroków procesu BFS krokami procesu \mathcal{B}_- , przydatny będzie poniższy lemat oszacowujący liczbę wierzchołków w podgrafie drzewa BFS składającym się z kopii wierzchołków, które dodają $d - 1$ elementów do $D(V_{BFS})$.

Lemat 8.6. *Niech $b = o\left(\frac{n_{max}}{\ln n}\right)$, $\frac{d^2 n}{m} = O(\ln n)$ i $i_0 = O(\ln n)$ będzie najmniejszym indeksem takim, że*

$$|\Gamma_{i_0}(v)| \geq b,$$

wtedy z prawdopodobieństwem $1 - o(1/n)$ dla dużych n

$$|\{v' \in \Gamma_{i_0}(v) : |D^*(v')| = d - 1\}| \geq b',$$

gdzie

$$b' = \begin{cases} b - 1 & \text{dla } b = o\left(\frac{\sqrt[4]{n}}{\sqrt{\ln n}}\right); \\ \frac{2}{3}b & \text{dla } \frac{b}{\ln n} \rightarrow \infty. \end{cases}$$

Dowód. Niech $|\Gamma_i(v)| < b$ dla każdego $i < i_0$. Załóżmy, że zaczęliśmy procedurę BFS w wierzchołku v oraz t i t' są takie, że $v_{t+1} \in \Gamma_{i_0-1}(v)$ oraz po t krokach i t' podkrokach procedury BFS dodaliśmy do drzewa BFS dokładnie j , $j < b$, wierzchołków z $\Gamma_{i_0}(v)$. Niech v' będzie wierzchołkiem z $\mathcal{V} \setminus V_{BFS}$, który będzie rozpatrywany w kolejnym podkroku. Zdefiniujmy A , B , $D^*(v_{t+1})$ i $D^*(v')$ jak w (8.1), (8.3) i (8.4). Wtedy z powyższych założeń wynika, że

$$\begin{aligned} |B| &< d i_0 b; \\ |A| &< 2d b. \end{aligned}$$

Prawdopodobieństwo, że kopia wierzchołka v' została dodana do drzewa BFS w rozpatrywanym podkroku wynosi nie więcej niż

$$\frac{\binom{|D^*(v_{t+1})|}{1} \binom{m-|B|}{d-1}}{\binom{m-|B|}{d}}.$$

Prawdopodobieństwo, że wierzchołek v' został dodany do drzewa BFS w rozpatrywanym podkroku oraz $|D^*(v')| = d - 1$ wynosi

$$\frac{\binom{|D^*(v_{t+1})|}{1} \binom{m-|B|-|A|}{d-1}}{\binom{m-|B|}{d}}.$$

Dlatego, prawdopodobieństwo warunkowe zajścia zdarzenia $|D^*(v')| = d - 1$ pod warunkiem, że v' został dodany do drzewa BFS w tym podkroku wynosi co najmniej

$$\begin{aligned} \frac{\binom{m-|B|-|A|}{d-1}}{\binom{m-|B|}{d-1}} &= \prod_{l=0}^{d-2} \left(1 - \frac{|A|}{m - |B| - l}\right) > \\ &> 1 - \frac{d|A|}{m - |B| - d} > 1 - \frac{2d^2b}{m - di_0b - d} = 1 - \xi, \end{aligned}$$

gdzie

$$\xi = \frac{2d^2b}{m - di_0b - d} = O\left(\frac{d^2b}{m}\right) = O\left(\frac{b \ln n}{n}\right).$$

Założmy teraz, że $|\Gamma_{i_0}(v)| \geq b$. Z powyższych rozważań zmienna losowa X o rozkładzie dwumianowym $\text{Bin}(b, 1 - \xi)$ jest stochastycznie zdominowana przez zmienną losową określającą moc zbioru

$$\{v' \in \Gamma_{i_0}(v) : |D^*(v')| = d - 1\}.$$

Dlatego wystarczy udowodnić, że z prawdopodobieństwem $1 - o(1/n)$ zachodzi zdarzenie $X \geq b'$.

Niech $b = o\left(\frac{\sqrt[4]{n}}{\sqrt{\ln n}}\right)$, wtedy, gdy podstawimy $b' = b - 1$, mamy

$$\begin{aligned} \Pr\{X \geq b'\} &= \binom{b}{b-1} \xi(1-\xi)^{b-1} + (1-\xi)^b \geq \\ &\geq b\xi(1-(b-1)\xi) + 1 - b\xi = 1 - b(b-1)\xi^2 = 1 - O\left(\frac{b^4 \ln^2 n}{n^2}\right) = 1 - o\left(\frac{1}{n}\right). \end{aligned}$$

Dla przypadku $\frac{b}{\ln n} \rightarrow \infty$, teza zachodzi po standardowym zastosowaniu nierówności Chernoffa (3.1). \square

Aproksymacja procesami gałęzkowymi okazuje się przydatna w przypadku, gdy dla pewnego k moc zbioru $\Gamma_k(v)$ jest dostatecznie duża. W dalszej części pokażemy, że dla każdego wierzchołka z prawdopodobieństwem $1 - o(1/n)$ proces gałęzkowy rozpoczęty w tym wierzchołku wygasa szybko lub dla pewnego k moc zbioru $\Gamma_k(v)$ jest dostatecznie duża, aby kolejne kroki procesu można było przybliżać kolejnymi krokami procesu gałęzkowego. W tym celu korzystać będziemy z poniższego rozumowania.

Uwaga 8.1. Niech $P \in (0; 1]$, λ będzie stałą dodatnią, T będzie nieskończonym drzewem opisanym w podrozdziale 8.2. Niech $n = |\mathcal{V}|$ oraz niech

N_- będzie takie, że $N_- \leq n - \lambda$. Rozważmy teraz procedurę BFS w grafie $G^{Jed}(\mathcal{V}, m, d)$ o początku w wierzchołku v . Rozważmy dowolny podkrok, w którym rozważamy (dodajemy lub nie) wierzchołek $w \in \mathcal{V}$, którego kopia nie została jeszcze dodana do drzewa BFS. Załóżmy, że jak długo drzewo BFS ma mniej wierzchołków niż λ tak długo, w każdym podkroku liczba wierzchołków dodana do drzewa BFS w tym podkroku może zostać ograniczona z dołu przez zmienną losową o rozkładzie Bernoulliego z parametrem P niezależną od dotychczas zbudowanego drzewa BFS (będzie to naturalne założenie dla $G^{Jed}(\mathcal{V}, m, d)$ wynikające z niezależności wyborów zbiorów $D(v)$ oraz z (8.7)). W dodatku, jeśli przed rozważanym krokiem drzewo miało mniej niż λ wierzchołków, wtedy co najwyżej λ wierzchołków zostało dodane w tym kroku do drzewa BFS z prawdopodobieństwem 0 (tzn. kopie co najmniej $n - \lambda$ wierzchołków nie zostały jeszcze dodane do drzewa BFS). Zatem liczbę wierzchołków dodanych w podkroku ograniczymy przez zmienną losową o rozkładzie Bernoulliego z parametrem P (w co najmniej N_- podkrokach) lub zmienną losową, która jest równa zeru z prawdopodobieństwem równym jedynce. Niech X będzie zmienną losową o rozkładzie dwumianowym $\text{Bin}(kN_-, P)$. Z faktu, że suma niezależnych zmiennych losowych o rozkładzie Bernoulliego z parametrem P jest zmienną losową o rozkładzie dwumianowym, po zastosowaniu rozumowania podobnego do tego z dowodu twierdzenia 5.4 z [32], wynika, że

$$\Pr\{X \geq b + k - 1 = \lambda\} \leq \Pr\{\mathcal{A}\} \leq \Pr\{\mathcal{A}'\},$$

gdzie \mathcal{A} jest zdarzeniem, że procedura BFS kończy się przed k -tym krokiem lub λ -ty wierzchołek został dodany do drzewa BFS przed $k+1$ -wszym krokiem oraz \mathcal{A}' jest zdarzeniem, że albo procedura BFS kończy się w co najwyżej $k-1$ krokach lub istnieje $k_0 \leq k$ takie, że po k_0 -tym kroku istnieje b nierozpatrzonych wierzchołków.

Teraz wykorzystamy powyższe rozważania, aby przyjrzeć się bliżej pierwszym krokom procedury BFS. Poniższe lematy będą mówić o tym, że procedura BFS w $G^{Jed}(\mathcal{V}, m, d)$ albo kończy się szybko albo drzewo BFS szybko się rozrasta.

Lemat 8.7. Niech $v \in \mathcal{V}$, $b > 0$ będzie dowolną ustaloną stałą oraz n , m , d będą zdefiniowane jak w twierdzeniu 8.1(i). Jeśli dla P_- i N_- zdefiniowanych jak w (8.11) zachodzi $P_- N_- / \ln N_- \rightarrow C$, wtedy z prawdopodobieństwem $1 - o(1/n)$ procedura BFS $G^{Jed}(\mathcal{V}, m, d)$ zaczynająca od wierzchołka v albo skończy się w co najwyżej $\lfloor \frac{1}{C} \rfloor$ krokach albo istnieje $1 \leq i_0 \leq \lfloor \frac{1}{C} \rfloor + 1$, takie, że

$$|\Gamma_{i_0}(v)| \geq b.$$

Dowód. Niech $N_- P_- / \ln N_- = C_N$, $b = O(1)$, $\lambda = \lambda = 2b + \lfloor \frac{1}{C} \rfloor$, $k = \lfloor \frac{1}{C} \rfloor + 1$ i $X \sim \text{Bin}(kN_-, P_-)$. W dodatku, skoro $\frac{1}{C} - \lfloor \frac{1}{C} \rfloor < 1$, możemy ustalić stałą $0 < \varepsilon < \frac{1}{3}$ taką, że $\lfloor \frac{1}{C} \rfloor + 1 \geq \frac{1+3\varepsilon}{C}$. W dodatku możemy ustalić dostatecznie duże n , aby zachodziło $\frac{C_N}{C} \geq 1 - \varepsilon$. Dlatego dla dostatecznie dużych n

$$\left(\left\lfloor \frac{1}{C} \right\rfloor + 1\right) C_N \geq \frac{C_N}{C} (1 + 3\varepsilon) \geq (1 - \varepsilon)(1 + 3\varepsilon) = 1 + 2\varepsilon - 3\varepsilon^2 > 1 + \varepsilon.$$

Dlatego, dla dostatecznie dużych n ,

$$\begin{aligned}
\Pr \{X \leq \lambda\} &= \sum_{i=0}^{\lambda} \binom{kN_-}{i} P_-^i (1 - P_-)^{kN_- - i} = \\
&= (1 - P_-)^{kN_-} \left(1 + \sum_{i=1}^{\lambda} \binom{kN_-}{i} \left(\frac{P_-}{1 - P_-} \right)^i \right) \leq \\
&\leq \exp(-kP_-N_-) \left(1 + \sum_{i=1}^{\lambda} \left(\frac{kN_-P_-}{1 - P_-} \right)^i \right) = \\
&= \exp \left(- \left(\left\lfloor \frac{1}{C} \right\rfloor + 1 \right) C_N \ln N_- \right) \frac{\left(\frac{kN_-P_-}{1 - P_-} \right)^{\lambda+1} - 1}{\frac{kN_-P_-}{1 - P_-} - 1} \leq \\
&\leq \frac{1}{N_-^{1+\varepsilon}} O((\ln N_-)^\lambda) = o\left(\frac{1}{n}\right).
\end{aligned}$$

Dlatego z uwagi 8.1 z prawdopodobieństwem $1 - o(1/n)$ proces BFS kończy się w co najwyżej $\lfloor \frac{1}{C} \rfloor$ krokach lub istnieje $k_0 \leq \lfloor \frac{1}{C} \rfloor + 1$ takie, że po k_0 -tym kroku mamy co najmniej $2b$ nierozpatrzonych wierzchołków. Załóżmy, że wierzchołek rozpatrywany w k_0 -tym kroku należy do zbioru $\Gamma_j(v)$. Oczywiście $j \leq \lfloor \frac{1}{C} \rfloor$ oraz wszystkie nierozpatrzone po tym kroku wierzchołki należą do $\Gamma_j(v)$ lub $\Gamma_{j+1}(v)$. Czyli

$$|\Gamma_j(v)| + |\Gamma_{j+1}(v)| \geq 2b.$$

Dlatego istnieje $1 \leq i_0 \leq \lfloor \frac{1}{C} \rfloor + 1$ ($i_0 = j$ lub $i_0 = j + 1$) takie, że

$$|\Gamma_{i_0}(v)| \geq b.$$

□

Lemat 8.8. Niech $v \in \mathcal{V}$ oraz n, m, d będą zdefiniowane jak w twierdzeniu 8.1(ii). Jeżeli dla P_- i N_- zdefiniowanych jak w (8.11) zachodzi $P_-N_- \rightarrow C > 1$, to z prawdopodobieństwem $1 - o(1/n)$ procedura BFS w $G^{Jed}(\mathcal{V}, m, d)$ zaczynająca się w wierzchołku v kończy się w co najwyżej $\frac{16C}{(\sqrt{C}-1)^2(C-1)} \ln N - 1$ krokach lub istnieje $1 \leq i_0 \leq \frac{16C}{(\sqrt{C}-1)^2(C-1)} \ln N$, takie, że

$$|\Gamma_{i_0}(v)| \geq \frac{4C}{(\sqrt{C}-1)^2} \ln N + 1.$$

Dowód. Niech $P_-N_- = C_N$, $k = \frac{16C}{(\sqrt{C}-1)^2(C-1)} \ln N$ i $b = \frac{4C}{(\sqrt{C}-1)^2} \ln N = \frac{(C-1)k}{4}$. W dodatku niech X będzie zdefiniowane jak w dowodzie lematu 8.7. Dowód jest analogiczny do dowodu lematu 8.7. Jedyną rzeczą jaką musimy udowodnić to to, że z prawdopodobieństwem $o(1/n)$

$$(8.13) \quad X \leq k - 1 + 2b + 1 = k - 1 + \frac{(C-1)k}{2} + 1.$$

Wówczas z prawdopodobieństwem $1 - o(1/n)$ procedura BFS kończy się szybko albo istnieje odpowiednie j takie, że

$$|\Gamma_j(v)| + |\Gamma_{j+1}(v)| \geq 2b + 1.$$

Udowodnijmy, więc (8.13). Zauważmy, że

$$\mathbb{E}X = kN_-P_- = C_N k.$$

Z nierówności Chernoffa (3.1), dla dostatecznie dużych n

$$\begin{aligned} \Pr \left\{ X \leq k + \frac{(C-1)k}{2} \right\} &\leq \exp \left(-\frac{(2C_N - C - 1)^2 k}{8C_N} \right) \leq \\ &\leq \exp \left(-\frac{9}{8} \frac{(2C_N - C - 1)^2 C}{(C-1)^2 C_N} \ln N \right) = o \left(\frac{1}{n} \right). \end{aligned}$$

Nierówności wynikają z tego, że

$$k = \frac{16C}{(\sqrt{C}-1)^2(C-1)} \ln N > \frac{9C}{(C-1)^2} \ln N$$

oraz

$$\frac{(2C_N - C - 1)^2 C}{(C-1)^2 C_N} \rightarrow 1.$$

□

8.5. Dowód twierdzenia o długości średnicy

Aby udowodnić główne twierdzenie tego rozdziału w poprzednich podrozdziałach zebraliśmy wyniki pomocnicze, które teraz wykorzystamy.

8.5.1. Własności $G^{Jed}(\mathcal{V}, m, d)$ na progu spójności

Z poniższego lematu wynika oszacowanie dolne na średnicę składowej o $\Omega(n)$ wierzchołkach. Pokażemy, że istnieje $k_- \sim \ln n / \ln \ln n$ takie, że asymptotycznie prawie na pewno dla dowolnego wierzchołka v istnieje zbiór o mocy $n - o(n_{max})$, składający się z wierzchołków w odległości większej niż k_- od v .

Lemat 8.9. *Niech $d^2 n / (m \ln n) = c_n \rightarrow c$. Wtedy istnieje $n_1 = o(n_{max})$ i $k_- \sim \ln n / \ln \ln n$, takie, że w $G^{Jed}(\mathcal{V}, m, d)$*

$$\Pr \left\{ \forall v \in \mathcal{V} |\mathcal{N}_{k_-}(v)| \leq n_1 \right\} = 1 - o(1).$$

Dowód. Zdefiniujmy P_+ jak w podrozdziale 8.2. Z (8.10) wynika, że

$$N_+ P'_+ = C'_{N_+} \ln N_+, \quad \text{i} \quad N_+ P_+ = C_{N_+} \ln N_+,$$

gdzie

$$N_+ = n, \quad C'_{N_+} \rightarrow C' = c \quad \text{i} \quad C_{N_+} \rightarrow C = \begin{cases} \frac{d-1}{d} c & \text{dla } d = O(1), \\ c & \text{dla } d \rightarrow \infty. \end{cases}$$

Niech

$$k_- = \left\lfloor \frac{\ln N_+ - \frac{3}{2} \ln \ln N_+ - \ln \left(4 + \frac{2}{C} \right) - \ln \frac{C'_{N_+}}{C_{N_+}}}{\ln C_{N_+} + \ln \ln N_+} \right\rfloor.$$

Oczywiście $k_- \sim \ln n / \ln \ln n$. Z lematu 8.3 i (8.12) dla danego $v \in \mathcal{V}$ z prawdopodobieństwem $1 - o(1/n)$ mamy

$$(8.14) \quad |\mathcal{N}_{k_-}(v)| \leq 1 + \sum_{i=1}^{k_-} \frac{C'_{N_+}}{C_{N_+}} \left(4 + \frac{2}{C}\right) (C_{N_+} \ln N_+)^i = o(n_{max}).$$

Zatem dla

$$n_1 = 1 + \sum_{i=1}^{k_-} \frac{C'_{N_+}}{C_{N_+}} \left(4 + \frac{2}{C}\right) (C_{N_+} \ln N_+)^i$$

teza jest prawdziwa. □

Poniższy lemat daje oszacowanie górne na średnicę składowej.

