

POLITECHNIKA WROCŁAWSKA
WYDZIAŁ INFORMATYKI I ZARZĄDZANIA

KIERUNEK: Informatyka (INF)
SPECJALNOŚĆ: Systemy Baz Danych (SBD)

**PRACA DYPLOMOWA
MAGISTERSKA**

Analiza łańcucha transakcji w sieci Bitcoin

The analysis of Bitcoin transactions blockchain

AUTOR:
Bartosz Zychal

PROWADZĄCY PRACĘ:
dr inż. Radosław Michalski

OCENA PRACY:

Spis treści

Wprowadzenie	1
1. Kryptowaluty - wprowadzenie	2
1.1. Wprowadzenie	2
1.2. Definicja	2
1.3. Bezpieczeństwo w walutach i kryptowalutach	2
1.4. Kryptografia	3
1.5. Zastosowanie kryptografii w sieci Bitcoin	3
1.6. Metoda tworzenia kluczy publicznych na przykładzie sieci Bitcoin	4
1.7. Podsumowanie	6
2. Blockchain - rejestr transakcji	7
2.1. Wprowadzenie	7
2.2. Definicja Blockchain'a	7
2.3. Struktura i zawartość bloku	10
2.4. Transakcje i opłaty	11
2.5. Dołączania kolejnego bloku z transakcjami	13
2.6. Podsumowanie	14
3. Przegląd metod analiz sieci złożonych (też temporalnych) oraz analiz blockchaina	15
4. Analiza blockchaina Bitcoin	16
4.1. Wprowadzenie	16
4.2. Plan badań	16
4.3. Eksperyment	19
4.3.1. Badanie średnicy sieci	19
4.3.2. Badanie średniej długość ścieżki	21
4.3.3. Badanie średniego stopnia węzła	23
4.3.4. Badanie średniej centralności węzła	25
4.3.5. Badanie średniej wartości transakcji	27
4.3.6. Badanie ilości bloków	29
4.3.7. Badanie średniej różnicy czasów kolejnych transakcji	30
4.3.8. Badanie różnicy czasów granicznych transakcji	32
4.4. Wnioski	34
Podsumowanie	35
Bibliografia	35

Wprowadzenie

Do napisania na końcu. Ma zawierać informacje o motywacji i celu pracy.

Rozdział 1.

Kryptowaluty - wprowadzenie

1.1. Wprowadzenie

Rola niniejszego rozdziału jest wyjaśnienie istoty oraz sposobu działania kryptowalut. Przedstawiono w nim, czym są kryptowaluty oraz kriptografia, w jaki sposób kryptowaluty korespondują z kriptografią oraz w jaki sposób kriptografia pozwala zapewnić wysokie bezpieczeństwo kryptowaluty. Na przykładzie jednej z najpopularniejszych walut cyfrowych, tj. Bitcoin, zaprezentowano zastosowanie kriptografii na potrzeby jej zabezpieczenia.

1.2. Definicja

Kryptowaluta to cyfrowy zasób mogący odpowiadać pewnej wartości środków finansowych. Zasób ten został zaprojektowany w sposób pozwalający określić go jako medium wymiany przy użyciu kriptografii, z którą jest ściśle powiązany. Kriptografia pozwala na zabezpieczenie transakcji, kontrolowanie tworzenia nowych jednostek kryptowaluty oraz weryfikację ilości posiadanych jednostek. Aktualnie kryptowaluty klasyfikowane są do trzech grup:

- walut cyfrowych,
- walut alternatywnych,
- walut wirtualnych.

1.3. Bezpieczeństwo w walutach i kryptowalutach

Wszystkie waluty muszą być w jakiś sposób kontrolowane i podlegać różnego rodzaju zabezpieczeniom, tak aby zapobiegać oszustwom. W przypadku walut fiduciarnych, tj. walut nie mających pokrycia w dobrach materialnych, organizacje takie jak banki kontrolują podaż pieniądza oraz oznaczają fizycznie walutę, w celu uniemożliwienia jej podrobienia. Takie zabezpieczenia w pewnym stopniu ograniczają możliwości fałszerstwa, jednakże nie dają stuprocentowej pewności. Kryptowaluty podobnie jak tradycyjne waluty muszą posiadać miary zabezpieczeń w celu uniemożliwienia wpływu na stan systemu i tworzenia niekonkurencyjnych danych. Dodatkowo muszą one posiadać zabezpieczenia niepozwalające na wielokrotne użycie tych samych środków. W przeciwieństwie do walut fiduciarnych zasady bezpieczeństwa kryptowalut mogą bazować wyłącznie na istniejących

technologiach i nie mogą podlegać kontroli ze strony jakiejkolwiek centralnej instytucji [1].

1.4. Kryptografia

Kryptowaluty bardzo silnie bazują na kryptografii, która oferuje mechanizm bezpiecznego kodowania zasad ich systemu. Kryptografia pozwala nie tylko bronić system przed manipulacjami i mactactwami, ale również dobrze może zostać użyta w celu kodowania zasad tworzenia nowych jednostek kryptowaluty przy pomocy określonego matematycznego protokołu[2].

Kryptografię można sklasyfikować jako dziedzinę wiedzy o zabezpieczeniach przed nieautoryzowanym dostępem do informacji. W dzisiejszych czasach uważa się ją nie tylko za gałąź matematyki, ale i informatyki. Kryptografię można podzielić na:

- A. Symetryczną - polega na możliwości odczytania wiadomości przy pomocy tego samego klucza, którym została podpisana. Znaczącym problemem bezpieczeństwa w tym podejściu jest przekazanie odbiorcy klucza.
- B. Niesymetryczną - polega na istnieniu co najmniej dwóch kluczy:
 - Prywatny - klucz ten nie powinien być nigdy nikomu udostępniony. Przy pomocy klucza prywatnego można odszyfrować wiadomość podpisana kluczem publicznym. Pozwala również na podpisanie wiadomości, która może być później zweryfikowana za pomocą klucza publicznego.
 - Publiczny - klucz ten może zostać bez żadnych zastrzeżeń upubliczony. Klucz publiczny tworzy się na podstawie klucza prywatnego, jednakże odtworzenie klucza prywatnego z klucza publicznego jest bardzo trudne. Klucz publiczny używany jest do szyfrowania wiadomości oraz weryfikacji wiadomości podpisanej kluczem prywatnym.

Kryptografia oparta na kluczu publicznym została opracowana w latach 70. XX wieku i cały czas stanowi solidną podstawę bezpieczeństwa komputerowego i informacyjnego. Od czasu jej powstania odkryto funkcje matematyczne, które są praktycznie nieodwracalne, (np. *potęgowanie liczby pierwszej i mnożenie krzywych eliptycznych*), co oznacza, że łatwo obliczyć je w jednym kierunku, jednakże operacja odwrotna jest praktycznie niewykonywalna. Bitcoin korzysta z mnożenia krzywej eliptycznej jako podstawy przy wyliczaniu klucza publicznego. W podrozdziale 1.6. przedstawiono sposób użycia tej funkcji.

1.5. Zastosowanie kryptografii w sieci Bitcoin

Aktualnie na rynku dostępne jest ponad tysiąc różnych kryptowalut, a wraz z rosnącym zainteresowaniem oraz zaufaniem społecznym ilość walut cyfrowych cały czas rośnie[3]. Niepodważalny jest fakt, że jedną z najbardziej powszechnych i popularnych kryptowalut jest Bitcoin. Bitcoin jest całkowicie zdecentralizowaną, zdigitalizowaną walutą bez globalnego emitenta, który miałby nią zarządzać oraz ją rozpowszechniać. Bazując na specjalistycznym otwartym oprogramowaniu pewna ilość bitcoinów przekazywana jest użytkownikom w zamian za działania pozwalające na poprawne funkcjonowanie systemu Bitcoin. Użytkownicy Ci zwani są kopaczami lub górnikami, a operacje przez nich wykonywane, w celu

podtrzymania systemu, zwane są kopaniem. Kopanie bitcoinów, poza zyskiem ze strony kopaczy, daje olbrzymi zysk dla systemu, pozwalając weryfikować zlecone transakcje.

Właściciele bitcoinów ustalani są na podstawie kluczy cyfrowych, adresów Bitcoin oraz podpisów cyfrowych. Klucze cyfrowe nie są przechowywane w sieci, jednakże są tworzone przez użytkowników oraz przetrzymywane w ich portfelach, w plikach lub bazie danych. Klucz cyfrowy jest całkowicie niezależny od protokołu sieci Bitcoin, dlatego też może być tworzony przez różnie oprogramowania. Oprogramowanie to musi zapewniać użycie bezpiecznego źródła entropii w celu wygenerowania unikalnego klucza[4]. Wygenerowanie istniejącego lub zbyt słabego klucza może spowodować, iż w przyszłości użytkownik utraci zebrane środki. Klucze zapewniają w sieci Bitcoin:

- zdecentralizowane zaufanie,
- zaświadczenie o własności,
- odporny na kryptografię model bezpieczeństwa.

Transakcje w sieci Bitcoin wymagają dodania prawidłowego podpisu do łańcucha bloków, co dokładniej zostało opisane w rozdziale 2.. Podpis ten może być wygenerowany przy pomocy ważnych kluczy cyfrowych. Każdy kto posiada kopię tych kluczy może kontrolować środki dostępne na koncie. Protokół sieci Bitcoin korzysta z szyfrowania asymetrycznego, co w konsekwencji oznacza, że w transakcji klucz publiczny odbiorcy jest prezentowy przez jego odcisk palca, zwany adresem Bitcoin. Adresy te są ogólnodostępne i widoczne przez wszystkich[5].

Z klucza publicznego korzysta się w celu odebrania Bitcoinów, natomiast klucz prywatny wymagany jest do wydawania Bitcoinów. Osoba wydająca Bitcoiny musi zaprezentować swój klucz publiczny oraz podpis w transakcji. Podpis za każdym razem jest inny, lecz tworzony z jednego klucza prywatnego, co pozwala na udaną weryfikację przy pomocy dołączonego klucza publicznego. Poprzez załączenie obu tych informacji każdy w sieci może zweryfikować oraz zaakceptować transakcję jako poprawną lub ją odrzucić w przypadku stwierdzenia, braku środków na adresie nadawcy.

Klucz prywatny w sieci Bitcoin powiązany jest ścisłe z adresem, dlatego też jego utrata powoduje nieodwracalną utratę środków. Pomimo iż środki są cały czas dostępne, nie mogą zostać użyte bez prawidłowego podpisu generowanego z klucza prywatnego.

1.6. Metoda tworzenia kluczy publicznych na przykładzie sieci Bitcoin

Jak już wcześniej wspomniano klucz publiczny obliczany jest z klucza prywatnego przy pomocy *mnożenia krzywej eliptycznej*. Metoda ta została szczegółowo opisana przez *Andreas M. Antonopoulos*[4], jednakże w celu przedstawienia siły zabezpieczenia, jakie daje szyfrowanie asymetryczne, przedstawiono ją poniżej w skróconej formie.

Klucz publiczny jest praktycznie nieodwracalny i sposób jego wyliczania można zapisać jako:

$$K = k * G, \quad (1.1.)$$

gdzie:

k jest wartością klucza prywatnego,

G jest stałym punktem zwanym punktem generującym,

K jest wynikowym kluczem publicznym.

Kryptografia krzywej eliptycznej jest rodzajem kryptografi asymetrycznej bazującej na problemie logarytmu dyskretnego wyrażonego jako sumy i iloczyny punktów na tej krzywej eliptycznej. Dlatego też, operacją odwrotną do mnożenia krzywej eliptycznej jest *odnalezienie logarytmu dyskretnego*, co wymaga zastosowania wyszukiwania przy pomocy algorytmu typu *brute-force*, czyli przeglądu zupełnego.