Lemat 8.10. *Niech $d^2 n / (m \ln n) = c_n \rightarrow c$. Wtedy asymptotycznie prawie na pewno $G^{Jed}(\mathcal{V}, m, d)$ składa się ze składowych o rozmiarze co najwyżej $\lfloor \frac{1}{C} \rfloor$, gdzie*

$$C = \begin{cases} \frac{d-1}{d}c & \text{for } d = O(1), \\ c & \text{for } d \rightarrow \infty \end{cases}$$

oraz z co najwyżej jednej składowej G o rozmiarze $\Omega(n / \ln n)$. W dodatku istnieje $k_+ \sim \ln n / \ln \ln n$ takie, że asymptotycznie prawie na pewno

$$\text{diam} G \leq k_+.$$

Dowód. Niech $c > 0$ będzie dowolną stałą i $d^2 n / (m \ln n) = c_n \rightarrow c$. Zdefiniujmy P_- jak w podrozdziale 8.2. Z (8.11)

$$N_- P'_- = C'_{N_-} \ln N_-, \quad \text{i} \quad N_- P_- = C_{N_-} \ln N_-,$$

gdzie

$$N_- = n - n_{max}, \quad C'_{N_-} \rightarrow C' = c \quad \text{i} \quad C_{N_-} \rightarrow C = \begin{cases} \frac{d-1}{d}c & \text{dla } d = O(1), \\ c & \text{dla } d \rightarrow \infty. \end{cases}$$

Niech w dodatku

$$(8.15) \quad b = \begin{cases} \frac{3}{c^2} \left(\frac{d}{d-1}\right)^3 & \text{dla } d = O(1); \\ \frac{3}{c^2} & \text{dla } d \rightarrow \infty \end{cases}$$

i b' będzie takie, że

$$b = b' - \sqrt{\frac{4b'}{C}}.$$

Niech $v \in \mathcal{V}$. Z lematu 8.7 z prawdopodobieństwem $1 - o(1/n)$ v jest zawarte w składowej o co najwyżej $\lfloor \frac{1}{C} \rfloor$ wierzchołkach lub

$$(8.16) \quad \exists_{i_0 \leq \lfloor \frac{1}{C} \rfloor + 1} |\Gamma_{i_0}(v)| \geq b' + 1.$$

Niech \mathcal{V}^* będzie zbiorem tych wierzchołków z \mathcal{V} , dla których zachodzi (8.16). Niech także $i_0 = i_0(v)$ i $b_{i_0} = b'$ dla każdego $v \in \mathcal{V}^*$. Z lematu 8.6 wiemy,

że jeśli $|\Gamma_{i_0}(v)| \geq b' + 1$ (tzn. jeśli $v \in \mathcal{V}^*$), wtedy z prawdopodobieństwem $1 - o(1/n)$ dla co najmniej b_{i_0} wierzchołków z $\Gamma_{i_0}(v)$ zachodzi $|D^*(\cdot)| = d - 1$.
Niech

$$k' = \left\lceil \frac{1 \ln N_- - \ln \ln N_- + \ln b - 2 \ln b}{2 \ln \ln N_- + \ln C_{N_-}} \right\rceil + \left\lfloor \frac{1}{C} \right\rfloor + 1.$$

Zauważmy, że $k' \sim \ln n / (2 \ln \ln n)$, stąd $k' \leq k_-$ dla dużych n . Z powyższych rozważań, lematu 8.9, (9.11) oraz z faktu, że $k' \leq k_-$, dla dużych n , z prawdopodobieństwem $1 - o(1/n)$, kolejne kroki procesu BFS mogą być przybliżone procesem \mathcal{B}_- z $g_{i_0}^- \geq b_{i_0}$. Z lematu 8.1, dla danego $v \in \mathcal{V}^*$ z prawdopodobieństwem $1 - o(1/n)$

$$|\Gamma_{k'}(v)| \geq g_{k'}^- \geq b (C_{N_-} \ln N_-)^{k' - i_0} \geq \lambda,$$

gdzie

$$\lambda = \lambda(n) = \sqrt{b \frac{N_-}{\ln N_-}}.$$

Stąd

$$\Pr \{ \forall v \in \mathcal{V}^* |\Gamma_{k'}(v)| \geq \lambda \} = 1 - o(1).$$

W dodatku wiemy, że wynik został uzyskany z przybliżenia procesem gałęzkowym podgrafu drzewa BFS, w którym to podgrafie każdy wierzchołek dodaje $d - 1$ nowych elementów do $D(V_{BFS})$. Dlatego

$$\Pr \{ \forall v \in \mathcal{V}^* |D(\Gamma_{k'}(v)) \setminus D(\mathcal{N}_{k'-1}(v))| \geq (d - 1)\lambda \} = 1 - o(1).$$

Dla dowolnych dwóch wierzchołków $v, v' \in \mathcal{V}^*$ zdefiniujmy

$$\begin{aligned} B &= B(v, v') = D(\mathcal{N}_{k'-1}(v) \cup \mathcal{N}_{k'-1}(v')); \\ A &= A(v) = D(\Gamma_{k'}(v)) \setminus D(\mathcal{N}_{k'-1}(v)); \\ A' &= A'(v') = D(\Gamma_{k'}(v')) \setminus D(\mathcal{N}_{k'-1}(v')). \end{aligned}$$

Z powyższych rozważań asymptotycznie prawie na pewno dla dowolnych $v, v' \in \mathcal{V}^*$

$$|B| + |A| + |A'| \leq 2dn_{max}, \quad |A| \geq (d - 1)\lambda, \quad |A'| \geq (d - 1)\lambda.$$

Z lematu 8.9

$$|\mathcal{V} \setminus (\mathcal{N}_{k'}(v) \cup \mathcal{N}_{k'}(v'))| \geq n - 2n_{max}.$$

Czyli z (8.5) i niezależności wyborów zbiorów własności, drzewa BFS o korzeniach w v i v' ($v, v' \in \mathcal{V}^*$) połączyły się zanim zostały rozpatrzone wierzchołki w odległości $k' + 1$ lub

$$\begin{aligned} &\Pr \left\{ \forall w \in \mathcal{V} \setminus (\mathcal{N}_{k'}(v) \cup \mathcal{N}_{k'}(v')) D(w) \cap A = \emptyset \text{ lub } D(w) \cap A' = \emptyset \right\} \leq \\ &\leq \left(1 - \frac{\binom{|A|}{1} \binom{|A'|}{1} \binom{m - |B| - |A| - |A'|}{d-2}}{\binom{m - |B|}{d}} \right)^{n - 2n_{max}} \leq \\ &\leq \exp \left(- \frac{\lambda^2 (d-1)^3 dn}{(m - |B| - d + 1)(m - |B| - d + 3)} \frac{\binom{m - |B| - |A| - |A'|}{d-2}}{\binom{m - |B|}{d-2}} \left(1 - \frac{2n_{max}}{n} \right) \right) \leq \\ &\leq \exp \left(- \frac{\lambda^2 (d-1)^3 d^4 n^2}{d^3 n m^2} \left(1 - O \left(\frac{n_{max}}{n} + \frac{dn_{max}}{m} \right) \right) \right) = \\ &= \exp(-3(1 + o(1)) \ln n) = o \left(\frac{1}{n^2} \right), \end{aligned}$$

ponieważ

$$\lambda = (1 + o(1)) \sqrt{\frac{3}{c^2} \left(\frac{d}{d-1}\right)^3 \frac{n}{\ln n}}.$$

Dlatego, asymptotycznie prawie na pewno wierzchołki z \mathcal{V}^* wchodzą w skład jednej składowej spójności. Niech G będzie podgrafem $G^{Jed}(\mathcal{V}, m, d)$ indukowanym na \mathcal{V}^* . Ponieważ asymptotycznie prawie na pewno dowolne $v, v' \in \mathcal{V}^*$, v i v' są połączone ścieżką o długości co najwyżej $2k' + 2$, więc

$$(8.17) \quad \text{diam}G \leq k_+,$$

gdzie $k_+ = 2k' + 2 \sim \ln n / \ln \ln n$. □

8.5.2. Dowód twierdzenia 8.1(i)

Dowód twierdzenia 8.1(i). Niech G będzie składową spójności zawierającą $\Omega(n)$ wierzchołków w $G^{Jed}(\mathcal{V}, m, d)$. Ograniczenie górne na długość średnicy wynika z lematu 8.10. W dodatku z lematu 8.9 asymptotycznie prawie na pewno dla dowolnego wierzchołka v ze składowej G istnieje $n - o(n_{max})$ wierzchołków w odległości większej niż $k_- \sim \ln / \ln \ln n$. Istnieje więc $\Omega(n)$ wierzchołków w G w odległości co najmniej k_- od v . To daje oszacowanie dolne na długość średnicy. □

8.5.3. Własności $G^{Jed}(\mathcal{V}, m, d)$ na progu przejścia fazowego

Wyniki analogiczne do tych z lematów 8.9 i 8.10 można uzyskać dla grafu $G^{Jed}(\mathcal{V}, m, d)$ przy $d(d-1)/m \sim c$ i $c > 1$.

Lemat 8.11. *Niech $d(d-1)n/m = c_n \rightarrow c$ Wtedy istnieje $n_1 = o(n_{max})$ i $k_- \asymp \ln n$, takie, że w $G^{Jed}(\mathcal{V}, m, d)$*

$$\Pr \left\{ \forall v \in \mathcal{V} |\mathcal{N}_{k_-}(v)| \leq n_1 \right\} = 1 - o(1)$$

Dowód. Zdefiniujmy P_+ , N_+ i C_{N_+} tak jak w podrozdziale 8.2. Niech

$$k_3 = \left\lfloor \frac{\ln N_+ - \frac{9}{2} \ln \ln N_+}{\ln C_{N_+}} \right\rfloor.$$

Z lematu 8.4 i (9.11) dla danego $v \in \mathcal{V}$ z prawdopodobieństwem $1 - o(1/n)$

$$|\mathcal{N}_{k_1}(v)| \leq 1 + \sum_{i=1}^{k_3} \frac{C'_{N_+}}{C_{N_+}} i^2 C_{N_+}^i \ln N_+ = O\left(k_3^2 C_{N_+}^{k_3} \ln N_+\right) = o(n_{max}).$$

Powyższe oszacowania wynikają z tego, że

$$\begin{aligned} \sum_{i=1}^{k_3} i^2 x^i &= x^2 \left(\sum_{i=2}^{k_3} i(i-1)x^{i-2} \right) + x \left(\sum_{i=2}^{k_3} ix^{i-1} \right) = \\ &= x^2 \left(\frac{x^{k_3+1} - 1}{x-1} \right)'' + x \left(\frac{x^{k_3+1} - 1}{x-1} \right)'. \end{aligned}$$

□

Lemat 8.12. Niech $d(d-1)n/m = c_n \rightarrow c$, wtedy asymptotycznie prawie na pewno dla dużych n graf $G^{Jed}(\mathcal{V}, m, d)$ składa się ze składowych o co najwyżej $\frac{16c}{(\sqrt{c}-1)^2(c-1)} \ln n$ wierzchołkach i co najwyżej jednej składowej o co najmniej $\sqrt{n \ln n}$ wierzchołkach. Jeśli G jest składową o $\Omega(\sqrt{n \ln n})$ wierzchołkach, to istnieje k_+ , $k_+ \asymp \ln n$, takie, że asymptotycznie prawie na pewno

$$\text{diam}G \leq k_+.$$

Dowód. Niech $c > 0$ będzie stałą i $d(d-1)n/m = c_n \rightarrow c$. Zdefiniujemy P_- i P_+ tak jak w podrozdziale 8.2. Z (8.10) i (8.11)

$$\begin{aligned} N_+P'_+ &= C'_{N_+}, & N_-P'_- &= C'_{N_-}, \\ N_+P_+ &= C_{N_+}, & N_-P_- &= C_{N_-}, \end{aligned}$$

dla $N_+ = n$, $N_- = n - n_{max}$ i pewnych

$$\begin{aligned} C'_{N_+}, C'_{N_-} &\rightarrow C' = \begin{cases} \frac{d}{d-1}c & \text{dla } d = O(1); \\ c & \text{dla } d \rightarrow \infty, \end{cases} \\ C_{N_+}, C_{N_-} &\rightarrow C = c. \end{aligned}$$

Niech

$$\begin{aligned} b &= \frac{C}{2(\sqrt{C}-1)^2}, \\ t &= \frac{16C}{(\sqrt{C}-1)^2(C-1)} \ln N_- \end{aligned}$$

i

$$k_4 = \left\lceil \frac{1 \ln N_- - \ln \ln N_- + 2 \ln 2 - \ln b}{2 \ln C_{N_-}} \right\rceil + \left\lceil \frac{16C}{(\sqrt{C}-1)^2(C-1)} \ln N_- \right\rceil + 1.$$

Niech $v \in \mathcal{V}$. Z lematu 8.8 z prawdopodobieństwem $1 - o(1/n)$ v jest zawarte w składowej o co najwyżej $t - 1$ wierzchołkach lub

$$(8.18) \quad \exists_{i_0 \leq t} |\Gamma_{i_0}(v)| \geq \frac{4C}{(\sqrt{C}-1)^2} \ln N + 1.$$

Niech \mathcal{V}^* będzie zbiorem wierzchołków z \mathcal{V} , dla których jest spełnione (8.18).

Z lematu 8.6 wiemy, że z prawdopodobieństwem $1 - o(1/n)$ co najmniej $\frac{4C}{(\sqrt{C}-1)^2} \ln N$ wierzchołków z $\Gamma_{i_0}(v)$ dodało $d - 1$ elementów do $D(V_{BFS})$. Dlatego następane kroki możemy ograniczyć z dołu przez kolejne kroki procesu gałązkowego \mathcal{B}_- , w którym zachodzi $g_{i_0} \geq \frac{4C}{(\sqrt{C}-1)^2} \ln N$. Możemy stosować to ograniczenie tak długo jak $|\mathcal{N}(v)| \leq n_{max}$. Zdefiniujmy teraz zdarzenia

$$\begin{aligned} A_0 &- \exists_{i_0 \leq t} \forall_{i < i_0} |\Gamma_i(v)| < 3\sqrt{N_- \ln N_-} \text{ i } |\Gamma_{i_0}(v)^-| \geq 3\sqrt{N_- \ln N_-}; \\ B_0 &- \exists_{i_0 \leq t} \forall_{i < i_0} |\Gamma_i(v)| < 3\sqrt{N_- \ln N_-} \\ &\text{ i } \frac{4C}{(\sqrt{C}-1)^2} \ln N \leq |\Gamma_{i_0}(v)^-| < 3\sqrt{N_- \ln N_-}. \end{aligned}$$

Dla danej wartości i_0 oraz $i > 0$ zdefiniujemy zdarzenia

$$\begin{aligned} A_i &- |\Gamma_{i+i_0}(v)| \geq 3\sqrt{N_- \ln N_-}; \\ B_i &- |\Gamma_{i+i_0}(v)| < 3\sqrt{N_- \ln N_-} \text{ i } |\Gamma_{i+i_0}(v)^-| \geq b_{i+i_0} C_{N_-}^i \ln N_-, \end{aligned}$$

gdzie b_j jest zdefiniowane jak w lemacie 8.2 i $\Gamma_j(v)^-$ jest zbiorem tych wierzchołków z $\Gamma_j(v)$, które dodały $d-1$ elementów do $D(V_{BFS})$. Dlatego, jeśli $v \in \mathcal{V}^*$, to

$$\Pr\{A_0 \cup B_0\} \geq 1 - p \quad \text{dla pewnego } p = o\left(\frac{1}{n}\right).$$

Zauważmy, że dla $j = O(\ln n)$ zdarzenie $\bigcap_{i=0}^j B_i$ implikuje

$$|\Gamma_{j+i_0}(v)^-| \geq b_{j+i_0} C_{N_-}^j \ln N_- \quad \text{i} \quad |\mathcal{N}_{j+i_0}(v)| = O\left(\ln n \sqrt{n \ln n}\right) = o(n_{max}).$$

Dlatego kolejne kroki procedury BFS można ograniczyć z dołu przez \mathcal{B}_- z $g_j \geq b_{j+i_0} C_{N_-}^j$. Zatem z lematu 8.2

$$\Pr\left\{A_{j+1} \cup B_{j+1} \mid \bigcap_{i=i_0}^j B_i\right\} \geq 1 - q \quad \text{dla } q = \frac{1}{N_-^{\frac{3}{2}}}.$$

Dla danego $i_0 \leq t$, niech k będzie najmniejszym indeksem takim, że zachodzi $b_k C_{N_-}^{k-i_0} \ln N_- > 3\sqrt{N_- \ln N_-}$. Oczywiście $k \leq k_4$ i $\bigcap_{i=i_0}^{k-i_0} B_i = \emptyset$, dlatego z lematu 3.8, ponieważ A_i i B_i są rozłączne,

$$\Pr\left\{\exists_{i \leq k_4} |\Gamma_i(v)| \geq 3\sqrt{N_- \ln N_-}\right\} \geq (1-p)(1-q)^{k_4} \geq 1-p-k_4q = 1-o\left(\frac{1}{n}\right).$$

Dla danego wierzchołka $v \in \mathcal{V}^*$, niech $k(v)$ będzie najmniejszym indeksem takim, że

$$|\Gamma_{k(v)}(v)| \geq 3\sqrt{N_- \ln N_-}.$$

Z powyższych rozważań wynika, że z prawdopodobieństwem $1 - o(1/n)$ zachodzi $k(v) \leq k_4$. Dlatego z lematu 8.5 i ograniczenia z dołu przez proces gałązkowy \mathcal{B}_-

$$|\mathcal{N}_{k(v)}(v)| = O(k_4 \sqrt{N_- \ln N_-}) = o(n_{max}).$$

W dodatku, z lematu 8.6 z prawdopodobieństwem $1 - o(1/n)$ co najmniej $\frac{2}{3}3\sqrt{N_- \ln N_-}$ wierzchołków z $\Gamma_{k(v)}(v)$ dodaje co najmniej $d-1$ elementów do $D(V_{BFS})$, dlatego z prawdopodobieństwem $1 - o(1/n)$

$$|D(\Gamma_{k(v)}(v)) \setminus D(\mathcal{N}_{k(v)-1}(v))| \geq 2(d-1)\sqrt{N_- \ln N_-}.$$

Dla dowolnych dwóch wierzchołków $v, v' \in \mathcal{V}^*$, niech

$$\begin{aligned} B &= B(v, v') = D(\mathcal{N}_{k(v)-1}(v) \cup \mathcal{N}_{k(v)-1}(v')); \\ A &= A(v) = D(\Gamma_{k(v)}(v)) \setminus D(\mathcal{N}_{k(v)-1}(v)); \\ A' &= A'(v') = D(\Gamma_{k(v')}(v')) \setminus D(\mathcal{N}_{k(v')-1}(v')). \end{aligned}$$

Wtedy asymptotycznie prawie na pewno dla wszystkich par $v, v' \in \mathcal{V}^*$

$$|B| + |A| + |A'| \leq 2dn_{max}, \quad |A| \geq (d-1)\lambda, \quad |A'| \geq (d-1)\lambda,$$

gdzie $\lambda = 2\sqrt{N_- \ln N_-}$. W dodatku

$$|\mathcal{V} \setminus (\mathcal{N}_{k_2}(v) \cup \mathcal{N}_{k_2}(v'))| \geq n - 2n_{max}.$$

Dlatego z (8.5), drzewa BFS o korzeniach w v i v' połączyły się zanim zostały rozpatrzone wierzchołki w odległości $k(v)$ i $k(v')$ lub

$$\begin{aligned} & \Pr \left\{ \forall w \in \mathcal{V} \setminus (\mathcal{N}_{k_2}(v) \cup \mathcal{N}_{k_2}(v')) D(w) \cap A = \emptyset \text{ or } D(w) \cap A' = \emptyset \right\} \leq \\ & \leq \left(1 - \frac{\binom{|A|}{1} \binom{|A'|}{1} \binom{m-|B|-|A|-|A'|}{d-2}}{\binom{m-|B|}{d}} \right)^{n-2n_{max}} \leq \\ & \leq \exp \left(-\frac{\lambda^2 (d-1) d^2 (d-1)^2 n^2}{dn m^2} \left(1 - O \left(\frac{n_{max}}{n} + \frac{dn_{max}}{m} \right) \right) \right) = \\ & = \exp \left(-4 \frac{d-1}{d} c^2 (1+o(1)) \ln n \right) = o \left(\frac{1}{n^2} \right), \end{aligned}$$

ponieważ

$$\lambda^2 = (1+o(1))4n \ln n \quad \text{i} \quad c > 1.$$

Dlatego asymptotycznie prawie na pewno dowolne dwa wierzchołki z \mathcal{V}^* są połączone ścieżką długości co najwyżej $2k_4 + 2$. Stąd, jeśli G jest składową spójności indukowaną na \mathcal{V}^*

$$(8.19) \quad \text{diam} G \leq 2k_4 + 2.$$

To kończy dowód, gdyż $2k_4 + 2 \asymp \ln n$. □

8.5.4. Dowód twierdzenia 8.1(ii)

Dowód twierdzenia 8.1(ii). Niech G będzie składową spójności zawierającą $\Omega(n)$ wierzchołków w $G^{Jed}(\mathcal{V}, m, d)$. Jak w dowodzie twierdzenia 8.1(i) ograniczenie górne i dolne na długość średnicy wynikają z lematów 8.12 i 8.11. □

9. Wierzchołki izolowane

W grafach losowych zagadnienie spójności jest ściśle związane z zagadnieniem istnienia wierzchołka izolowanego. Często funkcje progowe dla obu własności się pokrywają. Dlatego pierwszym krokiem w kierunku zbadania spójności grafu $G_s(\mathcal{V}, m, \mathcal{P}_{(m)})$ i określenia funkcji progowej dla tej własności jest badanie rozkładu liczby wierzchołków izolowanych. Wyniki w tym rozdziale są uogólnieniem wyniku dotyczącego liczby wierzchołków izolowanych w $G^{Jed}(\mathcal{V}, m, d)$ udowodnionego przez Godeharda i Jaworskiego w [26].

9.1. Twierdzenie o liczbie wierzchołków izolowanych

Przypomnijmy, że w losowym grafie przecięć $G_s(\mathcal{V}, m, \mathcal{P}_{(m)})$ przez $D(v)$ oznaczamy zbiór własności wierzchołka $v \in \mathcal{V}$ oraz $Z_n = |D(v)|$ jest zmienną losową o rozkładzie $\mathcal{P}_{(m)}$. Załóżmy, że $s \geq 1$. Zauważmy, że jeśli $|D(v)| < s$, to wierzchołek v jest izolowany w $G_s(\mathcal{V}, m, \mathcal{P}_{(m)})$. W dodatku liczba wierzchołków w $G_s(\mathcal{V}, m, \mathcal{P}_{(m)})$ takich, że $|D(v)| < s$ ma rozkład dwumianowy $\text{Bin}(n, q)$, gdzie $q = \Pr\{Z_n < s\}$. W związku z tym, że chcemy wykluczyć istnienie takich wierzchołków izolowanych będziemy zakładać, że $\Pr\{Z_n \geq s\} = 1$.

Twierdzenie 9.1. *Niech Z_n będzie zmienną losową o rozkładzie $\mathcal{P}_{(m)}$, $\Pr\{Z_n \geq s\} = 1$,*

$$\frac{n(d_n)_s^2}{s! m^s} = \ln n + c_n, \quad c_n \rightarrow c$$

dla pewnej stałej $c > 0$,

$$\frac{(Z_n)_s - (d_n)_s}{(d_n)_s \omega(n)} = \mathbb{A}_n,$$

gdzie \mathbb{A}_n jest zbieżna według rozkładu do \mathbb{A} oraz \mathbb{A}_n i \mathbb{A} mają wartości w skończonych zbiorach o tym samym rozmiarze.

- (i) Jeżeli $\omega(n) = o(\frac{1}{\ln n})$, to w $G_s(\mathcal{V}, m, \mathcal{P}_{(m)})$ liczba wierzchołków izolowanych dąży według rozkładu do zmiennej losowej o rozkładzie Poissona $\text{Po}(\lambda)$ z parametrem $\lambda = e^{-c}$.
- (ii) Jeżeli $\omega(n) = \frac{1}{\ln n}$, to w $G_s(\mathcal{V}, m, \mathcal{P}_{(m)})$ liczba wierzchołków izolowanych dąży według rozkładu do zmiennej losowej o rozkładzie Poissona $\text{Po}(\lambda)$ z parametrem $\lambda = e^{-c - \mathbb{E}\mathbb{A}} \mathbb{E}(e^{-\mathbb{A}})$.
- (iii) Jeżeli $\omega(n) \ln n \rightarrow \infty$, $\mathbb{E}(\mathbb{A}_n) = 0$ i \mathbb{A} ma co najmniej dwie wartości, to asymptotycznie prawie na pewno $G_s(\mathcal{V}, m, \mathcal{P}_{(m)})$ ma nieskończenie wiele wierzchołków izolowanych

Dowód został podzielony na kilka części. W podrozdziale 9.2 udowodnimy pomocnicze lematy i fakty. W podrozdziale 9.3 podamy dowód głównego

twierdzenia wykorzystujący wyniki z podrozdziału 9.2. Natomiast w podrozdziale 9.4 pokażemy i udowodnimy wnioski z niego wynikające i inne twierdzenie dotyczące liczby wierzchołków izolowanych, którego dowód przebiega podobnie do dowodu twierdzenia 9.1.

9.2. Wyniki pomocnicze

Dowody w tym podrozdziale będą dotyczyły losowego grafu $G_s(\mathcal{V}, m, \mathcal{P}_{(m)})$ w przestrzeni warunkowej z danymi mocami zbiorów $D(v_i)$. W związku z tym podamy pomocniczą definicję.

Definicja 9.1. Niech $\mathcal{V} = \{v_1, \dots, v_n\}$ i $\mathcal{W} = \{w_1, \dots, w_{m(n)}\}$ oraz niech $s \geq 1$ będzie stałą liczbą całkowitą. Dla danego ciągu wektorów liczb naturalnych $\bar{d} = \overline{d(n)} = (d_1(n), d_2(n), \dots, d_n(n))$, niech $D(v_1), \dots, D(v_n)$ będzie rodziną podzbiorów zbioru \mathcal{W} taką, że dla każdego $1 \leq i \leq n$, $D(v_i)$ jest wybrany w sposób jednostajny spośród d_i -elementowych podzbiorów zbioru \mathcal{W} , tzn. dla danego d_i -elementowego podzbioru $A \subseteq \mathcal{W}$

$$\Pr(D(v_i) = A) = \binom{m}{d_i}^{-1}.$$

Losowym grafem przecięć o danych mocach zbiorów własności $G_s(\mathcal{V}, m, \bar{d})$ nazywamy graf o zbiorze wierzchołków \mathcal{V} i zbiorze krawędzi $E = \{\{v_i, v_j\} : |D(v_i) \cap D(v_j)| \geq s\}$.

Oczywiście

$$(9.1) \quad \Pr \{G_s(\mathcal{V}, m, \bar{d}) = G\} = \\ = \Pr \{G_s(\mathcal{V}, m, \mathcal{P}_{(m)}) = G \mid (|D(v_1)|, \dots, |D(v_n)|) = \bar{d}\}.$$

Dodatkowo oznaczmy przez A_{ij} zdarzenie, że wierzchołki v_i i v_j są połączone krawędzią w $G_s(\mathcal{V}, m, \bar{d})$. Z definicji wynika, że $d_i = d_i(n)$ jest liczbą własności wierzchołka v_i w grafie $G_s(\mathcal{V}, m, \bar{d})$ ($1 \leq i \leq n$). W dalszych rozważaniach będziemy zakładać, że $s \leq d_i$, gdyż wierzchołki z mniejszą liczbą własności są zawsze izolowane. Niech $\Delta = \Delta(n) = \max\{|D(v)| : v \in \mathcal{V}\}$ będzie maksymalną mocą zbioru własności wierzchołków. Będziemy zakładać, że $\Delta^2/m \rightarrow 0$. Także $O(\cdot)$ i $o(\cdot)$ będą jednostajnie ograniczone względem wszystkich możliwych wektorów \bar{d} takich, że $\max\{d_i : 1 \leq i \leq n\} \leq \Delta$

Dla tak określonego \bar{d} i $G_s(\mathcal{V}, m, \bar{d})$ z lematu 3.6

$$(9.2) \quad \Pr\{A_{ij}\} = \frac{(d_i)_s (d_j)_s}{m^s s!} \left(1 + O\left(\frac{\Delta^2}{m}\right)\right) = O\left(\frac{\Delta^{2s}}{m^s}\right) = O\left(\frac{\Delta^2}{m}\right).$$

9.2.1. Podzbiory wierzchołków izolowanych

Dla danej stałej k chcemy oszacować prawdopodobieństwo, że wierzchołki $v_1, v_2, v_3, \dots, v_k$ są izolowane w $G_s(\mathcal{V}, m, \bar{d})$. Najpierw, dla danego \bar{d} , oszacujemy prawdopodobieństwo, że żadne dwa wierzchołki ze zbioru $\{v_1, \dots, v_k\}$ nie są połączone krawędzią.

Fakt 9.1. W $G_s(\mathcal{V}, m, \bar{d})$

$$\Pr \left\{ \bigcap_{1 \leq i < j \leq k} A_{ij}^c \right\} = 1 - \left(1 + O \left(\frac{k^2 \Delta^2}{m} \right) \right) \sum_{1 \leq i < j \leq k} \Pr \{ A_{ij} \}.$$

Dowód. Zauważmy, że dla parami różnych indeksów i_1, i_2, i_3, i_4 zdarzenia $A_{i_1 i_2}$ i $A_{i_3 i_4}$ są niezależne. Tak samo niezależne są $A_{i_1 i_2}$ i $A_{i_2 i_3}$. Dlatego

$$\begin{aligned} \Pr \left\{ \bigcup_{1 \leq i < j \leq k} A_{ij} \right\} &\geq \sum_{1 \leq i < j \leq k} \Pr \{ A_{ij} \} - \frac{1}{2} \sum_{\substack{1 \leq i_1 < j_1 \leq k \\ 1 \leq i_2 < j_2 \leq k \\ (i_1, j_1) \neq (i_2, j_2)}} \Pr \{ A_{i_1 j_1} \cap A_{i_2 j_2} \} = \\ &= \sum_{1 \leq i < j \leq k} \Pr \{ A_{ij} \} - \frac{1}{2} \sum_{\substack{1 \leq i_1 < j_1 \leq k \\ 1 \leq i_2 < j_2 \leq k \\ (i_1, j_1) \neq (i_2, j_2)}} \Pr \{ A_{i_1 j_1} \} \Pr \{ A_{i_2 j_2} \} \geq \\ &\geq \left(\sum_{1 \leq i < j \leq k} \Pr \{ A_{ij} \} \right) \left(1 - \frac{1}{2} \sum_{1 \leq i < j \leq k} \Pr \{ A_{ij} \} \right) \end{aligned}$$

oraz

$$\Pr \left\{ \bigcup_{1 \leq i < j \leq k} A_{ij} \right\} \leq \sum_{1 \leq i < j \leq k} \Pr \{ A_{ij} \}.$$

Z (9.2) $\sum_{1 \leq i < j \leq k} \Pr \{ A_{ij} \} = O(k^2 \Delta^{2s} / m^s) = O(k^2 \Delta^2 / m)$, więc

$$\begin{aligned} \Pr \left\{ \bigcap_{1 \leq i < j \leq k} A_{ij}^c \right\} &= 1 - \Pr \left\{ \bigcup_{1 \leq i < j \leq k} A_{ij} \right\} = \\ &= 1 - \left(1 + O \left(\frac{k^2 \Delta^2}{m} \right) \right) \sum_{1 \leq i < j \leq k} \Pr \{ A_{ij} \}. \end{aligned}$$

□

W następnym fakcie oszacujemy prawdopodobieństwo, że v_1, \dots, v_k są izolowane w $G_s(\mathcal{V}, m, \bar{d})$ pod warunkiem, że v_1, \dots, v_k nie są między sobą połączone krawędziami.

Fakt 9.2. W $G_s(\mathcal{V}, m, \bar{d})$

$$\begin{aligned} \Pr \left\{ \bigcap_{k < j \leq n} \bigcap_{1 \leq i \leq k} A_{ij}^c \mid \bigcap_{1 \leq i < i' \leq k} A_{ii'}^c \right\} &= \\ &= \prod_{k < j \leq n} \left(1 - (1 + O(k \Delta^2 / m)) \sum_{i=1}^k \Pr \{ A_{ij} \} \right). \end{aligned}$$

Dowód. Niech \mathcal{D} będzie rodziną wszystkich ciągów zbiorów (z powtórzeniami) $\{D(v_1), \dots, D(v_k)\}$, dla których zachodzi: $|D(v_i)| = d_i$, $D(v_i) \subseteq \mathcal{W}$ oraz $|D(v_i) \cap D(v_j)| \leq s - 1$ dla każdych $1 \leq i, j \leq k$. Dla danego ciągu $\mathbb{D} \in \mathcal{D}$ oznaczmy

$$\Pr_{\mathbb{D}} \{ \cdot \} = \Pr \{ \cdot \mid \mathbb{D} = \{D(v_1), \dots, D(v_k)\} \}.$$

Oszacujemy wartość $\Pr_{\mathbb{D}}\{\cap_{1 \leq i \leq k} A_{ij}^c\}$. Dla danego $k \leq j \leq n$

$$\Pr_{\mathbb{D}}\left\{\bigcup_{1 \leq i \leq k} A_{ij}\right\} \geq \sum_{i=1}^k \Pr_{\mathbb{D}}\{A_{ij}\} - \frac{1}{2} \sum_{1 \leq i_1 \neq i_2 \leq k} \Pr_{\mathbb{D}}\{A_{i_1 j} \cap A_{i_2 j}\}.$$

oraz

$$\Pr_{\mathbb{D}}\{A_{ij}\} = \Pr\{A_{ij}\}.$$

Niech $\mathbb{B}_{ij}(t)$ będzie zdarzeniem, że zbiory $D(v_i), D(v_j)$ przecinają się na dokładnie t wierzchołkach ($t \leq s-1$). Wówczas

$$\begin{aligned} \Pr\{A_{i_1 j} \cap A_{i_2 j} | \mathbb{B}_{i_1 i_2}(t)\} &\leq \frac{\binom{d_{i_1}}{s} \binom{d_{i_2}-t}{1} \binom{m-s-1}{d_j-s-1}}{\binom{m}{d_j}} \leq \\ &\leq \frac{(d_{i_1})_s d_{i_2} (m-s-1)_{d_j-s-1} d_j!}{s! (m)_{d_j} (d_j-s-1)!} = \\ &= \frac{d_{i_2} (d_j-s) (d_{i_1})_s (d_j)_s}{m s! m^s} \left(1 + O\left(\frac{\Delta^2}{m}\right)\right) \leq \\ &\leq \frac{\Delta^2 (d_{i_1})_s (d_j)_s}{m s! m^s} \left(1 + O\left(\frac{\Delta^2}{m}\right)\right) \sim \\ &\sim \frac{\Delta^2}{m} \Pr\{A_{i_1 j}\}. \end{aligned}$$

Stąd

$$\begin{aligned} \Pr_{\mathbb{D}}\{A_{i_1 j} \cap A_{i_2 j}\} &= \\ &= \sum_{t=0}^{s-1} \Pr\{A_{i_1 j} \cap A_{i_2 j} | \mathbb{B}_{i_1 i_2}(t)\} \Pr_{\mathbb{D}}\{\mathbb{B}_{i_1 i_2}(t)\} = O\left(\frac{\Delta^2}{m}\right) \Pr\{A_{i_1 j}\} \end{aligned}$$

oraz $O(\cdot)$ jest jednostajnie ograniczone dla wszystkich wartości \mathbb{D} . Zatem

$$\begin{aligned} \sum_{i=1}^k \Pr\{A_{ij}\} &\geq \Pr_{\mathbb{D}}\{\bigcup_{1 \leq i \leq k} A_{ij}\} \geq \\ &\geq \sum_{i=1}^k \Pr\{A_{ij}\} - \frac{1}{2} \sum_{1 \leq i_1 \neq i_2 \leq k} \Pr_{\mathbb{D}}\{A_{i_1 j} \cap A_{i_2 j}\} \geq \\ &\geq \left(\sum_{i=1}^k \Pr\{A_{ij}\}\right) \left(1 + O\left(\frac{k\Delta^2}{m}\right)\right). \end{aligned}$$

To oznacza, że

$$\Pr_{\mathbb{D}}\{\cap_{1 \leq i \leq k} A_{ij}^c\} = 1 - \left(1 + O\left(\frac{k\Delta^2}{m}\right)\right) \sum_{i=1}^k \Pr\{A_{ij}\}.$$

Oczywiście $\cap_{1 \leq i < i' \leq k} A_{ii'}^c = \cup_{\mathbb{D} \in \mathcal{D}} \{\mathbb{D} = \{D(v_1), \dots, D(v_k)\}\}$. W dodatku dla danego $\{D(v_1), \dots, D(v_k)\}$ zdarzenia $\cap_{1 \leq i \leq k} A_{ij_1}^c, \dots, \cap_{1 \leq i \leq k} A_{ij_t}^c$ są niezależne dla $j_1, \dots, j_t > k$, zatem

$$\begin{aligned}
& \Pr \left\{ \bigcap_{k < j \leq n} \bigcap_{1 \leq i \leq k} A_{ij}^c \mid \bigcap_{1 \leq i < i' \leq k} A_{ii'}^c \right\} = \\
& = \sum_{\mathbb{D} \in \mathcal{D}} \Pr_{\mathbb{D}} \left\{ \bigcap_{k < j \leq n} \bigcap_{1 \leq i \leq k} A_{ij}^c \right\} \cdot \\
& \quad \cdot \Pr \left\{ \mathbb{D} = \{D(v_1), \dots, D(v_k)\} \mid \bigcap_{1 \leq i < i' \leq k} A_{ii'}^c \right\} = \\
& = \sum_{\mathbb{D} \in \mathcal{D}} \prod_{k < j \leq n} \Pr_{\mathbb{D}} \left\{ \bigcap_{1 \leq i \leq k} A_{ij}^c \right\} \cdot \\
& \quad \cdot \Pr \left\{ \mathbb{D} = \{D(v_1), \dots, D(v_k)\} \mid \bigcap_{1 \leq i < i' \leq k} A_{ii'}^c \right\} = \\
& = \sum_{\mathbb{D} \in \mathcal{D}} \prod_{k < j \leq n} \left(1 - (1 + O(k\Delta^2/m)) \sum_{i=1}^k \Pr\{A_{ij}\} \right) \cdot \\
& \quad \cdot \Pr \left\{ \mathbb{D} = \{D(v_1), \dots, D(v_k)\} \mid \bigcap_{1 \leq i < i' \leq k} A_{ii'}^c \right\} = \\
& = \prod_{k < j \leq n} \left(1 - (1 + O(k\Delta^2/m)) \sum_{i=1}^k \Pr\{A_{ij}\} \right).
\end{aligned}$$

□

9.2.2. Wierzchołki izolowane w $G_s(\mathcal{V}, m, \bar{d})$

Fakt 9.3. Niech k będzie stałą oraz $I = \{i_1, \dots, i_k\} \subseteq \{1, \dots, n\} = [n]$ będzie podzbiorem indeksów. Prawdopodobieństwo, że wierzchołki $\{v_{i_1}, \dots, v_{i_k}\}$ są izolowane w $G_s(\mathcal{V}, m, \bar{d})$ jest równe

$$\begin{aligned}
& \Pr \left\{ \bigcap_{i, i' \in I} A_{ii'}^c \cap \bigcap_{j \in [n] \setminus I} \bigcap_{i \in I} A_{ij}^c \right\} = \\
& = \exp \left(-O \left(\frac{\Delta^2}{m} \right) \right) \prod_{j \in [n] \setminus I} \exp \left(- \sum_{i \in I} \Pr\{A_{ij}\} \left(1 + O \left(\frac{\Delta^2}{m} \right) \right) \right).
\end{aligned}$$

Dowód. Z faktów 9.1 i 9.2 wynika, że dla danych d_1, \dots, d_n , prawdopodobieństwo, że wierzchołki v_1, \dots, v_k są izolowane jest równe

$$\begin{aligned}
& \Pr \left\{ \bigcap_{1 \leq i < i' \leq k} A_{ii'}^c \cap \bigcap_{k < j \leq n} \bigcap_{1 \leq i \leq k} A_{ij}^c \right\} = \\
& = \Pr \left\{ \bigcap_{k < j \leq n} \bigcap_{1 \leq i \leq k} A_{ij}^c \mid \bigcap_{1 \leq i < i' \leq k} A_{ii'}^c \right\} \Pr \left\{ \bigcap_{1 \leq i < i' \leq k} A_{ii'}^c \right\} = \\
& = \left(\prod_{k < j \leq n} \left(1 - \left(1 + O \left(\frac{k\Delta^2}{m} \right) \right) \sum_{i=1}^k \Pr\{A_{ij}\} \right) \right) \cdot \\
& \quad \cdot \left(1 - \left(1 + O \left(\frac{k^2\Delta^2}{m} \right) \right) \sum_{1 \leq i < i' \leq k} \Pr\{A_{ii'}\} \right).
\end{aligned}$$

Jeżeli dodatkowo założymy, że k jest stałą oraz skorzystamy z (9.2), to otrzymujemy, że

$$\begin{aligned}
& \Pr \left\{ \bigcap_{1 \leq i < i' \leq k} A_{ii'}^c \cap \bigcap_{k < j \leq n} \bigcap_{1 \leq i \leq k} A_{ij}^c \right\} = \\
& = \exp \left(-O \left(\frac{\Delta^2}{m} \right) \right) \prod_{k < j \leq n} \exp \left(- \sum_{i=1}^k \Pr\{A_{ij}\} \left(1 + O \left(\frac{\Delta^2}{m} \right) \right) \right).
\end{aligned}$$

Analogiczny wynik można uzyskać dla dowolnego k -elementowego podzbioru zbioru indeksów – wystarczy tylko przenumerować wierzchołki. \square

Fakt 9.4. Niech $k > 0$ będzie stałą całkowitą, $\Delta^{2s} = (\Delta(n))^{2s} = O\left(\frac{m^s \ln n}{n}\right)$ oraz $I = \{i_1, \dots, i_k\} \subseteq \{1, \dots, n\} = [n]$ będzie podzbiorem indeksów. Dla dowolnego \bar{d} takiego, że $\max\{d_i : 1 \leq i \leq n\} \leq \Delta$,

$$\begin{aligned} \Pr \left\{ \bigcap_{i, i' \in I} A_{ii'}^c \cap \bigcap_{j \in [n] \setminus I} \bigcap_{i \in I} A_{ij}^c \right\} &= \\ &= \exp(o(1)) \exp \left(- \sum_{j \in [n]} \sum_{i \in I} \frac{(d_i)_s (d_j)_s}{s! m^s} \right), \end{aligned}$$

gdzie $o(1)$ jest jednostajnie ograniczone dla wszystkich wektorów \bar{d} takich, że $\max\{d_i : 1 \leq i \leq n\} \leq \Delta$.