W przypadku Bitcoina parametry krzywej eliptycznej są ścisłe określone i zdefiniowane przy pomocy standardu zwanego *secp256k1*. Standard ten został ustalony przez amerykańską Narodową Instytucję Standaryzacji i Technologii. Zastosowana w tej kryptowalucie krzywa eliptyczna tworzona jest na podstawie określonego zbioru stałych matematycznych i jest wyrażana przy pomocy funkcji:

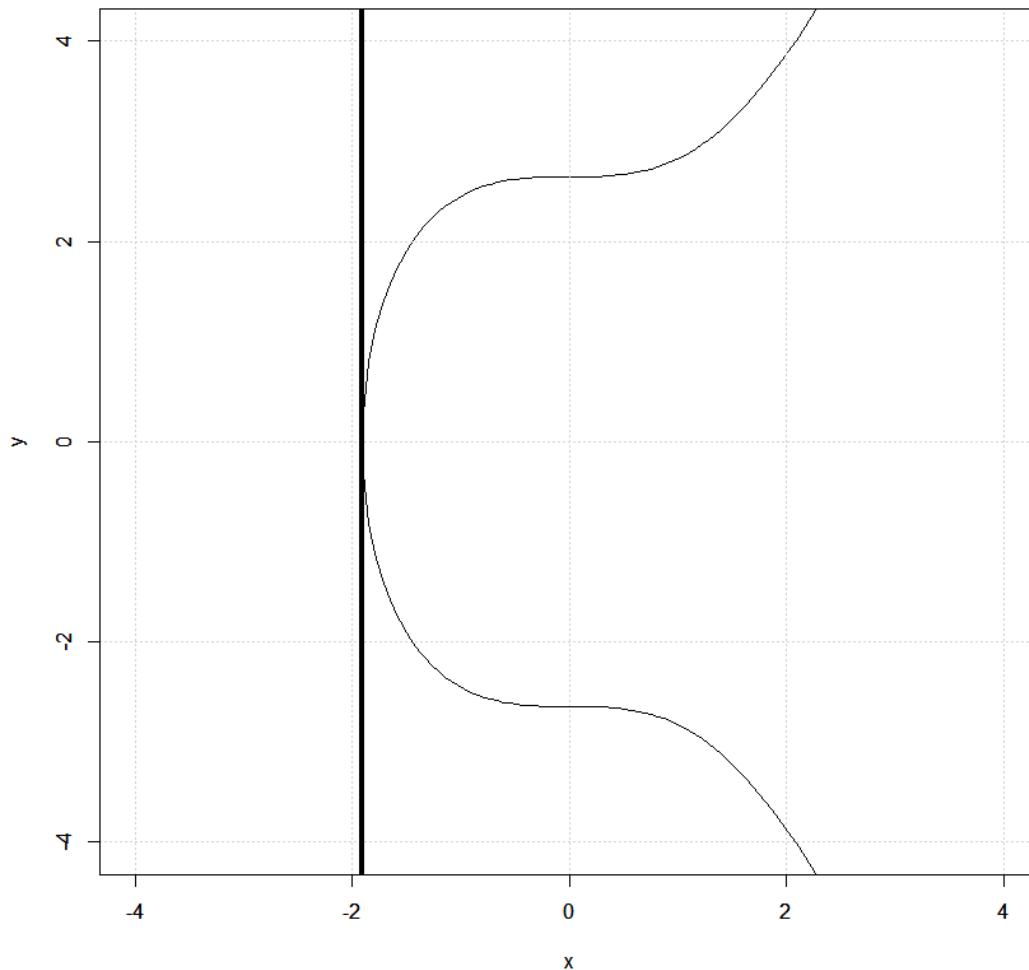
$$y^2 = x^3 + 7 \quad (1.2.)$$

lub:

$$y^2 \mod p = (x^3 + 7) \mod p \quad (1.3.)$$

Moduł liczby pierwszej $\mod p$ implikuje właściwość krzywej jako znajdującej się nad skończonym polem pierwszego rzędu p . Można ją również zapisać jako funkcję F_p , gdzie $p = 2^{256} - 2^{32} - 2^9 - 2^8 - 2^7 - 2^6 - 2^4 - 1$ jest olbrzymią liczbą pierwszą. Oznacza to, że od pewnego momentu krzywa zdefiniowana jest przy pomocy liczb zespolonych, a nie rzeczywistych. Utrudnia to jej wizualizację, gdyż wykres takiej funkcji musiałby zostać przedstawiony w dwóch wymiarach i składałby się z wielu pojedynczych punktów w przestrzeni. Na potrzeby graficznego przedstawienia uproszczono wykres 1.1. funkcji 1.6. przedstawiając go tylko w świecie liczb rzeczywistych. Wykres ten został podzielony na dwa obszary przy pomocy pionowej kreski, która nie jest częścią wykresu funkcji. Lewy obszar obejmuje wartości funkcji w świecie liczb zespolonych, który pominięto, natomiast prawa strona wykresu przedstawia funkcję 1.6. w świecie liczb rzeczywistych.

Parametr G został wybrany na podstawie krzywej eliptycznej - przedstawionej na wykresie 1.1. - przez protokół Bitcoina przy użyciu standardu *secp256k1*. Parametr ten jest przypisywany do wszystkich użytkowników sieci. Determinuje to wygenerowanie za każdym razem takiego samego klucza publicznego na podstawie tego samego klucza prywatnego. Klucz prywatny może mieć wartość od 1 do prawie 2^{256} , a zależność pomiędzy jego wartością k i wartością klucza publicznego K jest stała. Oznacza to, że aby przy znajomości stałej wartości G odtworzyć wartość klucza publicznego K wymagany jest przegląd wszystkich możliwych wartości klucza prywatnego k .



Rysunek 1.1.: Krzywa eliptyczna zastosowana do wyznaczenia wartości G w protokole Bitcoin

1.7. Podsumowanie

Reasumując, za każdą walutą musi stać określony system zabezpieczeń. W przypadku tradycyjnych walut są to centralne instytucje nadzorujące obrót i podaż określonego pieniądza. W przypadku kryptowalut bezpieczeństwo zapewnione jest poprzez zastosowanie prostej w użyciu, aczkolwiek skomplikowanej w budowie kriptografii. Zapewnia to możliwości bardzo szybkiej weryfikacji posiadanych środków przy bardzo niskim nakładzie pracy. Dodatkowo kriptografia, w porównaniu do tradycyjnych metod zabezpieczania pieniądza, nie pozwala na kontrolowanie przez jedną osobę, organizację czy instytucję prawa do danych. To zadanie wykonywane jest przez wszystkich kopaczów w sieci Bitcoin.

Rozdział 2.

Blockchain - rejestr transakcji

2.1. Wprowadzenie

W tym rozdziale przedstawiono oraz wyjaśniono ideę rejestrów transakcji, tzw. blockchaina. Opisano strukturę bloku, transakcji oraz wejść i wyjście w transakcji. Rozdział ten opisuje również w jaki sposób tworzone są transakcje, jak są weryfikowane i jakie koszty ponosi się za przekazanie Bitcoinów innemu uczestnikowi sieci. Dodatkowo przedstawiono proces tworzenia oraz dodawania nowego bloku do blockchaina.