Dowód. Najpierw zauważmy, że skoro

$$\Delta^{2s} = (\Delta(n))^{2s} = O\left(\frac{m^s \ln n}{n}\right),$$

to

$$\frac{(d_i)_s (d_j)_s}{s! m^s} \leq \frac{\Delta^{2s}}{s! m^s} = O\left(\frac{\ln n}{n}\right) \quad \text{i} \quad \frac{\Delta^2}{m} = o\left(\frac{1}{\ln n}\right) = o(1).$$

Natomiast z (9.2)

$$\Pr\{A_{ij}\} = O\left(\frac{\Delta^{2s}}{m^s}\right),$$

czyli z faktu 9.3 i (9.2) mamy

$$\begin{aligned} \Pr \left\{ \bigcap_{i, i' \in I} A_{ii'}^c \cap \bigcap_{j \in [n] \setminus I} \bigcap_{i \in I} A_{ij}^c \right\} &= \\ &= \exp \left(O\left(\frac{\Delta^2}{m}\right) \right) \prod_{j \in [n] \setminus I} \exp \left(- \sum_{i \in I} \Pr\{A_{ij}\} \left(1 + O\left(\frac{\Delta^2}{m}\right) \right) \right) = \\ &= \exp(o(1)) \prod_{j \in [n] \setminus I} \exp \left(- O\left(\frac{\Delta^2}{m}\right) \sum_{i \in I} \Pr\{A_{ij}\} \right) \cdot \\ &\quad \cdot \exp \left(- \sum_{j \in [n] \setminus I} \sum_{i \in I} \Pr\{A_{ij}\} \right) = \\ &= \exp \left(o(1) + O\left(n \cdot \frac{\Delta^{2s}}{m^s} \cdot \frac{\Delta^2}{m} \right) \right) \cdot \\ &\quad \cdot \exp \left(- \sum_{j \in [n] \setminus I} \sum_{i \in I} \frac{(d_i)_s (d_j)_s}{s! m^s} \left(1 + O\left(\frac{\Delta^2}{m}\right) \right) \right) = \\ &= \exp(o(1)) \exp \left(- \sum_{j \in [n]} \sum_{i \in I} \frac{(d_i)_s (d_j)_s}{s! m^s} + \right. \\ &\quad \left. + \sum_{j \in I} \sum_{i \in I} \frac{(d_i)_s (d_j)_s}{s! m^s} + O\left(n \cdot \frac{\Delta^{2s}}{s! m^s} \cdot \frac{\Delta^2}{m} \right) \right) = \\ &= \exp(o(1)) \exp \left(- \sum_{j \in [n]} \sum_{i \in I} \frac{(d_i)_s (d_j)_s}{s! m^s} \right). \end{aligned}$$

\square

Lemat 9.1. Niech M będzie stałą całkowitą, $\delta(n)$ ciągiem liczb dodatnich, $I \subseteq [n]$ skończonym podzbiorem indeksów, $\bar{\beta} = (\beta_1(n), \dots, \beta_M(n)) \in \mathbb{R}^n$ oraz $\bar{p} = (p_1(n), \dots, p_M(n)) \in [0, 1]^n$ będą ciągami wektorów. Jeżeli wektor $\bar{d} = \{d_1, \dots, d_n\}$ jest taki, że $d_i \in \{\beta_1(n), \dots, \beta_M(n)\}$ dla każdego $1 \leq i \leq n$, $\beta_l(n)^{2s} = O\left(\frac{m^s \ln n}{n}\right)$, $\mathcal{V}_l = \{i \in [n] : d_i = \beta_l(n)\}$, $|k_l| = |\mathcal{V}_l \cap I|$ oraz $p_l(n)n(1 - \frac{\delta(n)}{\ln n}) \leq |\mathcal{V}_l| \leq p_l(n)n(1 + \frac{\delta(n)}{\ln n})$ dla każdego $1 \leq l \leq M$, to

$$\begin{aligned} \Pr \left\{ \bigcap_{i, i' \in I} A_{ii'}^c \cap \bigcap_{j \in [n] \setminus I} \bigcap_{i \in I} A_{ij}^c \right\} &= \\ &= \exp(o(1) + O(\delta(n))) \exp \left(- \sum_{l'=1}^M k_{l'} \sum_{l=1}^M p_l(n) \frac{n(\beta_l)_s (\beta_{l'})_s}{s! m^s} \right). \end{aligned}$$

Dowód. Ustalmy wartości $\Delta = \max\{\beta_1(n), \dots, \beta_M(n)\}$ oraz $k_l = |\mathcal{V}_l \cap I|$. Skoro $\Delta^{2s} = O\left(\frac{m^s \ln n}{n}\right)$, to $\frac{n(\Delta)_s^2}{s! m^s} = O\left(\frac{\ln n}{n}\right)$. Dla I , takiego, że $|I| = k$ jest stałą, z faktu 9.4 wynika, że

$$\begin{aligned} \Pr \left\{ \bigcap_{i, i' \in I} A_{ii'}^c \cap \bigcap_{j \in [n] \setminus I} \bigcap_{i \in I} A_{ij}^c \right\} &= \\ &= \exp(o(1)) \exp \left(- \sum_{j \in [n]} \sum_{i \in I} \frac{(d_i)_s (d_j)_s}{s! m^s} \right) = \\ &= \exp(o(1)) \exp \left(- \sum_{l=1}^M \sum_{l'=1}^M |\mathcal{V}_l| k_{l'} \frac{(\beta_l)_s (\beta_{l'})_s}{s! m^s} \right) = \\ &= \exp(o(1) + O(\delta(n))) \exp \left(- \sum_{l'=1}^M k_{l'} \sum_{l=1}^M p_l(n) \frac{n(\beta_l)_s (\beta_{l'})_s}{s! m^s} \right). \end{aligned}$$

□

Niech $Y(\bar{d}) = Y(\bar{d}, n)$ będzie zmienną losową liczącą wierzchołki izolowane w $G_s(\mathcal{V}, m, \bar{d})$. Wtedy

$$(9.3) \quad Y(\bar{d}) = \sum_{i=1}^n \mathbb{I}_i, \text{ gdzie } \mathbb{I}_i = \begin{cases} 1 & \text{jeśli } v_i \text{ jest} \\ & \text{izolowany w } G_s(\mathcal{V}, m, \bar{d}); \\ 0 & \text{jeśli } v_i \text{ nie jest} \\ & \text{izolowany w } G_s(\mathcal{V}, m, \bar{d}). \end{cases}$$

Wtedy

$$\begin{aligned} \mathbb{E}(Y(\bar{d}))_k &= \\ &= k! \sum_{i_1 < \dots < i_k} \mathbb{E}(\mathbb{I}_{i_1} \cdot \dots \cdot \mathbb{I}_{i_k}) = k! \sum_{\substack{I \subseteq [n] \\ |I|=k}} \Pr \left\{ \bigcap_{i, i' \in I} A_{ii'}^c \cap \bigcap_{j \in [n] \setminus I} \bigcap_{i \in I} A_{ij}^c \right\}. \end{aligned}$$

Przyjmijmy oznaczenia z lematu 9.1. Niech $\overline{N(n)} = (N_1(n), \dots, N_M(n))$ będzie rodziną wektorów liczb całkowitych nieujemnych. Będziemy mówić, że \bar{d} realizuje $\overline{N(n)}$ (oznaczymy to przez $\bar{d} \triangleleft \overline{N(n)}$) jeśli dla \bar{d} mamy $|\mathcal{V}_l| = N_l(n)$ dla każdego $1 \leq l \leq M$. Zauważmy, że dla $\bar{d} \triangleleft \overline{N(n)}$ i $\bar{d}' \triangleleft \overline{N(n)}$, skoro \bar{d} i \bar{d}' różnią się jedynie kolejnością składowych, więc

$$(9.4) \quad \mathbb{E}(Y(\bar{d}))_k = \mathbb{E}(Y(\bar{d}'))_k = \mathbb{E}(Y(\overline{d_N}))_k,$$

gdzie $\overline{d_N}$ jest takie, że dla każdego $(N_1 + \dots + N_l) - N_l < i \leq (N_1 + \dots + N_l)$ i $1 \leq l \leq M$ zachodzi $d_i = \beta_l(n)$.

Lemat 9.2. Niech M będzie stałą, $\delta(n)$ będzie ciągiem liczb dodatnich dążącym do 0, $\overline{\beta} = (\beta_1(n), \dots, \beta_M(n)) \in \mathbb{R}^n$ i $\overline{p} = (p_1(n), \dots, p_M(n)) \in [0, 1]^n$ będą ciągami wektorów. Załóżmy dodatkowo, że ciągi $p_1(n), \dots, p_M(n)$ są ograniczone z dołu przez pewną stałą dodatnią. Mając dany wektor liczb $\overline{N}(n) = (N_1(n), \dots, N_M(n))$ taki, że

$$(9.5) \quad \forall_{1 \leq l \leq M} \quad p_l(n)n \left(1 - \frac{\delta(n)}{\ln n}\right) \leq N_l(n) \leq p_l(n)n \left(1 + \frac{\delta(n)}{\ln n}\right),$$

oraz $\sum_{l=1}^M N_l(n) = n,$

mamy

$$\mathbb{E}(Y(\overline{d_N}))_k = \exp(o(1) + O(\delta(n))) (A(\overline{p}(n), \overline{\beta}(n)))^k,$$

gdzie

$$A(\overline{p}(n), \overline{\beta}(n)) = \sum_{l_1=1}^M \exp\left(-\sum_{l_2=1}^M \left(p_{l_2}(n) \frac{n(\beta_{l_1})_s(\beta_{l_2})_s}{s!m^s}\right) + \ln(np_{l_1}(n))\right).$$

Dowód. Najpierw zauważmy, że ponieważ $N_l \rightarrow \infty$, to dla dowolnej stałej $k_l \in \mathbb{N}$,

$$\prod_{l=1}^M \left(\frac{(N_l)_{k_l}}{N_l^{k_l}} \cdot \frac{N_l^{k_l}}{(np_l)^{k_l}}\right) = \exp(o(1) + o(\delta(n))),$$

gdzie $\exp(o(1) + o(\delta(n)))$ jest jednostajnie ograniczone względem wszystkich wyborów $k_l \leq k$. Z lematu 9.1 otrzymujemy więc

$$\begin{aligned} & \mathbb{E}(Y(\overline{d_N}))_k = \\ & = k! \sum_{\substack{k_1, \dots, k_M \geq 0 \\ k_1 + \dots + k_M = k}} \left(\prod_{l_0=1}^M \binom{N_{l_0}}{k_{l_0}} \right) \cdot \\ & \quad \cdot \left(\exp(o(1) + O(\delta(n))) \exp\left(-\sum_{l_1=1}^M k_{l_1} \sum_{l_2=1}^M p_{l_2}(n) \frac{n(\beta_{l_1})_s(\beta_{l_2})_s}{s!m^s}\right) \right) = \\ & = \exp(o(1) + O(\delta(n))) \sum_{\substack{k_1, \dots, k_M \geq 0 \\ k_1 + \dots + k_M = k}} \frac{k!}{\prod_{l_0=1}^M (k_{l_0})!} \cdot \\ & \quad \cdot \left(\prod_{l'_0=1}^M (np_{l'_0}(n))^{k_{l'_0}} \right) \exp\left(-\sum_{l_1=1}^M k_{l_1} \sum_{l_2=1}^M \left(p_{l_2}(n) \frac{n(\beta_{l_1})_s(\beta_{l_2})_s}{s!m^s}\right)\right) \end{aligned}$$

$$\begin{aligned}
&= \exp(o(1) + O(\delta(n))) \\
&\quad \sum_{\substack{k_1, \dots, k_M \geq 0 \\ k_1 + \dots + k_M = k}} \frac{k!}{\prod_{l_0=1}^M (k_{l_0})!} \\
&\quad \cdot \exp\left(-\sum_{l_1=1}^M k_{l_1} \left(\sum_{l_2=1}^M \left(p_{l_2}(n) \frac{n(\beta_{l_1})_s (\beta_{l_2})_s}{s! m^s}\right) - \ln(np_{l_1}(n))\right)\right) = \\
&= \exp(o(1) + O(\delta(n))) \\
&\quad \sum_{\substack{k_1, \dots, k_M \geq 0 \\ k_1 + \dots + k_M = k}} \frac{k!}{\prod_{l_0=1}^M (k_{l_0})!} \\
&\quad \cdot \prod_{l_1=1}^M \left(\exp\left(-\sum_{l_2=1}^M \left(p_{l_2}(n) \frac{n(\beta_{l_1})_s (\beta_{l_2})_s}{s! m^s}\right) - \ln(np_{l_1}(n))\right)\right)^{k_{l_1}} = \\
&= \exp(o(1) + O(\delta(n))) \cdot \\
&\quad \cdot \left(\sum_{l_1=1}^M \exp\left(-\sum_{l_2=1}^M \left(p_{l_2}(n) \frac{n(\beta_{l_1})_s (\beta_{l_2})_s}{s! m^s}\right) + \ln(np_{l_1}(n))\right)\right)^k,
\end{aligned}$$

skąd wynika teza lematu. \square

9.3. Dowód twierdzenia o liczbie wierzchołków izolowanych

Założmy, że $\mathcal{P}_{(m)}$ jest takie, że $Z_n \in \{\beta_1(n), \dots, \beta_M(n)\}$ z prawdopodobieństwem równym 1. Dodatkowo ustalmy $p_l(n) = \Pr\{Z_n = \beta_l(n)\}$, $\bar{p} = \{p_1(n), \dots, p_M(n)\}$ i założmy, że w ciągach $(p_l(n))_{n \in \mathbb{N}}$ wszystkie wyrazy są większe od pewnej stałej dodatniej. Niech $\bar{\mathbb{N}} = (\mathbb{N}_1, \dots, \mathbb{N}_M)$ będzie wektorem losowym, takim że $\mathbb{N}_l = |\{i \in [n] : |D(v_i)| = \beta_l(n)\}|$, dla $1 \leq l \leq M$. Jeżeli $\delta(n) \rightarrow 0$ jest takie, że $\sqrt{\frac{\ln^3 n}{n}} = o(\delta(n))$, to z nierówności Chernoffa ((3.1) i (3.2)) asymptotycznie prawie na pewno:

$$(9.6) \quad p_l(n)n \left(1 - \frac{\delta(n)}{\ln n}\right) \leq \mathbb{N}_l \leq p_l(n)n \left(1 + \frac{\delta(n)}{\ln n}\right)$$

dla każdego $1 \leq l \leq M$. Niech $B_{\delta(n)}$ będzie zbiorem wektorów $\bar{\mathbb{N}}$ takich, że spełniona jest zależność (9.5).

Analogicznie jak w (9.3) możemy zdefiniować Y jako zmienną losową liczącą wierzchołki izolowane w $G_s(\mathcal{V}, m, \mathcal{P}_{(m)})$

$$Y = \sum_{i=1}^n \mathbb{J}_i \quad , \text{gdzie} \quad \mathbb{J}_i = \begin{cases} 1 & \text{jeśli } v_i \text{ jest izolowany w } G_s(\mathcal{V}, m, \mathcal{P}_{(m)}) \\ 0 & \text{jeśli } v_i \text{ nie jest izolowany w } G_s(\mathcal{V}, m, \mathcal{P}_{(m)}) \end{cases}$$

Z (9.1) wynika, że

$$\mathbb{E}(Y(\bar{d}))_k = \mathbb{E}((Y)_k \mid (|D(v_1)|, \dots, |D(v_n)|) = \bar{d}).$$

Zatem z (9.6), (9.4) i lematu 9.2 otrzymujemy, że

$$\begin{aligned}
\mathbb{E}(Y)_k &= \sum_{\bar{d}} \mathbb{E}(Y(\bar{d}))_k \Pr\{(Z_1, \dots, Z_n) = \bar{d}\} = \\
&= \sum_{\bar{N}} \left(\sum_{\bar{d} \leq \bar{N}} \mathbb{E}(Y(\bar{d}))_k \Pr\{(Z_1, \dots, Z_n) = \bar{d}\} \right) = \\
&= \sum_{\bar{N} \in B_{\delta(n)}} \mathbb{E}(Y(\bar{d}_{\bar{N}}))_k \Pr\{\bar{N} = \bar{N}\} + \sum_{\bar{N} \notin B_{\delta(n)}} \mathbb{E}(Y(\bar{d}_{\bar{N}}))_k \Pr\{\bar{N} = \bar{N}\} = \\
&= \exp(o(1))(A(\bar{p}, \bar{\beta}))^k + o(1).
\end{aligned}$$

Ostatecznie więc

$$(9.7) \quad \mathbb{E}(Y)_k \sim (A(\bar{p}, \bar{\beta}))^k.$$

Niech

$$\frac{n(d_n)_s^2}{s! m^s} = \ln n + c_n \quad i \quad c_n \rightarrow c$$

i

$$\frac{(Z_n)_s - (d_n)_s}{(d_n)_s \omega(n)} = \mathbb{A}_n$$

dla $\omega(n) > 0$ i \mathbb{A}_n dążącej według rozkładu do \mathbb{A} .

Założmy, że dla każdego n zmienne \mathbb{A}_n i \mathbb{A} mają wartości w zbiorach skończonych $\{\alpha_1(n), \dots, \alpha_M(n)\} \in \mathbb{R}$ i $\{\alpha'_1, \dots, \alpha'_M\} \in \mathbb{R}$, odpowiednio (czyli $\Pr\{\mathbb{A} = \alpha'_l\} > 0$ dla każdego $1 \leq l \leq M$). Niech $(\beta_l(n))_s = (d_n)_s (\alpha_l \omega(n) + 1)$ dla każdego $1 \leq l \leq M$. W związku z tym ze zbieżności według rozkładu (jeśli ustalimy indeksy tak, że $\alpha_l(n) < \alpha_{l+1}(n)$ i $\alpha'_l < \alpha'_{l+1}$ dla każdego l) mamy $p_l(n) \rightarrow \Pr\{\mathbb{A} = \alpha'_l\}$, $(p_l(n))$ dla dużych n są większe od pewnej stałej dodatniej, $\alpha_l(n) \rightarrow \alpha'_l$, $\mathbb{E}\mathbb{A}_n \rightarrow \mathbb{E}\mathbb{A}$ i $\mathbb{E} \exp(-\mathbb{A}_n) \rightarrow \mathbb{E} \exp(-\mathbb{A})$.

Założmy, że $\omega(n) = \frac{C_n}{\ln n}$ oraz $C_n = O(\ln n)$. Skoro $\frac{n(d_n)_s^2}{s! m^s} = \ln n + c_n$, $(Z_n)_s = (d_n)_s (\mathbb{A}_n \omega(n) + 1)$, więc także $(\Delta)_s^2 = O((d_n)_s^2) = O(\frac{\ln n m^s}{n})$. Czyli założenia lematu 9.2 są spełnione a wyrażenie z jego tezy jest postaci

$$A(\bar{p}(n), \bar{\beta}(n)) = \sum_{l_1=1}^M p_{l_1}(n) \exp(B(l_1)),$$

gdzie

$$\begin{aligned}
(9.8) \quad B(l_1) &= - \sum_{l_2=1}^M \left(p_{l_2}(n) \frac{n(\beta_{l_1})_s (\beta_{l_2})_s}{s! m^s} \right) + \ln n = \\
&= - \sum_{l_2=1}^M \left(p_{l_2}(n) \frac{n(d_n)_s^2}{s! m^s} (\alpha_{l_1} \omega(n) + 1) (\alpha_{l_2} \omega(n) + 1) \right) + \ln n = \\
&= - \frac{n(d_n)_s^2}{s! m^s} (\alpha_{l_1} \omega(n) + 1) \sum_{l_2=1}^M (p_{l_2}(n) (\alpha_{l_2} \omega(n) + 1)) + \ln n = \\
&= - (\ln n + c_n) (\omega(n) \alpha_{l_1} + 1) (\omega(n) \mathbb{E}(\mathbb{A}_n) + 1) + \ln n = \\
&= - (\ln n + c_n) \left(\frac{C_n \alpha_{l_1}}{\ln n} + 1 \right) \left(\frac{C_n \mathbb{E}(\mathbb{A}_n)}{\ln n} + 1 \right) + \ln n = \\
&= - \frac{C_n^2 \alpha_{l_1} \mathbb{E}(\mathbb{A}_n)}{\ln n} - C_n \alpha_{l_1} - C_n \mathbb{E}(\mathbb{A}_n) \\
&\quad - \frac{c_n C_n^2 \alpha_{l_1} \mathbb{E}(\mathbb{A}_n)}{\ln^2 n} - \frac{c_n C_n \alpha_{l_1}}{\ln n} - \frac{c_n C_n \mathbb{E}(\mathbb{A}_n)}{\ln n} - c_n.
\end{aligned}$$

Punkty (i), (ii) i (iii) twierdzenia odpowiadają trzem przypadkom:

(i) $C_n = o(1)$

$$B(l_1) = o(1) - c_n,$$

czyli

$$A(\overline{p(n)}, \overline{\beta(n)}) = \sum_{l_1=1}^M p_{l_1}(n) \exp(o(1) - c_n) = (1 + o(1)) \exp(-c_n);$$

(ii) $C_n = 1$

$$B(l_1) = o(1) - \alpha_{l_1} - \mathbb{E}\mathbb{A}_n - c_n,$$

czyli

$$\begin{aligned}
A(\overline{p(n)}, \overline{\beta(n)}) &= \sum_{l_1=1}^M p_{l_1}(n) \exp(o(1) - \alpha_{l_1} - \mathbb{E}\mathbb{A}_n - c_n) = \\
&= (1 + o(1)) \exp(-c_n - \mathbb{E}\mathbb{A}_n) \mathbb{E}(\exp(-\mathbb{A}_n));
\end{aligned}$$

(iii) $C_n \rightarrow \infty$, $\mathbb{E}\mathbb{A} = 0$ i \mathbb{A} ma co najmniej dwie wartości dla każdego $\alpha_l < 0$

$$B(l_1) \rightarrow \infty$$

a dla każdego $\alpha_l > 0$

$$B(l_1) \rightarrow -\infty,$$

więc

$$A(\overline{p(n)}, \overline{\beta(n)}) \rightarrow \infty.$$

Zatem twierdzenie 9.1(i) i (ii) wynika z powyższych wyliczeń, (9.7) i metody momentów (twierdzenie 3.2). Natomiast twierdzenie 9.1(iii) jest wnioskiem z (9.7), dla $k = 1$ i 2 , oraz z metody drugiego momentu (lemat 3.5).

9.4. Wnioski z twierdzenia o liczbie wierzchołków izolowanych

9.4.1. Jednostajny losowy graf przecięć

Wniosek 9.1. *Jeżeli*

$$(9.9) \quad \frac{n(d_n)_s^2}{s!m^s} = \ln n + c_n \quad i \quad c_n \rightarrow c,$$

to liczba wierzchołków izolowanych w $G_s^{Bin}(\mathcal{V}, m, p)$ dąży według rozkładu do zmiennej losowej o rozkładzie Poissona $Po(e^{-c})$.

Dowód. Wystarczy podstawić $\Pr\{\mathbb{A}_n = 0\} = 1$ oraz ciąg $\omega(n)$, dla którego $\omega(n) = o(1/\ln n)$. Wtedy wniosek wynika bezpośrednio z twierdzenia 9.1. \square

9.4.2. Dwumianowy losowy graf przecięć

Wniosek 9.2. *Jeżeli $m = n^\delta$ dla $\delta > 1/s$ oraz*

$$(9.10) \quad \frac{n(\lfloor mp \rfloor)_s^2}{s!m^s} = \ln n + c_n \quad i \quad c_n \rightarrow c,$$

to liczba wierzchołków izolowanych w grafie $G_s^{Bin}(\mathcal{V}, m, p)$ dąży według rozkładu do zmiennej losowej o rozkładzie Poissona $Po(e^{-c})$.

Dowód. Z (9.10)

$$mp \sim \sqrt[2s]{\frac{s!m^s \ln n}{n}},$$

czyli, dla $m = n^\delta$ i $\delta > 1/s$, zarówno d_+ jak i d_- zdefiniowane jak w lemacie 4.3 spełniają (9.9). Zatem wniosek wynika z lematu 4.3. \square

9.4.3. Nieskończona liczba typów

Wniosek 9.3. *Niech*

$$\frac{n(d_n)_s^2}{s!m^s} = \ln n + c_n \quad i \quad c_n \rightarrow c$$

dla stałej c i

$$\frac{(Z_n)_s - (d_n)_s}{(d_n)_s \omega(n)} = \mathbb{A}_n,$$

gdzie zmienna losowa \mathbb{A}_n dąży według rozkładu do \mathbb{A} ,

$$\mathbb{E}\mathbb{A}_n \rightarrow \mathbb{E}\mathbb{A}, \quad \mathbb{E} \exp(-\mathbb{A}_n) \rightarrow \mathbb{E} \exp(-\mathbb{A})$$

i \mathbb{A} ma rozkład prawdopodobieństwa o ograniczonym nośniku (ang. *bounded support*).

(i) *Jeśli $\omega(n) = o(\frac{1}{\ln n})$, to liczba wierzchołków izolowanych w $G_s(\mathcal{V}, m, \mathcal{P}_{(m)})$ dąży do rozkładu Poissona $Po(\lambda)$ z parametrem $\lambda = e^{-c}$.*

(ii) *Jeśli $\omega(n) = \frac{1}{\ln n}$, to liczba wierzchołków izolowanych w $G_s(\mathcal{V}, m, \mathcal{P}_{(m)})$ dąży według rozkładu do rozkładu Poissona $Po(\lambda)$ z $\lambda = e^{-c - \mathbb{E}\mathbb{A}} \mathbb{E}(e^{-\mathbb{A}})$.*

Dowód. Niech

$$\frac{n(d_n)_s^2}{s! m^s} = \ln n + c_n \quad \text{oraz} \quad c_n \rightarrow c$$

dla stałej c i

$$\frac{(Z_n)_s - (d_n)_s}{(d_n)_s \omega(n)} = \mathbb{A}_n.$$

Dodatkowo załóżmy, że

$$\omega(n) = \frac{C_n}{\ln n}, \quad \mathbb{A}_n \xrightarrow{D} \mathbb{A}, \quad \mathbb{E}\mathbb{A}_n \rightarrow \mathbb{E}\mathbb{A}, \quad \mathbb{E}\exp(-\mathbb{A}_n) \rightarrow \mathbb{E}\exp(-\mathbb{A})$$

i \mathbb{A} na rozkład $\tilde{\mathcal{P}}$, gdzie $\tilde{\mathcal{P}}$ jest rozkładem o ograniczonym nośniku. Ustalmy $\varepsilon > 0$ i M . Wybierzmy

$$\Delta_- = y_0 < y_1 < \dots < y_M = \Delta_+$$

takie, że

$$\tilde{\mathcal{P}}((\Delta_-; \Delta_+)) = 1, \quad \tilde{\mathcal{P}}(\{y_i\}) = 0 \quad \text{dla każdego } 1 \leq i \leq M$$

i

$$\varepsilon/2 < |y_{i+1} - y_i| < \varepsilon \quad \text{dla każdego } 1 \leq i \leq M-1.$$

Można zdefiniować dwie rodziny rozkładów $\tilde{\mathcal{P}}_n^-, \tilde{\mathcal{P}}^-$ i $\tilde{\mathcal{P}}_n^+, \tilde{\mathcal{P}}^+$, dla których:

$$\tilde{\mathcal{P}}_n^-(\{y_i\}) = \Pr\{\mathbb{A}_n \in (y_i; y_{i+1}]\}, \quad \tilde{\mathcal{P}}^-(\{y_i\}) = \Pr\{\mathbb{A} \in (y_i; y_{i+1}]\},$$

i

$$\tilde{\mathcal{P}}_n^+(\{y_{i+1}\}) = \Pr\{\mathbb{A}_n \in (y_i; y_{i+1}]\}, \quad \tilde{\mathcal{P}}^+(\{y_{i+1}\}) = \Pr\{\mathbb{A} \in (y_i; y_{i+1}]\},$$

dla każdego $0 \leq i \leq M-1$. Zdefiniujmy $\mathbb{A}_n^-, \mathbb{A}^-, \mathbb{A}_n^+$ oraz \mathbb{A}^+ jako zmienne losowe z rozkładami $\tilde{\mathcal{P}}_n^-, \tilde{\mathcal{P}}^-, \tilde{\mathcal{P}}_n^+$ i $\tilde{\mathcal{P}}^+$, odpowiednio. Można zatem skonstruować dwa losowe grafy przecięć $G_s(\mathcal{V}, m, \mathcal{P}_{(m)}^-)$ i $G_s(\mathcal{V}, m, \mathcal{P}_{(m)}^+)$, dla których, podobnie jak to zostało zrobione w dowodzie lematu 4.1, można skonstruować coupling, w którym

$$G_s(\mathcal{V}, m, \mathcal{P}_{(m)}^-) \subseteq G_s(\mathcal{V}, m, \mathcal{P}_{(m)}) \subseteq G_s(\mathcal{V}, m, \mathcal{P}_{(m)}^+).$$

Oznaczmy przez Y_n^-, Y_n, Y_n^+ , zmienne losowe określające liczbę wierzchołków izolowanych w $G_s(\mathcal{V}, m, \mathcal{P}_{(m)}^-)$, $G_s(\mathcal{V}, m, \mathcal{P}_{(m)})$, $G_s(\mathcal{V}, m, \mathcal{P}_{(m)}^+)$, odpowiednio. Z couplingu i z tego, że posiadanie co najwyżej x wierzchołków izolowanych jest własnością malejącą, wynika, że

$$(9.11) \quad \Pr\{Y_n^+ \leq x\} \leq \Pr\{Y_n \leq x\} \leq \Pr\{Y_n^- \leq x\}$$

dla każdego n oraz $x \in \mathbb{R}$. Mamy do rozpatrzenia dwa przypadki:

(i) $C_n = o(1)$

Z twierdzenia 9.1 Y_n^+ i Y_n^- dążą według rozkładu do rozkładu Poissona z parametrem $\lambda = e^{-c}$, czyli z (9.11) Y_n dąży według rozkładu do rozkładu Poissona z parametrem $\lambda = e^{-c}$.

(ii) $C_n = 1$

Oznaczmy:

$$\lambda^- = e^{-c-\mathbb{E}A^-} \mathbb{E}(e^{-A^-}), \lambda = e^{-c-\mathbb{E}A} \mathbb{E}(e^{-A}), \lambda^+ = e^{-c-\mathbb{E}A^+} \mathbb{E}(e^{-A^+}).$$

Zauważmy, że dla dowolnego n :

$$\mathbb{E}A_n^+ - \varepsilon \leq \mathbb{E}A_n \leq \mathbb{E}A_n^+,$$

$$\mathbb{E} \exp(-A_n^+) \leq \mathbb{E} \exp(-A_n) \leq \exp(\varepsilon) \mathbb{E} \exp(-A_n^+)$$

i

$$\mathbb{E}A_n^- \leq \mathbb{E}A_n \leq \mathbb{E}A_n^- + \varepsilon,$$

$$\exp(-\varepsilon) \mathbb{E} \exp(-A_n^-) \leq \mathbb{E} \exp(-A_n) \leq \mathbb{E} \exp(-A_n^-).$$

Stąd ze zbieżności wartości oczekiwanych:

$$\lambda^+ \leq \lambda \leq \exp(2\varepsilon)\lambda^+,$$

$$\exp(-2\varepsilon)\lambda^- \leq \lambda \leq \lambda^-,$$

oraz

$$e^{-2\varepsilon}\lambda \leq \lambda^+, \quad e^{-\lambda} \leq e^{-\lambda^+},$$

$$\lambda^- \leq e^{2\varepsilon}\lambda, \quad e^{-\lambda^-} \leq e^{-\lambda}.$$

Z twierdzenia 9.1 Y_n^+ i Y_n^- dążą według rozkładów do zmiennych losowych o rozkładzie Poissona z parametrami λ^+ and λ^- , odpowiednio. Zatem z (9.11), dla każdego $k \in \mathbb{Z}$ i $x \in (k; k+1)$, mając dane $\varepsilon > 0$, dla dostatecznie dużego n mamy

$$\begin{aligned} e^{-2k\varepsilon} \sum_{i=0}^k \frac{(\lambda)^i}{i!} e^{-\lambda} - \varepsilon &\leq \sum_{i=0}^k \frac{(\lambda^+)^i}{i!} e^{-\lambda^+} - \varepsilon \leq \\ &\leq \Pr\{Y_n^+ \leq x\} \leq \Pr\{Y_n \leq x\} \leq \Pr\{Y_n^- \leq x\} \leq \\ &\leq \sum_{i=0}^k \frac{(\lambda^-)^i}{i!} e^{-\lambda^-} + \varepsilon \leq e^{2k\varepsilon} \sum_{i=0}^k \frac{(\lambda)^i}{i!} e^{-\lambda} + \varepsilon. \end{aligned}$$

Biorąc $\varepsilon \rightarrow 0$ otrzymujemy zbieżność dystrybuanty zmiennych losowych Y_n do dystrybuanty zmiennej losowej o rozkładzie Poissona $\text{Po}(\lambda)$. □

9.4.4. Skończona liczba typów

Twierdzenie 9.2. Niech d będzie stałą i $L \subseteq \mathbb{N}$ będzie skończonym zbiorem, takim, że $\min\{l : l \in L\} \geq s$ i $\max\{l : l \in L\} = d$. Załóżmy, że rozkład $\mathcal{P}_{(m)} = (P_0, P_1, \dots, P_m)$ jest taki, że $P_l(n) = 0$ dla wszystkich $l \notin L$ oraz $P_l(n) \sim n^{\frac{(l)_s}{(d)_s} - 1}$ dla wszystkich $l \in L$. Jeżeli

$$m = \sqrt[s]{\frac{n(d)_s^2}{s!(\ln n + c_n)}} \quad \text{dla } c_n \rightarrow c,$$

to liczba wierzchołków izolowanych w $G_s(\mathcal{V}, m, \mathcal{P}_{(m)})$ dąży według rozkładu do rozkładu Poissona $\text{Po}(\lambda)$ z parametrem $\lambda = \sum_{l \in L} e^{-c \frac{(l)_s}{(d)_s}}$.

Dowód. Podstawmy w dowodzie twierdzenia 9.1 $d_n = d$, $\alpha_l = \frac{(l)_s}{(d)_s} - 1$ oraz $\omega(n) = 1$. W tym przypadku $\Pr\{\mathbb{A}_n = \alpha_l\} = P_l(n)$, $\beta_l = l$ i $\mathbb{E}\mathbb{A}_n = \frac{\mathbb{E}(Z_n)_s}{(d)_s} - 1$. W tym przypadku założmy jednak, że $(P_l(n))_{n \in \mathbb{N}}$ dla $l \in L$ nie są ograniczone z dołu przez pewną liczbę dodatnią, lecz $P_l(n) \sim n^{\frac{(l)_s}{(d)_s} - 1}$. Na przykład można podstawić dla $L = \{s, s+1, \dots, d\}$

$$\begin{cases} P_s(n) = n^{\frac{(s)_s}{(d)_s} - 1}; \\ P_l(n) = n^{\frac{(l)_s}{(d)_s} - 1} - P_{l-1}(n), \text{ dla } s < l \leq d. \end{cases}$$

Mimo, że pominiemy założenia o ograniczoności z dołu, wszystkie przekształcenia z dowodu twierdzenia 9.1 możemy powtórzyć, gdyż z nierówności Chernoffa ((3.1) i (3.2)) asymptotycznie prawie na pewno będzie spełniona zależność (9.6), jeśli odpowiednio podstawimy $\delta(n)$.