2.2. Definicja Blockchain'a

Łancuch bloków (ang. Blockchain) jest uporządkowaną strukturą, zwaną jednokierunkową listą składającą się z bloków transakcji. Listę tę charakteryzuje połączenie wsteczne, co oznacza, że blok następny wskazuje na blok poprzedni. Każdy kolejny blok ma przypisaną swoją wysokość w łańcuchu bloków. Wysokość ta ustalana jest na podstawie odległości bloku od pierwszego bloku w łańcuchu. Blok ten, zwany blokiem genezy, stanowi pierwszego *rodzica* oraz wspólnego przodka dla wszystkich bloków w całym łańcuchu[6]. Na ilustracji 2.1. przedstawiona została idea łańcucha bloków.

Pierwszy blok łańcucha bloków w sieci Bitcoin został wygenerowany w 2009 roku i jako jedyny z bloków zapisany jest w kodzie źródłowym oprogramowania - zapewnia to zaufanego wspólnego przodka dla wszystkich bloków łańcucha.

Bloki rozpoznaje się na podstawie unikalnego hash'a, który generowany jest przy użyciu algorytmu kryptograficznego SHA256. Wynikowy hash ma 256 bitów i jest prezentowany zazwyczaj w systemie heksadecymalnym, co ułatwia jego odczyt. Bloki nie zawierają swojego hash'a w nagłówku, co komplikuje szybkie odnajdowanie konkretnego bloku. W celu łatwej i szybkiej identyfikacji, dane o bloku muszą być przetrzymywane w bazie danych zawierającej dodatkowe informacje o łańcuchu bloków. Każdy z węzłów sieci oblicza hash bloku w trakcie jego odbierania. W nagłówku blok zawiera informację tylko o rodzicu, tzn. przechowuje hash rodzica. Poruszając się wstecz przy pomocy tych hash'ów powróci się do pierwszego bloku - bloku genezy.

Hash bloku powstaje na podstawie jego zawartości włączając w to nagłówek bloku, w którym znajduje się wspomniany hash poprzednika. Powoduje to, iż ingerencja w dane znajdujące się w blokach staje się niemożliwa ze względu na ilość obliczeń jakie należałyby wykonać, aby je zmienić. Jakakolwiek zmiana w bloku rodzica wymusza zmianę w bloku dziecka, dlatego im blok jest starszy tym jest bezpieczniejszy. Taka struktura stanowi podstawę bezpieczeństwa w sieci Bitcoin.

Im blok znajduje się dalej w łańcuchu tym jest stabilniejszy i pewniejszy, natomiast ostatnie bloki, tzn. najświeższe, mogą ulegać zmianom w wyścigu prowadzonym przez kopaczy. W tym samym momencie na ostatni blok może wskazywać wiele nowych bloków wyprodukowanych w trakcie kopania. Taka sytuacja prowadzi do rozgałęzienia się łańcucha bloków. Jest to nieakceptowne w opisywanej strukturze, dlatego ostatecznie do łańcucha bloków dołączany zostaje jeden z wygenerowanych bloków, a reszta bloków zostaje odrzucona. Bardziej szczegółowo problem ten został opisany w podrozdziale 2.5..

Hash bloku	0000000000000000a7b47a1e58e456 fd54ae5a30cb92a35ca3e5acee065287
Wysokość bloku	496201
Rozmiar:	1071.607 kB
Liczba transakcji:	2032
Nagłówek bloku	
Wersja:	0x20000000
Hash poprzedniego bloku:	0000000000000000c6dd215947b569 fa06de2cb856dec643daf5a7e8efc72e
Markle root:	b96da6d09865e36e4862b5a612fb1893 275d889989feb5a7a217503fd84019e3
Czas wykopania:	2017-11-26 13:51:02
Trudność:	1,347,001,430,558.57
Transakcje	



Hash bloku	0000000000000000c6dd215947b569 fa06de2cb856dec643daf5a7e8efc72e
Wysokość bloku	496200
Rozmiar:	1062.264 kB
Liczba transakcji:	1036
Nagłówek bloku	
Wersja:	0x20000000
Hash poprzedniego bloku:	0000000000000000c645c295fbd91d f2f53e0346a4ecf8821abc8cc18fa4b1
Markle root:	c6be2f384d7fa195dfac1ce367ee3a9f 9ccfaf9cd337b261be1a0ff0e5d73bc4
Czas wykopania:	2017-11-26 13:43:52
Trudność:	1,347,001,430,558.57
Transakcje	



Hash bloku	0000000000000000c645c295fbd91d f2f53e0346a4ecf8821abc8cc18fa4b1
Wysokość bloku	496199
Rozmiar:	1059.467 kB
Liczba transakcji:	1758
Nagłówek bloku	
Wersja:	0x20000000
Hash poprzedniego bloku:	00000000000000004f4c7ee2fad2a6 51b5bc1e7ce4254fe48490f7a1640dde
Markle root:	20e56ffedb72d4d729d668037acla187 0ef937e4ae9491eb7ac7c69f77924407
Czas wykopania:	2017-11-26 13:43:59
Trudność:	1,347,001,430,558.57
Transakcje	

Rysunek 2.1.: Przykładowy łańcuch 3 bloków sieci Bitcoin.

2.3. Struktura i zawartość bloku

Każdy blok ma ścisłe określona strukturę, która zawsze jest taka sama i składa się z:

- rozmiaru bloku,
- nagłówka bloku,
- licznika transakcji,
- transakcji.

Pierwsze trzy elementy mają stały lub ograniczony rozmiar, natomiast rozmiar potrzebny na zapis transakcji w bloku zależy od ich ilości i jest zmienny. Poniżej przedstawiono bardziej szczegółowo każdy z elementów bloku.

Rozmiar bloku zapisany jest jako podstawowa informacja na samym jego początku, co pozwala to na odczytanie odpowiedniej ilości danych.

Nagłówek bloku ma zawsze rozmiar 80 bajtów i zawiera informacje dotyczące:

- wersji protokołu sieci użytej do wygenerowania bloku - 4 bajty;
- bloku poprzedniego, tzn. jego hash - 32 bajty;
- podsumowania transakcji reprezentowanej przy pomocy drzewa Merkle - 32 bajty;
- szacowany czas wykopania bloku - 4 bajty;
- trudność wykopania bloku za pomocą określonego algorytmu proof-of-work - 4 bajty;
- licznik potrzebny dla algorytmu proof-of-work kopania bloku - 4 bajty.

Licznik transakcji wskazuje na ilość transakcji zapisanych w rejestrze bloku i zajmuje od 1 do 9 bajtów.

Ostatnią, ale najważniejszą częścią bloku są transakcje. Zawartość pojedynczej transakcji została opisana w osobnym podrozdziale 2.4..

Ilustracja 2.2. przedstawia rzeczywiste dane jednego z bloków. Dane te zaczerpnięto ze strony blockchain.info pozwalającej na eksplorowanie Bitcoinowego Blockchaina[5].

Hash bloku	0000000000000000a7b47a1e58e456 fd54ae5a30cb92a35ca3e5acee065287
Wysokość bloku	496201
Rozmiar:	1071.607 kB
Liczba transakcji:	2032
Nagłówek bloku	
Wersja:	0x20000000
Hash poprzedniego bloku:	0000000000000000c6dd215947b569 fa06de2cb856dec643daf5a7e8efc72e
Markle root:	b96da6d09865e36e4862b5a612fb1893 275d889989feb5a7a217503fd84019e3
Czas wykopania:	2017-11-26 13:51:02
Trudność:	1,347,001,430,558.57
Transakcje	

Rysunek 2.2.: Przykładowa zawartość jednego bloku sieci Bitcoin.

2.4. Transakcje i opłaty

Jednym z najważniejszych elementów w sieci Bitcoin są transakcje, które pozwalają na przekazywanie środków pomiędzy klientami sieci. Jak wspomniano w rozdziale 2.3. transakcje są zapisywane w blokach, w łańcuchu bloków. łańcuch ten jest publicznie dostępny, co implikuje właściwość transakcji jako publicznych.

Każda nowo powstała transakcja rozwijana jest w sieci, gdzie jest weryfikowana przez węzły, a w konsekwencji dodawana do nowo tworzonego bloku, który zostaje dołączony do blockchaina. Proces ten jest bardzo pracochłonny i wymaga rozesłania informacji o nowej transakcji w całej sieci. Sieć Bitcoin oparta jest na modelu komunikacji typu P2P (peer-to-peer). Oznacza to, że każdy z węzłów połączony jest z paroma innymi uczestnikami sieci, a z kolei Ci uczestnicy połączeni są z paroma kolejnymi uczestnikami. Powoduje to rozprzestrzenianie się informacji w sieci wykładowiczo. Każdy z węzłów sieci po otrzymaniu transakcji weryfikuje ją pod kątem poprawności. W przypadku wykrycia niepoprawnej transakcji jest ona natychmiastowo odrzucona i przestaje być dalej rozpoznawczana. Komunikacja w sieci Bitcoin jest synchroniczna, co oznacza, że węzeł, który rozgłosił transakcję dostaje informację zwrotną od węzłów z nim połączonymi o poprawności stworzonej transakcji. Kryptografia asymetryczna pozwala na anonimizację węzłów w sieci. Nie muszą się one znać, ani sobie ufać - wystarczy, że transakcja jest prawidłowo podpisana, a będzie przekazana dalej. Struktura transakcji oraz proces przekazywania bitcoinów przy pomocy transakcji są skonstruowane w sposób w pełni zabezpieczający środki. Przekaz nie potrzebuje żadnych dodatkowych zabezpieczeń takich jak, np. szyfrowanie łącza. Może być rozpowszechniany bez jakiegokolwiek szyfrowania, co w porównaniu z walutami fiduciarnymi jest ogromnym zyskiem[4].

Struktura transakcji składa się z elementów zaprezentowanych w tabeli 2.1., jednakże najważniejsze z nich to:

- adresy kluczy publicznych (wejścia do transakcji), reprezentujące źródło środków potrzebnych na pokrycie deklarowanej sumy przekazu;
- adresy kluczy publicznych (wyjścia z transakcji), reprezentujące cel przekazu.

Przekazywane bitcoiny nie są fizycznie dostępne na adresach wejściowych, a jedynie są zablokowane kluczem znanym tylko właścicielom tych adresów. Po podpisaniu transakcji środki z adresów wejściowych są zablokowane na rzecz adresów wyjściowych i tylko właściciele tych adresów mogą uwierzytelnić kolejne przekazy, używając ich jako adresów wejściowych.

W celu określenia aktualnie dostępnych środków na adresie wejściowym zliczane są *niewydatkowane bloki wynikowe transakcji* lub *niewydane wyjścia transakcyjne* zwane UTXO (ang. Unspend Transaction Output). UTXO są to niepodzielne oraz jeszcze nie wydane kawałki bitcoinów. Bitcoiny dzielą się na mniejsze kawałki zwane tzw. Satoshi i UTXO mogą składać się z ich wielokrotności.

W momencie potwierdzania transakcji weryfikowane są wszystkie UTXO przypisane do adresów wejściowych, a następnie po wpisaniu jej do łańcucha bloków rejestrowany jest nowy zablokowany UTXO przypisany do adresu wyjściowego. Jak pisano wcześniej, UTXO może zostać odblokowany jedynie przez jego właściciela. W sieci Bitcoin nie istnieje jeden główny bilans wszystkich adresów. Podczas realizacji każdej z transakcji sprawdzany jest cały łańcuch bloków i sumowane są wszystkie UTXO należące do określonego klucza publicznego. Właściciel UTXO podczas jego użycia musi wydać całą zablokowaną sumę Satoshi, dlatego też chęć przelania tylko części środków na UTXO

Tabela. 2.1.: Struktura transakcji.

Element	Rozmiar (bajty)	Opis
wersja protokołu	4	określa użytkę wersję protokołu użytą do stworzenia transakcji
ilość adresów wejściowych	1-9	ilość adresów reprezentujących źródło przekazywanych środków
adresy wejściowe	zmienna	adresy reprezentujące źródło przekazywanych środków
ilość adresów wyjściowych	1-9	ilość adresów reprezentujących cel przekazywanych środków
adresy wyjściowe	zmienna	adresy reprezentujące cel przekazywanych środków
czas blokady	4	określa czas, kiedy transakcja może zostać wykonana, zazwyczaj ustawiana na 0 w celu jak najszybszego podpisania

powoduje stworzenie wielu nowych UTXO. Część z nich wskazywać będzie na adresy wyjściowe, a część jako reszta z UTXO, będzie wskazywać na adres nadawcy transakcji. Bardzo często zdarza się sytuacja, w której właściciel wielu UTXO chce przelać środki na jeden adres. W takim przypadku dozwolone jest połączenie tych jednostek w celu osiągnięcia oczekiwanej sumy bitcoinów i przekazania jej na adres odbiorcy.