Zauważmy teraz, że skoro $\sum_{l \in L} P_l(n) = 1$ oraz dla $l \neq d$ mamy $\frac{(l)_s}{(d)_s} - 1 < 0$, to $P_d(n) = 1 - O\left(\frac{1}{n^\varepsilon}\right)$, gdzie $\varepsilon = 1 - \frac{(d-1)_s}{(d)_s}$, zatem

$$\mathbb{E}(Z_n)_s = (d)_s - O\left(\frac{1}{n^\varepsilon}\right).$$

W dodatku mamy

$$\omega(n) = 1, \quad \alpha_l = \frac{(l)_s}{(d)_s} - 1 \quad \text{i} \quad \mathbb{E}\mathbb{A}_n = \frac{\mathbb{E}(Z_n)_s}{(d)_s} - 1.$$

Stąd z (9.8)

$$\begin{aligned} B(l) &= -(\ln n + c_n)(\alpha_l + 1)(\mathbb{E}(\mathbb{A}_n) + 1) + \ln n = \\ &= -(\ln n + c_n) \frac{(l)_s}{(d)_s} \cdot \frac{\mathbb{E}(Z_n)_s}{(d)_s} + \ln n = \\ &= \left(1 - \frac{(l)_s \mathbb{E}(Z_n)_s}{(d)_s^2}\right) \ln n - c_n \frac{(l)_s \mathbb{E}(Z_n)_s}{(d)_s^2} = \\ &= \left(1 - \frac{(l)_s}{(d)_s}\right) \ln n - c_n \frac{(l)_s}{(d)_s} + o(1). \end{aligned}$$

Ponieważ $P_l(n) \sim n^{\frac{(l)_s}{(d)_s} - 1} = \exp\left(\left(\frac{(l)_s}{(d)_s} - 1\right) \ln n\right)$, więc

$$\begin{aligned} A(\overline{p(n)}, \overline{\beta(n)}) &= \sum_{l \in L} P_l(n) \exp(B(l)) \\ &\sim \sum_{l \in L} \exp\left(-c_n \frac{(l)_s}{(d)_s}\right) \sim \sum_{l \in L} \exp\left(-c \frac{(l)_s}{(d)_s}\right). \end{aligned}$$

By otrzymać tezę dowodzonego twierdzenia wystarczy teraz powtórzyć rozumowanie z dowodu twierdzenia 9.1 i zastosować metodę momentów (twierdzenie 3.2). \square

10. Spójność i składowe spójności

Wykorzystamy teraz wyniki zgromadzone w rozdziale 8 i rozdziale 9 dotyczące wielkości składowych spójności i liczby wierzchołków izolowanych w $G^{Jed}(\mathcal{V}, m, d)$, w celu określenia funkcji progowych dla spójności i przejścia fazowego w $G^{Jed}(\mathcal{V}, m, d)$. Pokażemy też wyniki dotyczące przejścia fazowego w $G_s^{Jed}(\mathcal{V}, m, d)$ i $G_s^{Bin}(\mathcal{V}, m, p)$.

10.1. Spójność $G^{Jed}(\mathcal{V}, m, d)$

Twierdzenie 10.1. *Niech $\{b_n\}$ będzie ciągiem, $b > 0$ będzie stałą oraz $d = d(n) \geq 2$ będzie takie, że*

$$\frac{d^2 n}{m} = \ln n + b_n.$$

- (i) *Jeśli $b_n \rightarrow -\infty$ to asymptotycznie prawie na pewno $G^{Jed}(\mathcal{V}, m, d)$ jest niespójny.*
- (ii) *Jeśli $b_n \rightarrow b$ to asymptotycznie prawie na pewno $G^{Jed}(\mathcal{V}, m, d)$ składa się z jednej dużej składowej spójności i wierzchołków izolowanych poza nią. Liczba wierzchołków izolowanych dąży według rozkładu do zmiennej losowej o rozkładzie Poissona $Po(e^{-b})$.*
- (iii) *Jeśli $b_n \rightarrow +\infty$ to asymptotycznie prawie na pewno $G^{Jed}(\mathcal{V}, m, d)$ jest spójny.*

W dowodzie twierdzenia wykorzystamy lemat dotyczący liczby wierzchołków izolowanych w $G^{Jed}(\mathcal{V}, m, d)$.

Lemat 10.1. *Niech*

$$\frac{d^2 n}{m} = \ln n + b_n$$

i X_n będzie zmienną losową określającą liczbę wierzchołków izolowanych w grafie $G^{Jed}(\mathcal{V}, m, d)$.

- (i) *Jeśli $b_n \rightarrow -\infty$, to asymptotycznie prawie na pewno jest nieskończenie wiele wierzchołków izolowanych w $G^{Jed}(\mathcal{V}, m, d)$.*
- (ii) *Jeśli $b_n \rightarrow b$, dla pewnej stałej b , to X_n dąży według rozkładu do rozkładu Poissona $Po(e^{-b})$.*
- (iii) *Jeśli $b_n \rightarrow \infty$, to asymptotycznie prawie na pewno nie ma wierzchołków izolowanych w $G^{Jed}(\mathcal{V}, m, d)$.*

Lemat ten został udowodniony w [26], ale także jest szczególnym przypadkiem twierdzenia przedstawionego w rozdziale 9 (a w zasadzie wniosku 9.1 z tego twierdzenia). Punkty (i) i (iii) lematu nie wynikają bezpośrednio

z wniosku 9.1 z twierdzenia 9.1. Jednakże w dowodzie twierdzenia 9.1 została zastosowana metoda momentów, więc obliczenia w nim zawarte dają oszacowania na pierwszy i drugi moment dla przypadków (i) i (iii). Z tych oszacowań i metody pierwszego i drugiego momentu (lemat 3.5) wynika (i) i (iii).

Potrzebny będzie też poniższy prosty fakt.

Fakt 10.1. *Niech $d = 2$, $\frac{d^2 n}{m} - \ln n = b_n$, $b_n = \Omega(1)$, wtedy asymptotycznie prawie na pewno nie ma izolowanych krawędzi w $G^{Jed}(\mathcal{V}, m, d)$.*

Dowód. Dla $d = 2$ prawdopodobieństwo, że istnieje krawędź izolowana w grafie $G^{Jed}(\mathcal{V}, m, d)$ wynosi co najwyżej

$$\begin{aligned} & \binom{n}{2} \left(\frac{1}{\binom{m}{2}} \left(\frac{\binom{m-2}{2}}{\binom{m}{2}} \right)^{n-2} + \frac{2(m-2)}{\binom{m}{2}} \left(\frac{\binom{m-3}{2}}{\binom{m}{2}} \right)^{n-2} \right) \leq \\ & \leq \binom{n}{2} \left(\frac{1}{\binom{m}{2}} \left(1 - \frac{2}{m} \right)^{2(n-2)} + \frac{2(m-2)}{\binom{m}{2}} \left(1 - \frac{3}{m} \right)^{2(n-2)} \right) = \\ & = O \left(\frac{n^2}{m^2} \exp \left(-\frac{4n}{m} \right) \right) + O \left(\frac{n^2}{m} \exp \left(-\frac{6n}{m} \right) \right) = \\ & = O \left(\frac{n^2}{m^2} \exp(-\ln n - b_n) \right) + O \left(\frac{n^2}{m} \exp \left(-\frac{3}{2} \ln n - \frac{3}{2} b_n \right) \right) = \\ & = O \left(\frac{n}{m^2} \right) + O \left(\frac{\sqrt{n}}{m} \right) = O \left(\frac{(\ln n)^2}{n} \right) + O \left(\frac{\ln n}{\sqrt{n}} \right) = o(1). \end{aligned}$$

□

Dowód twierdzenia 10.1.

(i) Jest bezpośrednią konsekwencją lematu 10.1(i).

(ii) i (iii) Załóżmy, że $b_n = o(\ln n)$. Zauważmy, że z założeń twierdzenia wynika, że

$$d = \sqrt{\frac{m(\ln n + b_n)}{n}},$$

czyli

$$\frac{d^2 n}{m \ln n} = c_n \rightarrow c,$$

gdzie $c = 1$. Z lematu 8.10 asymptotycznie prawie na pewno $G^{Jed}(\mathcal{V}, m, d)$ składa się z jednej dużej składowej spójności i składowych o co najwyżej $\lfloor \frac{1}{C} \rfloor$ wierzchołkach. Z definicji C wartość $\lfloor \frac{1}{C} \rfloor$ wynosi

$$\lfloor \frac{1}{C} \rfloor = \begin{cases} 2 & \text{dla } d = 2; \\ 1 & \text{dla } d \neq 2. \end{cases}$$

W dodatku z faktu 10.1 dla $d = 2$ asymptotycznie prawie na pewno graf $G^{Jed}(\mathcal{V}, m, d)$ nie zawiera składowych o dwóch wierzchołkach. Czyli dla dowolnego $d \geq 2$ asymptotycznie prawie na pewno $G^{Jed}(\mathcal{V}, m, d)$ składa się z jednej składowej spójności i wierzchołków izolowanych poza tym. Zatem

rezultat dla $b_n = o(\ln n)$ wynika z lematu 10.1. Dla większych wartości b_n wystarczy wykorzystać prosty coupling podobny do tego przedstawionego w rozdziale 4. Z tego couplingu wynika, że dla $d \leq d'$ jeśli $G^{Jed}(\mathcal{V}, m, d)$ jest spójny asymptotycznie prawie na pewno, to $G^{Jed}(\mathcal{V}, m, d')$ jest spójny asymptotycznie prawie na pewno. \square

10.2. Przejście fazowe

W tym rozdziale wykorzystamy stosowaną standardowo przy określaniu funkcji progowej przejścia fazowego metodę przybliżania procesu BFS procesami gałązkowymi. Jak wspominają Bollobás, Janson i Riordan w [14], najprawdopodobniej została ona wykorzystana po raz pierwszy przez Karpa w [41]. Prosty opis tej metody można znaleźć w rozdziale 5 [32]. W tym podrozdziale będziemy opierać się na rozumowaniu przedstawionym w dowodzie twierdzenia 5.4 z [32], jednakże ze względu na specyfikę losowych grafów przecięć będziemy musieli to rozumowanie zmodyfikować.

W odróżnieniu do grafu $G(n, p)$, którego dotyczył dowód z [32], w losowym grafie przecięć występują zależności między powstającymi krawędziami. W związku z tym przybliżenie procesu BFS w losowym grafie przecięć procesem gałązkowym wymaga dodatkowej pracy. W rozdziale 8 pokazaliśmy już, że w $G^{Jed}(\mathcal{V}, m, d)$ wielkość drzewa BFS można przybliżyć przez liczbę potomków odpowiednich procesów gałązkowych. Natomiast analogiczny wynik dotyczący $G_s^{Bin}(\mathcal{V}, m, p)$ będzie stanowił główną część dowodu zaprezentowanego w podrozdziale 10.2.3.

W dowodach obu twierdzeń dotyczących przejścia fazowego będziemy wykorzystywać znane fakty dotyczące prawdopodobieństwa wymarcia procesu gałązkowego, dlatego przedstawimy je w oddzielnym podrozdziale.

10.2.1. Prawdopodobieństwo wymarcia procesu gałązkowego

Dla danych $P = P(n) \in [0; 1]$, $P' = P'(n) \in [0; 1]$, $N = N(n) \sim n$ takich, że $NP \rightarrow c$, przyjmijmy oznaczenie na proces gałązkowy $\mathcal{B}(P', P, N)$ jak w podrozdziale 8.3. Oznaczmy także przez $\mathcal{B}_P(c)$ proces gałązkowy, w którym liczba dzieci każdego osobnika jest niezależną zmienną losową o rozkładzie Poissona $Po(c)$.

Oznaczmy przez $\rho(c)$ i $\rho'(P', P, N)$ prawdopodobieństwo wymarcia, odpowiednio, procesów gałązkowych $\mathcal{B}_P(c)$ i $\mathcal{B}(P', P, N)$ (tzn. prawdopodobieństwo, że procesy te składają się tylko ze skończonej liczby potomków).

Poniższy lemat, który wynika z rozważań przedstawionych w podrozdziale 5.2 w [32], jest prostym wnioskiem z klasycznego twierdzenia o prawdopodobieństwie wymarcia procesu gałązkowego i punktowej zbieżności funkcji tworzącej zmiennej losowej o rozkładzie dwumianowym $\text{Bin}(N, P)$ do funkcji tworzącej zmiennej losowej o rozkładzie Poissona $Po(c)$ dla $NP \rightarrow c$.

Lemat 10.2. *Jeśli $NP \rightarrow c$ i $c > 1$, to*

$$\rho'(P, P, N) \rightarrow \rho(c), \quad \text{przy } n \rightarrow \infty,$$

gdzie $\rho(c)$ jest jedynym pierwiastkiem równania

$$(10.1) \quad \rho(c) = e^{(1-\rho(c))c}$$

w przedziale $(0; 1)$.

Proces $\mathcal{B}(P', P, N)$ różni się nieznacznie od procesu $\mathcal{B}(P, P, N)$. Jednakże teoretycznie, jeśli P' jest większe od P , różnica ta może wpłynąć na prawdopodobieństwo wymarcia procesu. Dlatego podajemy poniższy prosty fakt.

Fakt 10.2. *Jeśli $NP \rightarrow c$, $NP' \rightarrow C'$, $c > 1$ oraz*

$$C' = \begin{cases} \frac{d-1}{d}c, & \text{dla } d = O(1); \\ c, & \text{dla } d \rightarrow \infty, \end{cases}$$

to

$$\rho'(P', P, N) \rightarrow \rho'(c, d),$$

gdzie

$$(10.2) \quad \rho'(c, d) = \begin{cases} e^{-\frac{d-1}{d}c(1-\rho(c))}, & \text{dla } d = O(1); \\ \rho(c), & \text{dla } d \rightarrow \infty \end{cases}$$

oraz $\rho(c)$ jest jedynym pierwiastkiem równania (10.1) w przedziale $(0; 1)$.

Dowód. Niech X_1 będzie zmienną losową liczącą liczbę potomków dodanych do procesu w pierwszym kroku procesu $\mathcal{B}(P', P, N)$. Pod warunkiem, że $X_1 = i$, $0 \leq i \leq N$, prawdopodobieństwo wymarcia procesu $\mathcal{B}(P', P, N)$ jest równe prawdopodobieństwu wymarcia i niezależnych procesów $\mathcal{B}(P, P, N)$. Zatem

$$\begin{aligned} \rho'(P', P, N) &= \sum_{i=0}^N (P')^i (1 - P')^{N-i} (\rho'(P, P, N))^i = \\ &= (1 - P'(1 - \rho'(P, P, N)))^N = e^{-C'(1-\rho'(P,P,N))+o(1)}. \end{aligned}$$

czyli teza wynika z lematu 10.2 i faktu, że $\rho(c)$ spełnia równanie (10.1). \square

10.2.2. Przejście fazowe w $G^{Jed}(\mathcal{V}, m, d)$

Twierdzenie 10.2. *Niech $c > 0$ będzie stałą, $d = d(n) \geq 2$ i*

$$\frac{d(d-1)n}{m} \sim c.$$

- (i) *Jeśli $c < 1$, to asymptotycznie prawie na pewno największa składowa spójności grafu $G^{Jed}(\mathcal{V}, m, d)$ ma wielkość $O(\ln n)$.*
- (ii) *Jeśli $c > 1$, to asymptotycznie prawie na pewno graf $G^{Jed}(\mathcal{V}, m, d)$ składa się z dokładnie jednej składowej spójności zawierającej $\Omega(n)$ wierzchołków i składowych spójności wielkości $O(\ln n)$.*

Dowód twierdzenia 10.2.

W dowodzie wartości N_+, P_+, P'_+ i N_-, P_-, P'_- będą dane wzorami (8.10) oraz (8.11).

(i) Zauważmy, że ograniczenie liczby wierzchołków drzewa BFS od góry przez liczbę osobników procesu gałązkowego $\mathcal{B}_+ = \mathcal{B}_+(P'_+, P_+, N)$ przedstawione w podrozdziale 8.4 jest prawdziwe także dla $c < 1$. Dlatego analogiczny dowód jak dowód twierdzenia 5.4(i) z [32] da tezę. Niech X_1, \dots, X_k będą niezależnymi zmiennymi losowymi takimi, że X_1 ma rozkład dwumianowy $\text{Bin}(N, P'_+)$ a X_2, \dots, X_k mają rozkład dwumianowy $\text{Bin}(N, P_+)$. Skoro dla dużych n zachodzi $P'_+ \leq P_+ + (1 - P_+)P_+ + (1 - P_+)^2 P_+$, to dla dużych n $X_1 \prec X'_1 + X''_1 + X'''_1$ (X_1 jest stochastycznie zdominowane przez $X'_1 + X''_1 + X'''_1$), gdzie X'_1, X''_1, X'''_1 są niezależnymi zmiennymi losowymi o rozkładzie dwumianowym $\text{Bin}(N, P_+)$. W związku z tym dla dostatecznie dużych n , prawdopodobieństwo, że $X'_1 + X''_1 + X'''_1 + \sum_{i=2}^k X_i \leq k - 1$ wynosi co najwyżej tyle co prawdopodobieństwo, że $\sum_{i=1}^k X_i \leq k - 1$. W dodatku rozumowanie analogiczne do tego zastosowanego w uwadze 8.1 doprowadzi nas do wniosku, że prawdopodobieństwo, że $\sum_{i=1}^k X_i \leq k - 1$ wynosi co najwyżej tyle co prawdopodobieństwo, że procedura BFS zakończy się w co najwyżej k krokach. Jako suma niezależnych zmiennych losowych o rozkładzie dwumianowym, zmienna losowa $X = X'_1 + X''_1 + X'''_1 + \sum_{i=2}^k X_i$ ma rozkład dwumianowy $\text{Bin}((k + 2)N, P_+)$. Podstawmy $k = k(n) = 3 \ln n / (1 - c)^2$. Skoro $NP_+ = C_N \rightarrow c < 1$, więc z nierówności Chernoffa (3.2), jeśli n jest dostatecznie duże, żeby $C_N < 1$, to

$$\begin{aligned} & \Pr\{w G^{Jed}(\mathcal{V}, m, d) \text{ istnieje składowa o więcej niż } k \text{ wierzchołkach}\} \leq \\ & \leq n \Pr\{\text{BFS zacząty w } v \text{ nie zakończy się w co najwyżej } k \text{ krokach}\} \leq \\ & \leq n \Pr\{X > k - 1\} = \\ & = n \Pr\{X > (k + 2)C_N + (1 - C_N)k - 1 - 2C_N\} \leq \\ & \leq n \exp\left(-\frac{((1 - C_N)k - 1 - 2C_N)^2}{2\left((k + 2)C_N + \frac{(1 - C_N)k - 1 - 2C_N}{3}\right)}\right) = o(1). \end{aligned}$$

Ostatnie przekształcenie wynika z faktu, że $c < 1$, $k = k(n) \rightarrow \infty$ oraz

$$\frac{((1 - C_N)k - 1 - 2C_N)^2}{2\left((k + 2)C_N + \frac{(1 - C_N)k - 1 - 2C_N}{3}\right)} \sim \frac{(1 - c)^2 k^2}{2\left(\frac{1 + 2c}{3}k\right)} = \frac{3 \ln n}{\frac{2(1 + 2c)}{3}} = \frac{9}{2(1 + 2c)} \ln n.$$

(ii) Z lematu 8.12 wynika, że asymptotycznie prawie na pewno $G^{Jed}(\mathcal{V}, m, d)$ zawiera tylko składowe o co najwyżej $k_- = 16c \ln n / ((\sqrt{c} - 1)^2 (c - 1))$ wierzchołkach i dokładnie jedną składową o co najmniej $k_+ = \sqrt{n \ln n}$ wierzchołkach. Niech Y będzie zmienną losową określającą liczbę wierzchołków w składowych o co najwyżej k_- wierzchołkach. Wtedy $Y = Y_n = \sum_{v \in \mathcal{V}} Y_n(v)$, gdzie

$$Y_n(v) = \begin{cases} 1, & \text{jeśli } v \text{ jest w składowej o co najwyżej } k_- \text{ wierzchołkach;} \\ 0, & \text{w przeciwnym przypadku.} \end{cases}$$

Z rozważań zawartych w podrozdziale 8.4 wynika, że wartość

$$\Pr \{Y_n(v) = 1\}$$

jest ograniczona z góry przez $\rho'(P'_+, P_+, N_+)$ natomiast z dołu jest ograniczona przez $\rho'(P'_-, P_-, N_-) + o(1)$, gdzie składnik $o(1)$ wynika z prawdopodobieństwa wymarcia procesu po k_- krokach. Stąd z faktu 10.2

$$\mathbb{E}Y \sim \rho'(c, d)n,$$

gdzie $\rho'(c, d)$ jest dane wzorem (10.2). Natomiast, jeśli v należy do składowej o co najwyżej k_- wierzchołkach, to v' należy do takiej składowej, jeśli jest zawarty w składowej wierzchołka v lub proces BFS rozpoczęty w v' nie połączył się z procesem BFS rozpoczętym w v i wygasł w co najwyżej k_- krokach. Stąd z faktu 10.2

$$\begin{aligned} \mathbb{E}Y(Y-1) &= \sum_{v \in \mathcal{V}} \sum_{v' \in \mathcal{V}, v' \neq v} \Pr \{Y_n(v) = 1, Y_n(v') = 1\} \leq \\ &\leq n\rho'(P'_+, P_+, N_+)(k_- + n\rho'(P'_+, P_+, N_+ - k_-)) \sim \\ &\sim (\rho'(c, d)n)^2. \end{aligned}$$

Zatem z metody drugiego momentu (lemat 3.5) dla dowolnego $\varepsilon > 0$ asymptotycznie prawie na pewno liczba wierzchołków w składowych o co najwyżej k_- wierzchołkach wynosi co najwyżej $(\rho'(c, d) + \varepsilon)n$. Czyli wierzchołków w największej składowej jest $\Omega(n)$. \square

10.2.3. Przejście fazowe w $G_s^{Jed}(\mathcal{V}, m, d)$ i $G_s^{Bin}(\mathcal{V}, m, p)$

Twierdzenie 10.3. *Niech $c > 0$ i $s \geq 2$. Niech N_s będzie wielkością największej składowej spójności w $G_s^{Jed}(\mathcal{V}, m, d)$. Załóżmy, że*

$$d^{2s} \sim cm^s s! / n$$

oraz

$$n = o(m^{s/(2s-1)}).$$

Wtedy z prawdopodobieństwem dążącym do 1 przy $n \rightarrow \infty$

- (i) $N_s \leq \frac{9}{(1-c)^2} \ln n$, dla $c < 1$;
- (ii) $N_s = (1 + o(1))(1 - \rho)n$, dla $c > 1$.