Każdy z bloków zawiera jedną transakcję stworzoną przez jednego uczestnika sieci - zwycięskiego kopacza, który otrzymuje profity z tytułu podpisania bloku. Jest to zawsze pierwsza transakcja tzw. *coinbase*. Proces ten został opisany w podrozdziale 2.5..

Wszystkie adresy wejściowe można traktować jako punktory do transakcji wyjściowych, zrealizowanych oraz niezrealizowanych (z nich powstają UTXO). Zrealizowane transakcje zawierają również odblokowane skrypty, co pozwala na przekazanie ich dalej. Należy podkreślić, że adresy wejściowe nie niosą informacji o ilości posiadanych środków, a jedynie informację o miejscu, w którym można się o tym dowiedzieć. Poniżej, w tabeli 2.2., przedstawiono strukturę wyjścia transakcji.

Tabela. 2.2.: Struktura wyjścia transakcji.

Element	Rozmiar (bajty)	Opis
kwota	8	ilość bitcoinów przedstawionych w Satoshi
rozmiar skryptu blokującego	1-9	długość skryptu blokującego przedstawiona w bajtach
skrypt blokujący	zmienna	skrypt określający warunki do spełnienia w celu odblokowania wyjścia

Transakcje zazwyczaj tworzone są za pomocą klienta Bitcoin, co jest wygodnym i bezpiecznym sposobem przekazywania środków. Podczas tworzenia transakcji aplikacja ta zapewnia, że w transakcji znajdują się wszystkie wymagane informacje. W tabeli 2.3. przedstawiono strukturę tworzenia transakcji. Pozwala ona spełnić wszystkie warunki określone

w wyjściu transakcji, której strukturę przedstawiono w tabeli 2.2..

Tabela. 2.3.: Struktura wprowadzania transakcji.

Element	Rozmiar (bajty)	Opis
hash transakcji	32	punktor do transakcji zawierającej UTXO do wydania
indeks wyjścia	4	numer indeksu UTXO do wydania
rozmiar skryptu odblokowującego	1-9	długość skryptu odblokowującego wyrażona w bajtach
skrypt odblokowujący	zmienna	skrypt spełniający wszystkie wymagane warunki skryptu blokującego UTXO
numer sekwencji	4	aktualnie nieużywana funkcjonalność wymiany Tx

Z każdą przeprowadzaną transakcją wiążą się koszty operacyjne. W przypadku Bitcoina jest to opłata dla górników mająca zachętać do kopania. Wielkość opłaty nie zależy od ilość przekazywanych bitcoinów, a od wielkości stworzonej transakcji. Oznacza to, że osoba, która posiada wiele UTXO i chce je skonsumować w jednej transakcji musi zapłacić wyższą opłatę. Wynika to z faktu, że górnik w celu weryfikacji posiadanych UTXO musi wykonać dużo więcej pracy, niż gdyby transfer odbywał się z jednego UTXO. Górnicy bardzo często ustalają kryteria i tworzą priorytety transakcji, które chcą przetwarzać. W niektórych przypadkach może się okazać, że przez zbyt niską opłatę żaden z górników nie będzie podejmował się sprawdzenia i podpisania konkretnej transakcji. Jeżeli transakcja nie zostanie podpisana przez górnika nie znajdzie się w wytwarzanym bloku. Transakcja ta będzie oczekiwana w sieci do czasu przetworzenia przez któregoś z kopaczy.

2.5. Dołączania kolejnego bloku z transakcjami

Wszystkie transakcje rozsypane po sieci trafiają do górników, którzy gromadzą je w celu zbudowania nowego bloku. Transakcje te są jedynie kandydatami do znalezienia się w bloku, a to które z nich znajdą się w zbudowanym bloku zależy od priorytetów górnika.

Przypuszczając sytuację, w której górnik otrzymuje wiadomość o powstaniu nowego bloku X , nad który też pracował, akceptuje otrzymane rozwiązanie i podejmuje się wyzwania skonstruowania bloku $X + 1$. Porównuje transakcje z bloku X z otrzymanymi transakcjami. Redukuje ich ilość, tak by nie próbować umieszczać tych samych transakcji w kolejnym bloku i próbuje wytworzyć kolejny blok. Sprawdza wszystkie transakcje przeznaczone do umieszczenia w bloku pod względem poprawności, a następnie generuje transakcję zwaną *coinbase*.

Transakcja *coinbase* zawiera w wejściu do transakcji nowe bitcoin'y, które są nagrodą za wygenerowany blok oraz opłaty wniesione przez zleceniodawców transakcji. Ilość nowych bitcoinów zależy od ilości istniejących bloków w sieci. Nagroda zmniejsza się wraz z ich ilością. Na początku istnienia sieci wynosiła ona 50 BTC i zmniejsza się o połowę co 210 tys. bloków. Aktualnie wynosi 12.5 BTC. Każdy górnik po otrzymaniu bloku sprawdza, czy górnik, który wytworzył blok nie oszukał generując transakcję *coinbase*. W przypadku oszustwa taki blok zostaje odrzucony, a praca i koszty poniesione przez górnika zostają

niepokryte. W przeciwej sytuacji, kiedy to górnik poprawnie wykonał swoją pracę, bitcoiny zwarte w tej transakcji trafiają na adres w niej określony. Zazwyczaj tym adresem jest adres górnika.

Po dodaniu transakcji *coinbase* tworzony jest nagłówek bloku, który zawiera wszystkie informacje opisane w rozdziale 2.3.. W nagłówku dostępny jest hash bloku poprzedniego oraz hash drzewa Merkle, który pozwala na szybką weryfikację *zaksięgowania* transakcji w bloku. Algorytm tworzenia drzewa Merkle wymaga istnienia parzystej liczby transakcji w bloku. W przypadku zawarcia nieparzystej ilości transakcji w bloku, ostatnia z nich jest duplikowana na potrzeby stworzenia drzewa.