Gdzie $\rho = \rho(c)$ jest dane wzorem (10.1).

Powyższe twierdzenie jest prostym wnioskiem z lematu 4.2 oraz poniższego twierdzenia dotyczącego grafu $G_s^{Bin}(\mathcal{V}, m, p)$.

Twierdzenie 10.4. *Niech $c > 0$ i $s \geq 2$ oraz niech N'_s oznacza wielkość największej składowej spójności w $G_s^{Bin}(\mathcal{V}, m, p)$. Załóżmy, że*

$$p^{2s} \sim cs! / m^s n$$

oraz

$$n = o(m^{s/(2s-1)}).$$

Wtedy z prawdopodobieństwem dążącym do 1 przy $n \rightarrow \infty$,

- (i) $N_s \leq \frac{9}{(1-c)^2} \ln n$, dla $c < 1$;
(ii) $N_s = (1 + o(1))(1 - \rho)n$, dla $c > 1$.
Gdzie $\rho = \rho(c)$ dane jest wzorem (10.1).

Dowód. Jak w rozdziale 8 będziemy porównywać procedurę BFS zaczynającą się w wierzchołku v_1 do odpowiedniego procesu gałązkowego. Zauważmy, że przy założeniu, że $|D(v_1)| = k$, liczba sąsiadów wierzchołka v_1 jest zmienną losową o rozkładzie dwumianowym $\text{Bin}(n-1, p_1)$, gdzie $p_1 = p(k)$ oraz dla danego k wartość $p(k)$ jest dana wzorem

$$(10.3) \quad p(k) = \sum_{j \geq 0} \binom{k}{s+j} p^{s+j} (1-p)^{k-s-j}.$$

Założmy, podobnie jak poprzednio, że jesteśmy po $t-1$ ($t \geq 1$) krokach procedury BFS i nie wykonaliśmy jeszcze żadnego podkroku kroku t . W takim razie do kroku $t-1$ do drzewa BFS zostały dodane kolejno wierzchołki $\{v_1, \dots, v_{t-1}, v_t, \dots, v_{t+s}\}$. Oznaczmy uporządkowany zbiór tych wierzchołków przez $V_{BFS}^{t-1}(v) = V_{BFS}^{t-1}$. Oznaczmy także przez $D_{[t-1]} = D(\{v_1, \dots, v_{t-1}\})$ dla $t \geq 2$ oraz $D_{[0]} = \emptyset$. Niech $D^*(v_t) = D(v_t) \setminus D_{[t-1]}$ oraz $D'(v_t) = D(v_t) \cap D_{[t-1]}$. Wprowadźmy zmienne losowe $Z_{[t]} = |D_{[t]}|$, $Z_t^* = |D^*(v_t)|$ i $Z_t' = |D'(v_t)|$. Dla $u \in \mathcal{V} \setminus V_{BFS}^{t-1}$, niech $\mathbb{I}(u)$ będzie zmienną indykatorową zdarzenia $\{|D(u) \cap D(v_t)| \geq s\}$. Wtedy $\sum_{u \in \mathcal{V} \setminus V_{BFS}^{t-1}} \mathbb{I}(u)$ jest zmienną losową liczącą wierzchołki dodane do drzewa BFS w t -tym kroku. Zauważmy, że pod warunkiem zdarzenia

$$\mathbb{D}_t := (V_{BFS}^{t-1}; D(v_1), \dots, D(v_t)),$$

zmienne losowe $\mathbb{I}(u)$, $u \in \mathcal{V} \setminus V_{BFS}^{t-1}$ są niezależne. W dodatku, pod warunkiem \mathbb{D}_t , zmienna X_t oznaczająca liczbę wierzchołków dodanych do drzewa BFS w kroku t ma rozkład dwumianowy $\text{Bin}(n_t, p_t)$, gdzie $n_t = n - |V_{BFS}^{t-1}|$ i

$$p_t = \Pr\left\{|D(u) \cap D(v_t)| \geq s \mid \mathbb{D}_t\right\},$$

dla $u \in \mathcal{V} \setminus V_{BFS}^{t-1}$.

Zajmiemy się teraz oszacowaniem wielkości p_t . Pierwszym istotnym spostrzeżeniem jest to, że każda własność jest dodawana niezależnie do zbioru własności wierzchołka. Zatem, pod warunkiem zdarzenia \mathbb{D}_t , zmienne losowe $\eta = |D(u) \cap D^*(v_t)|$ i $\xi = |D(u) \cap D'(v_t)|$ są niezależne. W dodatku z definicji tych zmiennych losowych zachodzi

$$p_t = \Pr\{\eta + \xi \geq s \mid \mathbb{D}_t\}.$$

Oczywiście z tego wynika, że

$$(10.4) \quad p(k^*) = \Pr\{\eta \geq s \mid \mathbb{D}_t\} \leq \Pr\{\eta + \xi \geq s \mid \mathbb{D}_t\} = p_t,$$

gdzie $k^* = Z_t^* = |D^*(v_t)|$ pod warunkiem zajścia \mathbb{D}_t .

Z drugiej strony, jeśli podstawimy $\xi_l := |D(v_t) \cap D(u) \cap D(v_l)|$ oraz wprowadzimy zdarzenie $\mathbb{D}_t^* = \{\xi_l \leq s-1, 1 \leq l < t\}$, dla $a = 0, 1, \dots, s-1$ zachodzi

$$\begin{aligned}
\Pr\left\{\eta + \xi = a \mid \mathbb{D}_t\right\} &= \Pr\left\{\eta + \xi = a \mid \mathbb{D}_t^*, D(v_1), \dots, D(v_t)\right\} \\
&= \frac{\Pr\left\{\eta + \xi = a, \mathbb{D}_t^* \mid D(v_1), \dots, D(v_t)\right\}}{\Pr\left\{\mathbb{D}_t^* \mid D(v_1), \dots, D(v_t)\right\}} \\
&= \frac{\Pr\left\{\eta + \xi = a \mid D(v_1), \dots, D(v_t)\right\}}{\Pr\left\{\mathbb{D}_t^* \mid D(v_1), \dots, D(v_t)\right\}} \\
&\geq \Pr\left\{\eta + \xi = a \mid D(v_1), \dots, D(v_t)\right\} \\
&= \Pr\left\{|D(u) \cap D(v_t)| = a \mid |D(v_t)| = Z_t\right\}.
\end{aligned}$$

Zauważmy teraz, że $\eta + \xi = a$ implikuje $\xi \leq a \leq s-1$, więc zachodzi $\xi_l \leq \xi \leq s-1$, dla $1 \leq l < t$. Zatem ostatecznie

$$\begin{aligned}
(10.5) \quad p_t = \Pr\{\eta + \xi \geq s \mid \mathbb{D}_t\} &\leq \\
&\leq \Pr\left\{|D(u) \cap D(v_t)| \geq s \mid |D(v_t)| = Z_t\right\} = p(Z_t),
\end{aligned}$$

gdzie $k = Z_t = |D(v_t)|$ pod warunkiem zajścia \mathbb{D}_t oraz dla danego k wartość $p(k)$ jest dana wzorem (10.3). Ostatecznie z (10.4) i (10.5) wynika

$$(10.6) \quad p(k^*) \leq p_t \leq p(k).$$

Czyli, aby oszacować p_t , wystarczy określić jakie wartości przyjmują zmienne Z_t^* i Z_t oraz określić $p(k)$ dla tych wartości. W tym celu dla ciągu $a = a(n) = 3\sqrt{\ln nmp} + 2nmp^2$ i $1 \leq l \leq t$ zdefiniujemy zdarzenia:

$$\begin{aligned}
B_{l1} &= \{Z_l^* \geq mp - a\}, \\
B_{l2} &= \{Z_l \leq mp + a\},
\end{aligned}$$

$$\mathbb{B}_l = \bigcap_{i=1}^l (B_{i1} \cap B_{i2}) = \{\forall_{1 \leq i \leq l} mp - a \leq Z_i^* \leq Z_i \leq mp + a\}.$$

Zauważmy także, że dla n, m i p spełniających założenia twierdzenia mamy $nmp^2 = o(mp)$ oraz $\sqrt{mp \ln n} = o(mp)$, więc $a = o(mp)$. Zatem pod warunkiem \mathbb{B}_{l-1} mamy $Z_{[l-1]} = |D_{[l-1]}| \leq (l-1)(mp+a) \leq 2nmp$ dla dostatecznie dużych n . Stąd wynika, że pod warunkiem \mathbb{B}_{l-1} , dla dużych wartości n , zmienna losowa Z_l^* dominuje stochastycznie nad zmienną losową Z^- o rozkładzie dwumianowym $\text{Bin}(m - 2nmp, p)$. Zatem, dla dużych wartości n , z nierówności Chernoffa (3.1) dla $1 \leq l \leq t$

$$\begin{aligned}
\Pr\left\{Z_l^* \leq mp - a \mid \mathbb{B}_{l-1}\right\} &\leq \Pr\{Z^- \leq mp - a\} = \\
&= \Pr\{Z^- \leq (m - 2nmp)p - 3\sqrt{\ln nmp}\} \leq \\
&\leq \exp\left(\frac{9mp \ln n}{2(m - 2nmp)p}\right) \leq n^{-4}.
\end{aligned}$$

Natomiast Z_l ma rozkład dwumianowy $\text{Bin}(m, p)$, więc z nierówności Chernoffa (3.2) dla dużych n mamy

$$\begin{aligned} \Pr\{Z_l \geq mp + a\} &\leq \Pr\left\{Z_l \geq mp + 3\sqrt{mp \ln n}\right\} \leq \\ &\leq \exp\left(\frac{27mp \ln n}{2(3mp + 3\sqrt{mp \ln n})}\right) \leq n^{-4}. \end{aligned}$$

Czyli można pokazać, że dla dużych wartości n (jednostajnie względem wszystkich t i wszystkich wierzchołków)

$$\begin{aligned} \Pr\{\mathbb{B}_t\} &= \Pr\{\mathbb{B}_{t-1} \cap (B_{t1} \cap B_{t2})\} = \\ (10.7) \quad &= \Pr\{(B_{t1} \cap B_{t2}) | \mathbb{B}_{t-1}\} \Pr\{\mathbb{B}_{t-1}\} \geq (1 - 2n^{-4}) \Pr\{\mathbb{B}_{t-1}\} \geq \\ &\geq \dots \geq (1 - 2n^{-4})^t \geq 1 - 2tn^{-4} = 1 - o(n^{-2}). \end{aligned}$$

Z drugiej strony mamy

$$\begin{aligned} p(k) &= \sum_{j \geq 0} \binom{k}{s+j} p^{s+j} (1-p)^{k-s-j} = \\ &= \binom{k}{s} p^s (1-p)^{k-s} \left(1 + \sum_{j \geq 1} \frac{(k-s)_j}{(s+j)_j} \left(\frac{p}{1-p}\right)^j\right). \end{aligned}$$

Stąd przy $k = (1 + o(1))mp$, ponieważ $kp = o(1)$, zachodzi

$$(1-p)^{k-s} = e^{kp + O((kp)^2)} = O(kp)$$

oraz

$$\sum_{j \geq 1} \frac{(k-s)_j}{(s+j)_j} \left(\frac{p}{1-p}\right)^j \leq \sum_{j \geq 1} \frac{(kp)^j}{j!} \leq e^{kp} - 1 = O(kp),$$

więc

$$p(k) = \frac{(k)_s p^s}{s!} (1 + o(1)) = \frac{(kp)^s}{s!} (1 + o(1)).$$

Ostatecznie po podstawieniu $kp = mp^2(1 + o(1))$ dla $p^* = m^s p^{2s}/s!$ asymptotycznie prawie na pewno

$$(10.8) \quad p_t = p^*(1 + o(1))$$

jednostajnie względem $v_1 \in \mathcal{V}$ i $v_t \in V_{BFS}^{t-1}$.

Podsumowując, udowodniliśmy, że pod warunkiem \mathbb{D}_t , liczba X_t wierzchołków dodanych do drzewa BFS w t -tym kroku ma rozkład dwumianowy $\text{Bin}(n_t, p_t)$, gdzie p_t spełnia (10.8). Zatem dopóki $V_{BFS}^t = O(n^{2/3})$ proces BFS może być ograniczony z dołu i z góry procesami gałęzowym $\mathcal{B}(P_-, P_-, N_-)$ i $\mathcal{B}(P_+, P_+, N_+)$ takimi, że $P_- \sim P_+ \sim p^*$ oraz $N_- \sim N_+ = n$.

Kolejne kroki dowodu (i) są identyczne z tymi z dowodu twierdzenia 5.4(i) z [32] (patrz też dowód twierdzenia 10.2(i)). Natomiast w dowodzie (ii), analogicznie jak w dowodzie twierdzenia 5.4(ii), można udowodnić, że asymptotycznie prawie na pewno dla każdego $v_1 \in \mathcal{V}$ proces BFS skończy się

w $k_- = k_-(n) = O(\ln n)$ krokach lub po $n^{2/3}$ krokach jest co najmniej $(c-1)n^{2/3}/2$ wierzchołków nierozpatrzonych w drzewie BFS. Rozważmy teraz dowolne dwa drzewa BFS o korzeniach w v_1 i v'_1 . Wiemy, że asymptotycznie prawie na pewno, jeśli drzewa te mają więcej niż k_- wierzchołków, to albo te drzewa połączą się przed wykonaniem kroku $n^{2/3}$ albo w każdym z nich jest $(c-1)n^{2/3}/2$ nierozpatrzonych wierzchołków. Oznaczmy zdarzenia:

$$\mathbb{D}(v_1) = \left(V_{BFS}^{n^{2/3}}(v_1); D(v_1), \dots, D(v_{n^{2/3}+(c-1)n^{2/3}/2}) \right)$$

$$\mathbb{D}(v'_1) = \left(V_{BFS}^{n^{2/3}}(v'_1); D(v'_1), \dots, D(v'_{n^{2/3}}) \right)$$

dla procesów BFS rozpoczętych w wierzchołkach v_1 i v'_1 . Niech $\mathbb{D}(v_1)$ oraz $\mathbb{D}(v'_1)$ będą takie, że $mp-a \leq |D^*(v_l)| \leq |D(v_l)| \leq mp+a$ i $|D(v'_l)| \leq mp+a$ dla $1 < l \leq n^{2/3} + (c-1)n^{2/3}/2$, $1 < l' \leq n^{2/3}$, $a = 3\sqrt{\ln nmp} + 2nmp^2$ oraz po kroku $n^{2/3}$ w każdym z procesów jest nadal $(c-1)n^{2/3}/2$ nierozpatrzonych wierzchołków. Zauważmy, że z (10.7) pierwszy warunek zachodzi z prawdopodobieństwem $1 - o(n^{-2})$, czyli z poprzednich rozważań wynika, że asymptotycznie prawie na pewno oba warunki zachodzą dla każdej pary drzew BFS w $G_s^{Bin}(\mathcal{V}, m, p)$ o co najmniej $n^{2/3}$ wierzchołkach. Oznaczmy zbiory składające się z tych $(c-1)n^{2/3}/2$ nierozpatrzonych wierzchołków przez A i A' , odpowiednio dla procesów rozpoczętych w v_1 i v'_1 . Z definicji zbiory $D^*(v_l)$ są rozłączne, więc pod warunkiem $\mathbb{D}(v_1)$, $\mathbb{D}(v'_1)$ i, że procesy się nie połączyły w $n^{2/3}$ pierwszych krokach, zdarzenia $\{|D^*(v_l) \cap D(v'_l)| < s\}$ są niezależne dla $n^{2/3} < l, l' \leq n^{2/3} + (c-1)n^{2/3}/2$. Czyli prawdopodobieństwo, że nie ma krawędzi między $v_l \in A$ i $v'_l \in A'$ (tzn. dla $n^{2/3} < l, l' \leq n^{2/3} + (c-1)n^{2/3}/2$) wynosi

$$\begin{aligned} & \Pr \left\{ \forall_{n^{2/3} < l, l' \leq n^{2/3} + (c-1)n^{2/3}/2} |D(v_l) \cap D(v'_l)| < s \mid \mathbb{D}(v_1), \mathbb{D}(v'_1) \right\} \leq \\ & \leq \Pr \left\{ \forall_{n^{2/3} < l, l' \leq n^{2/3} + (c-1)n^{2/3}/2} |D^*(v_l) \cap D(v'_l)| < s \mid \mathbb{D}(v_1), \mathbb{D}(v'_1) \right\} = \\ & = \prod_{n^{2/3} < l, l' \leq n^{2/3} + (c-1)n^{2/3}/2} \left(1 - \Pr \left\{ |D^*(v_l) \cap D(v'_l)| \geq s \mid \mathbb{D}(v_1), \mathbb{D}(v'_1) \right\} \right) \leq \\ & \leq \exp \left(-(1 + o(1)) \frac{m^s p^{2s} (c-1)n^{4/3}}{s!2} \right) = \\ & = \exp \left(-(1 + o(1)) \frac{c(c-1)^2 n^{1/3}}{4} \right) = o(n^{-2}). \end{aligned}$$

Czyli ostatecznie asymptotycznie prawie na pewno wszystkie wierzchołki zawarte w składowych o co najmniej $n^{2/3}$ wierzchołkach zawarte są w jednej składowej.

Podobnie jak w dowodzie twierdzenia 10.2(ii) określmy $Y_n = Y$ jako zmienną losową liczącą wierzchołki znajdujące się w składowych o co najwyżej k_- wierzchołkach. Postępując analogicznie jak w końcowej fazie dowodu twierdzenia 10.2(ii) można udowodnić, że z lematu 10.2, ograniczeń procesami gałęzkowymi, $\mathbb{E}Y \sim n\rho(c)$ i $\text{Var}Y \sim (\mathbb{E}Y)^2$, oraz z metody drugiego momentu (lemat 3.5) wynika teza twierdzenia. \square

11. Wnioski i problemy otwarte

W rozprawie badane były asymptotyczne własności losowych grafów przecięć z punktu widzenia analizowania sieci rzeczywistych. Rozważania skupiały się na dwóch rodzajach sieci rzeczywistych, dla których $G(\mathcal{V}, m, \mathcal{P}_{(m)})$ jest naturalnym modelem.

Po pierwsze wyznaczony został rozkład $\mathcal{P}_{(m)}$, dla którego $G(\mathcal{V}, m, \mathcal{P}_{(m)})$ dobrze modeluje sieci złożone. Pokazane zostało, że dla odpowiednio dobranych parametrów, graf $G(\mathcal{V}, m, \mathcal{P}_{(m)})$ ma rozkład stopni wierzchołków spełniający prawo potęgowe oraz asymptotycznie stały współczynnik skupienia. Dla tego rozkładu $\mathcal{P}_{(m)}$ pozostają otwarte pytania dotyczące średnicy, największej składowej i przejścia fazowego w $G(\mathcal{V}, m, \mathcal{P}_{(m)})$. W rozdziale 8 wspomniane zostało, że dokładna odpowiedź na pierwsze pytanie może być trudnym zadaniem. Z drugiej strony wydaje się, że metoda przybliżania procesem gałęzowym zaprezentowana w rozdziale 8 w połączeniu z metodą zastosowaną w pracach Bloznelisa [9, 10], bez większych problemów pozwoli udowodnić wyniki dotyczące średniej odległości między wierzchołkami i przejścia fazowego w rozpatrywanym w rozprawie losowym grafie przecięć przy $n \asymp m$. Z punktu widzenia modelowania sieci złożonych interesujące mogą być także problemy związane ze znalezieniem innych rozkładów $\mathcal{P}_{(m)}$, dla których $G(\mathcal{V}, m, \mathcal{P}_{(m)})$ ma dobre własności. Dla wszystkich tych rozkładów można się zapytać o globalny współczynnik skupienia i spróbować udowodnić analogiczne wyniki do tych zawartych w rozdziale 7.

Bardzo interesujące może być także zagadnienie zastosowania losowych grafów przecięć do skierowych sieci złożonych. Przykładem sieci złożonej, dla której naturalnym modelem może być graf skierowany jest sieć hiperlinków stron WWW. Wyniki zawarte w tej rozprawie dotyczą grafu podległego grafu skierowanego modelującego tę sieć. Ciekawe może być pytanie, czy można zdefiniować, bazujący na losowym grafie przecięć, model grafu skierowanego dobrze modelujący skierowane sieci złożone takie jak sieć hiperlinków stron WWW.

W rozprawie znajduje się także ostateczne rozwiązanie zagadnienia związanego ze spójnością grafu $G^{Jed}(\mathcal{V}, m, d)$, który jest teoretycznym modelem sieci sensorowej z losową predystrybucją kluczy. W dodatku określona została długość średnicy w tym grafie, co jest równoważne znalezieniu odległości, jaką musi pokonać informacja, aby dotrzeć do wszystkich sensorów połączonych w sieci. Pozostaje jednak nadal kilka interesujących problemów otwartych. W rozprawie przedstawione zostały wyniki dotyczące wielkości największej składowej w grafie $G_s^{Jed}(\mathcal{V}, m, d)$ przy dodatkowych założeniach o mocy zbioru \mathcal{W} . Dla $s = q$ graf ten jest teoretycznym modelem dla q -kompozytowego modelu sieci sensorowej z losową predystrybucją kluczy. Otwarte jednak nadal są zagadnienia dotyczące spójności i przejścia fazowego grafu

$G_s^{Jed}(\mathcal{V}, m, d)$ dla dowolnego doboru parametrów. W związku z większymi zależnościami między istniejącymi krawędziami w tym grafie nie można w tym wypadku zastosować technik przedstawionych w rozdziale 8. Istniejące wyniki sugerują, że funkcja $\tilde{d} = \tilde{d}(n)$ jest funkcją progową przejścia fazowego, jeśli

$$\tilde{d}^{2s} \asymp m^s s! / n.$$

o ile $\tilde{d} \geq d_{min}(s)$, gdzie $d_{min}(s)$ jest pewną funkcją zależną od s . Natomiast \tilde{d} jest funkcją progową spójności, jeśli

$$\tilde{d}^{2s} \asymp m^s s! \ln n / n$$

dla $\tilde{d} \geq d_{min}(s)$.

Kolejnym ciekawym kierunkiem dalszych prac nad sieciami sensorowymi jest badanie asymptotycznych własności modelu grafu, który jest połączeniem losowego grafu przecięć z losowym grafem geometrycznym. W przypadku, gdy do modelowania sieci sensorowej wykorzystuje się graf $G^{Jed}(\mathcal{V}, m, d)$, zakłada się, że sensory komunikują się wtedy i tylko wtedy, gdy mają wspólne klucze. Nie uwzględnia się zatem fizycznych ograniczeń komunikacji takich jak moc nadajników. Alternatywą jest rozpatrywanie sieci bezprzewodowej z ograniczoną mocą nadajników. Dobrym jej modelem jest losowy graf geometryczny, w którym wierzchołki są połączone wtedy i tylko wtedy, gdy leżą na płaszczyźnie w odległości co najwyżej r (r jest z góry ustalonym parametrem grafu). Z definicji tego modelu wynika, że rezultaty dla niego uzyskane nie mogą być stosowane do sieci sensorowych z losową predystrybucją kluczy. Z punktu widzenia modelowania sieci sensorowych interesujące byłyby zatem uzyskanie wyników o spójności i przejściu fazowym modelu grafu losowego, w którym krawędzie powstają wtedy i tylko wtedy gdy istnieją one w losowym grafie przecięć i w losowym grafie geometrycznym zdefiniowanych na tym samym zbiorze wierzchołków. Dotychczas znane są jedynie wyniki eksperymentalne dotyczące takiego modelu. Jedyne znane wyniki analityczne dotyczą modelu, w którym przyjęte jest dodatkowe założenie o rozłożeniu wierzchołków oraz dostosowanym do niego protokole przydzielania kluczy (patrz [36]).

Notacja

Notacja asymptotyczna

$a_n = O(b_n)$,	
$a_n = \Omega(b_n)$,	
$a_n = o(b_n)$,	patrz podrozdział 1.1
$a_n \sim b_n$,	
$a_n \asymp b_n$	

Rozkłady prawdopodobieństwa i zmienne losowe

$\text{Bin}(n, p)$	rozkład dwumianowy o parametrach n i p
$\text{Po}(\lambda)$	rozkład Poissona o wartości oczekiwanej λ
$X \prec Y$	zmienna losowa X jest stochastycznie zdominowana przez zmienną losową Y
$\mathcal{P} \prec \mathcal{P}'$	rozkład prawdopodobieństwa \mathcal{P} jest stochastycznie zdominowany przez rozkład prawdopodobieństwa \mathcal{P}'

Modele grafów losowych

$G(n, p), G(n, M)$	klasyczne grafy losowe (patrz przykład 1.1)
$G_s(\mathcal{V}, m, \mathcal{P}_{(m)})$	losowy graf przecięć (definicja 1.3)
$G_s^{\text{Bin}}(\mathcal{V}, m, p)$	dwumianowy losowy graf przecięć (definicja 1.4)
$G_s^{\text{Jed}}(\mathcal{V}, m, d)$	jednostajny losowy graf przecięć (definicja 1.5)
$G^{\text{Bin}}(\mathcal{V}, m, W)$	warunkowy dwumianowy losowy graf przecięć, W – zmienna losowa (definicja 1.6)
\mathcal{V}	zbiór wierzchołków losowego grafu przecięć
\mathcal{W}	pomocniczy zbiór własności losowego grafu przecięć
n	liczba wierzchołków losowego grafu przecięć (moc zbioru \mathcal{V})
$m(n)$	liczba własności losowego grafu przecięć (moc zbioru \mathcal{W})
$D(v)$	zbiór własności wierzchołka $v \in \mathcal{V}$
$Z_n = D(v) $	moc zbioru własności wierzchołka $v \in \mathcal{V}$
$\mathcal{P}_{(m)} = (P_0, \dots, P_m)$	rozkład prawdopodobieństwa decydujący o mocach zbiorów własności wierzchołków w grafie $G_s(\mathcal{V}, m, \mathcal{P}_{(m)})$

Bibliografia

- [1] R. Albert i A-L. Barabási, *Emergence of scaling in random networks*, Science **286** (1999), 509512.
- [2] R. Albert i A-L. Barabási, *Statistical mechanics of complex networks*, Reviews of modern physics **74** (2002), 47–97.
- [3] A. D. Barbour, L. Holst, i S. Janson, *Poisson approximation*, Oxford University Press, 1992.
- [4] A. Barrat i M. Weigt, *On the properties of small-world networks*, The European Physical Journal B **13** (2000), 547–560.
- [5] M. Behrisch, *Component evolution in random intersection graphs*, The Electronic Journal of Combinatorics **14** (2007), no. 1.
- [6] M. Behrisch i A. Taraz, *Efficiently covering complex networks with cliques of similar vertices*, Theoretical Computer Science **355** (2006), no. 1, 37–47.
- [7] M. Behrisch, A. Taraz, i M. Ueckerdt, *Coloring random intersection graphs and complex networks*, SIAM Journal on Discrete Mathematics **23** (2009), no. 1, 288–299.
- [8] S. R. Blackburn i S. Gerke, *Connectivity of the uniform random intersection graph*, Discrete Mathematics **309** (2009), 5130–5140.
- [9] M. Bloznelis, *Component evolution in general random intersection graphs*.
- [10] M. Bloznelis, *A note on loglog distances in a power law random intersection graph*, <http://arxiv.org/abs/0911.5127>.
- [11] M. Bloznelis, *Degree distribution of a typical vertex in a general random intersection graph*, Lithuanian Mathematical Journal **48** (2008), no. 1, 38–45.
- [12] M. Bloznelis, J. Jaworski, i K. Rybarczyk, *Component evolution in a secure wireless sensor network*, Networks **53** (2009), no. 1, 19–26.
- [13] B. Bollobás, *Random graphs*, Academic Press, 1985.
- [14] B. Bollobás, S. Janson, i O. Riordan, *The phase transition in inhomogeneous random graphs*, Random Structures and Algorithms **31** (2007), no. 1, 3–122.
- [15] H. Chan, A. Perrig, i D. Song, *Random key predistribution schemes for sensor networks*, SP '03: Proceedings of the 2003 IEEE Symposium on Security and Privacy (Washington, DC, USA), IEEE Computer Society, 2003, pp. 197–213.
- [16] F. Chung i L. Lu, *The diameter of sparse random graphs*, Adv. in Appl. Math. **26** (2001), no. 4, 257–279.
- [17] K. L. Chung, *A course in probability theory*, Harcourt, New York, 1968.
- [18] T. H. Cormen, *Wprowadzenie do algorytmów*, Wydawnictwa Naukowo – Techniczne, 2005.
- [19] M. Deijfen i W. Kets, *Random intersection graphs with tunable degree distribution and clustering*, Probability in the Engineering and Information Sciences **23** (2009), 661–674.
- [20] C. Efthymiou i P. G. Spirakis, *On the existence of hamiltonian cycles in random intersection graphs*, Automata, Languages and Programming 32nd International Colloquium, ICALP 2005, Lisbon, Portugal, July 11-15, 2005. Proceedings, 2005, pp. 690–701.

- [21] P. Erdős i A. Rényi, *On random graphs I*, Publicationes Mathematicae Debrecen **6** (1959), 290–297.
- [22] P. Erdős i A. Rényi, *On the evolution of random graphs*, Publ. Math. Inst. Hungar. Acad. Sci. **5** (1960), 17–61.
- [23] P. Erdős i J. Spencer, *Probabilistic methods in combinatorics*, Akadémia Kiadó, 1974.
- [24] L. Eschenauer i V. D. Gligor, *A key-management scheme for distributed sensor networks*, CCS '02: Proceedings of the 9th ACM conference on Computer and communications security (New York, NY, USA), ACM Press, 2002, pp. 41–47.
- [25] J. A. Fill, E. R. Scheinerman, i K. B. Singer-Cohen, *Random intersection graphs when $m = \omega(n)$: An equivalence theorem relating the evolution of the $G(n, m, p)$ and $G(n, p)$ models*, Random Structures and Algorithms **16** (2000), 156–176.
- [26] E. Godehardt i J. Jaworski, *Two models of random intersection graphs for classification*, Studies in Classification, Data Analysis and Knowledge Organization, Springer, 2003, pp. 67–81.
- [27] E. Godehardt, J. Jaworski, i K. Rybarczyk, *Random intersection graphs and classification*, Studies in Classification, Data Analysis, and Knowledge Organization, Springer, 2007, pp. 67–74.
- [28] E. Godehardt, J. Jaworski, i K. Rybarczyk, *Isolated vertices in random intersection graph*, Studies in Classification, Data Analysis, and Knowledge Organization, no. 36, Springer Verlag Heidelberg - Berlin, 2009, pp. 135–145.
- [29] J.-L. Guillaume i M. Latapy, *Bipartite structure of all complex networks*, Information Processing Letters **90** (2004), no. 5, 215–221.
- [30] J. Hwang i Y. Kim, *Revisiting random key pre-distribution schemes for wireless sensor networks*, SASN '04: Proceedings of the 2nd ACM workshop on Security of ad hoc and sensor networks, ACM, 2004, pp. 43–52.
- [31] J. Jakubowski i R. Sztencel, *Wstęp do rachunku prawdopodobieństwa*, Script, 2001.
- [32] S. Janson, T. Łuczak, i A. Ruciński, *Random graphs*, Wiley, 2001.
- [33] J. Jaworski i M. Karoński, *On the connectivity of graphs generated by a sum of random mappings*, Journal of Graph Theory **17** (1993), no. 2, 135–150.
- [34] J. Jaworski, M. Karoński, i D. Stark, *The degree of a typical vertex in generalized random intersection graph models*, Discrete Mathematics **306** (2006), 2152–2165.
- [35] J. Jaworski i Z. Palka, *Remarks on a general model of a random digraph*, Ars Combinatoria **65** (2002), 135–144.
- [36] J. Jaworski, M. Ren, i K. Rybarczyk, *Random key predistribution for wireless sensor networks using deployment knowledge*, Computing **85** (2009), no. 1–2, 57–76.
- [37] J. Jaworski i I. Smit, *On a random digraph*, Annals of Discrete Mathematics **33** (1987), 111–127.
- [38] J. Jaworski i D. Stark, *The vertex degree distribution of passive random intersection graph models*, Combinatorics, Probability and Computing **17** (2008), no. 4, 549–558.
- [39] M. Karoński i A. Ruciński, *Small subgraphs of k -partite random graphs*, Proceedings of the Third International Conference on Combinatorial Mathematics (New York, NY, USA), New York Academy of Sciences, 1989, pp. 230–240.
- [40] M. Karoński, E. R. Scheinerman, i K.B. Singer-Cohen, *On random intersec-*

- tion graphs: The subgraph problem*, *Combinatorics, Probability and Computing* **8** (1999), 131–159.
- [41] R. M. Karp, *The transitive closure of a random digraph*, *Random Structures and Algorithms* **1** (1990), 73–93.
- [42] A. N. Lagerås i M. Lindholm, *A note on the component structure in random intersection graphs with tunable clustering*, *The Electronical Journal of Combinatorics* **15(1)** (2008), N10.
- [43] T. Łuczak, *On the equivalence of two basic models of random graphs*, *Random Graphs 87'* (Karoński M., Jaworski J., i Ruciński A., eds.), John Wiley & Sons, 1990, pp. 151–158.
- [44] M. Molloy i B. Reed, *A critical point for random graphs with a given degree sequence*, *Random Structures and Algorithms* **6** (1995), 161–179.
- [45] M. Molloy i B. Reed, *The size of the giant component of a random graph with a given degree sequence*, *Combinatorics, Probability and Computing* **7** (1998), 295–305.
- [46] M. E. J. Newman, *Properties of highly clustered networks*, *Physical Review E* **68** (2003), 026121.
- [47] M. E. J. Newman, *The structure and function of complex networks*, *SIAM Review* **45** (2003), 167–256.
- [48] M. E. J. Newman, *Power laws, Pareto distributions and Zipf's law*, *Contemporary Physics* **46** (2005), no. 5, 323–351.
- [49] M. E. J. Newman, S. H. Strogatz, i D. J. Watts, *Random graphs with arbitrary degree distributions and their applications*, *Physical Review E* **64** (2001), 026118.
- [50] M. E. J. Newman i D. J. Watts, *Renormalization group analysis of the small-world network model*, *Physics Letters A* **263** (1999), 341–346.
- [51] M. E. J. Newman i D. J. Watts, *Scaling and percolation in the small-world network model*, *Physical Review E* **60** (1999), 7332–7342.
- [52] S. E. Nikolettseas, C. Raptopoulos, i P. G. Spirakis, *The existence and efficient construction of large independent sets in general random intersection graphs*, *Automata, Languages and Programming 31st International Colloquium, ICALP 2004, Turku, Finland, July 12-16, 2004. Proceedings, LNCS, 2004*, pp. 1029–1040.
- [53] S. E. Nikolettseas, C. Raptopoulos, i P. G. Spirakis, *Expander properties and the cover time of random intersection graphs*, *Mathematical Foundations of Computer Science 2007 32nd International Symposium, MFCS 2007 Cesk Krumlov, Czech Republic, August 26-31, 2007 Proceedings, 2007*.
- [54] Z. Palka i A. Ruciński, *Niekonstrukttywne metody matematyki dyskretnej*, Wydawnictwa Naukowo-Techniczne, 1996.
- [55] R. Di Pietro, L. V. Mancini, A. Mei, A. Panconesi, i J. Radhakrishnan, *Sensor networks that are provably resilient*, *Proc 2nd IEEE Int Conf Security Privacy Emerging Areas Commun Networks (SecureComm 2006)*, Baltimore, MD, 2006, 2006.
- [56] K. Rybarczyk, *O pewnych zastosowaniach hipergrafów losowych*, praca magisterska, Wydział Matematyki i Informatyki, Uniwersytet Adama Mickiewicza w Poznaniu, 2005.
- [57] K. Rybarczyk i D. Stark, *Poisson approximation of the number of cliques in the random intersection graph*, złożona do *Journal of Applied Probability*.
- [58] K. B. Singer-Cohen, *Random intersection graphs*, praca doktorska, Department of Mathematical Sciences, The Johns Hopkins University, 1995.

- [59] D. Stark, *The vertex degree distribution of random intersection graphs*, Random Structures and Algorithms **24** (2004), 249–258.
- [60] S. H. Strogatz i D. J. Watts, *Collective dynamics of small-world networks*, Nature **393** (1998), 440–442.