Kolejnym krokiem tworzenia bloku jest podpisanie go poprzez znalezienie rozwiązania określonego algorytmu *proof-of-work*. Algorytm ten polega na wielokrotnym mieszaniu nagłówków bloków oraz jednej zmiennej do momentu uzyskania hash'a o określonych właściwościach. Hash'e generowane są przy pomocy algorytmu kryptograficznego wspomnianego w rozdziale 2.2. zwanego funkcją skrótu SHA256. Warunki do spełnienia mogą się zmieniać wraz ze zmianami w sieci. Przykładową własnością docelową poszukiwanego hash'a może być zależność polegająca na tym, aby hash'a ten był mniejszy niż ustalony próg celu, tzn. musi być mniejszy niż określony hash.

Po odnalezieniu hash'a bloku $X + 1$, zostaje on rozsyłany po sieci w celu ogłoszenia jego wygenerowania. Klienci sieci otrzymują tę informację i weryfikują poprawność nowego bloku. Oznacza to, że każdy węzeł sieci musi przeprowadzić wiele testów weryfikacyjnych zanim wyśle blok do połączonych węzłów. Po zaakceptowaniu bloku przez węzeł, zostaje on dodany do lokalnego łańcucha i wysłany dalej. W przypadku kiedy węzeł otrzyma blok Y od innego węzła wskazujący na ten sam poprzedni blok(X), co blok $X + 1$, blok Y zostaje odrzucony.

W tak skonstruowanym systemie może się wydarzyć sytuacja, w której dwóch górników jednocześnie zacznie rozsyłać poprawny blok po sieci. Oznaczać to będzie, że węzły zatwierdzą blok *górnika 1* oraz blok *górnika 2* i powstanie rozwidlenie sieci. W takich sytuacjach sieć oczekuje na wyprodukowanie kolejnego bloku, który używa jednego z bloków wyprodukowanych przez *górnika 1* lub *górnika 2* w celu odnalezienia globalnego consensusu. Zachowanie sieci w tym przypadku zależy od wskazania, w tym generowanym bloku, hash'a bloku poprzedniego. Dłuższy łańcuch bloków jest przez sieć traktowany jako ważniejszy i nieużyty blok w tym łańcuchu zostaje przegłosowany, a konsekwencji staje się blokiem sierocym. łańcuchy bloków stworzone przez węzły na podstawie przegłosowanego bloku zostają skorygowane i zastosowany zostaje dłuższy łańcuch bloków.

2.6. Podsumowanie

Konkludując, Blockchain jest strukturą zbudowaną z bloków wstępnie połączonych. Każdy blok z łańcucha zawiera informacje o zrealizowanych transakcjach oraz o hash'u bloku poprzedniego. Transakcje składają się z wejść do transakcji oraz wyjść z transakcji. Wejścia są punktami do niewydanych wyjść transakcji z poprzednich bloków, a wyjścia transakcji wskazują na adresy uczestników sieci, do których mają trafić bitcoiny. Bloki generowane są z transakcji przez górników, którzy walczą w wyścigu o nagrodę w postaci nowych bitcoinów oraz opłat transakcyjnych. Nowo wygenerowane bloki dołączane są do łańcucha bloków w procesie znajdowania globalnego consensusu w sieci.

Rozdział 3.

Przegląd metod analiz sieci złożonych (też temporalnych) oraz analiz blockchaina

Czym jest analiza sieci, jakie były analizy, metody analiz grafów. Nie tylko analizuje się sieci statyczne, ale temporalne i wielowarstwowe. Jak analizować blockchain, co zostało zrobione w kontekście jego analizy?

Rozdział 4.

Analiza blockchaina Bitcoin

4.1. Wprowadzenie

W niniejszym rozdziale przedstawiono część eksperymentalną pracy. Sformułowano cel badawczy oraz określono zakres przeprowadzanych analiz. Przybliżono również sposób tworzenia sieci oraz określono związek pomiędzy węzłami sieci. Następnie opisano przeprowadzone analizy wraz z wnioskami.

4.2. Plan badań

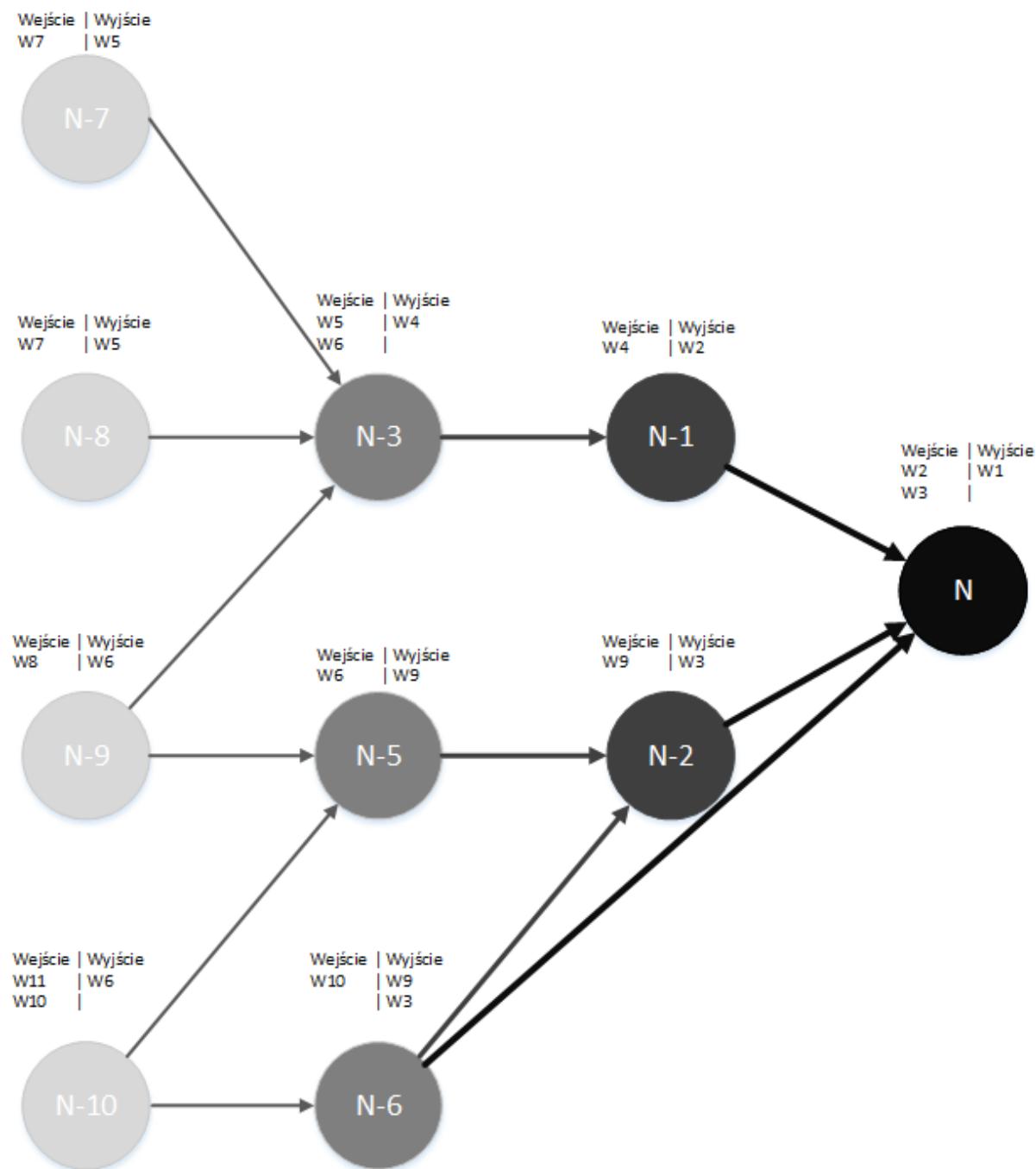
Celem przeprowadzanego eksperymentu jest analiza rejestru transakcji w sieci Bitcoin. Podjęto próbę określania cech sieci, które pozwalają na stworzenie charakterystyki jej rozwoju. Przy pomocy metod analizy sieci złożonych zbadano trendy zmian zachodzących w sieci oraz zaobserwowano zdarzenia nietypowe, odbiegające od wyznaczonego trendu.

Na potrzeby realizacji eksperymentu wyznaczono po dziesięć węzłów startowych w sieci w dziesięciu okresach. Wybrano bloki z łańcucha bloków dołączone jako ostatnie w dziesięciu kwartałach w okresie od *2015-03-31* do *2017-06-30*. Następnie z każdego z bloków, dla każdego okresu, wybrano po dziesięć transakcji, które stały się węzłami startowymi dla badanej sieci. Oznacza to, że analizę przeprowadzono na stu próbach po sto tysięcy węzłów każda. Ilość oraz wielkość prób zmniejsza ryzyko natrafienia na próbę nierepresentatywną.

Każdą sieć, która stanowi próbę do badań, stworzono zaczynając od węzła startowego wstecz. Każdy węzeł reprezentuje transakcję wskazującą na jej adres. Jak opisano w podrozdziale 2.4. transakcje zawierają adresy wejściowe oraz adresy wyjściowe. W podrozdziale 2.4. podkreślono również fakt, że adresy wejściowe są jedynie referencjami na adresy wyjściowe, co implikuje właściwość związków pomiędzy transakcjami. W celu łatwiejszego zrozumienia sposobu budowania sieci w tabeli 4.1. przedstawiono dwie połączone transakcje, na przykładzie których zauważać można, że transakcja $N - 1$ zawiera adres wyjściowy identyczny jak jeden z adresów wejściowych transakcji N . Oznacza to, że transakcje te są połączone i stanowią dwa węzły sieci połączone krawędzią. Wykorzystując tę zależność każdą próbę stworzono poprzez znalezienie stu tysięcy połączeń pomiędzy transakcjami na podstawie adresów wejściowych i adresów wyjściowych zaczynając od początkowego adresu transakcji. Na ilustracji 4.1. przedstawiono przykładowy generyczny wycinek każdej z stworzonych sieci zaczynając od węzła N , natomiast na ilustracji 4.2. przedstawiono rzeczywisty graf jednej z próbek.

Tabela. 4.1.: Związek między transakcjami.

Transakcja N-1		Transakcja N	
Adresy wejściowe	Adresy wyjściowe	Adresy wejściowe	Adresy wyjściowe
W4	W2	W2 W3	W1



Rysunek 4.1.: Przykładowy wycinek próbki sieci.



Rysunek 4.2.: Wizualizacja grafu składającego się ze stu tysięcy połączeń.

Przeprowadzone badania podzielono na dwie grupy pod względem reprezentowanych cech sieci. Pierwsza grupa reprezentuje analizę sieci przy pomocy algorytmów generycznych dla sieci złożonych. Algorytmy te nie są związane z problematyką fachową sieci, jednakże pozwalają na zbadanie podstawowych własności każdej z sieci. W ramach analizy dla każdej próbki zbadano:

- średnicę sieci,
- średnią długość ścieżek,
- średni stopień węzłów,
- średnią centralność węzłów.

Druga grupa przeprowadzonych badań pozwoliła na analizę sieci pod kątem jej specyficznych własności, które są bardzo mocno związane z jej problematyką. W ramach analizy dla każdej próbki zbadano:

- średnią wartość transakcji,
- liczbę bloków potrzebnych do stworzenia próbki,
- średnią różnicę czasów kolejnych transakcji,
- różnice czasu granicznych transakcji.

Badania z grupy pierwszej przeprowadzono przy pomocy biblioteki *Igraph*, natomiast algorytmy potrzebne do przeprowadzenia badań z grupy drugiej zostały przygotowane samodzielnie.

4.3. Eksperyment

W ramach przeprowadzonych badań dla każdej z metryk stworzono mapę cieplną przedstawiającą wartości poszczególnych właściwości sieci. Na wykresie zaprezentowano również średnią wartość dla każdego z okresu wraz z odchyleniem standardowym, z którego stworzono regresję liniową. Na podstawie tych danych przeprowadzono analizę oraz określono trend zmian w sieci. Pierwsze cztery analizy przeprowadzono dla właściwości sieci nie bezpośrednio powiązanych z typem sieci. Kolejne cztery przeprowadzone analizy związane są wyłącznie z fachowością badanej sieci.

4.3.1. Badanie średnicy sieci

Średnica sieci jest długością

$$\max_{u,v} d(u, v) \quad (4.1.)$$

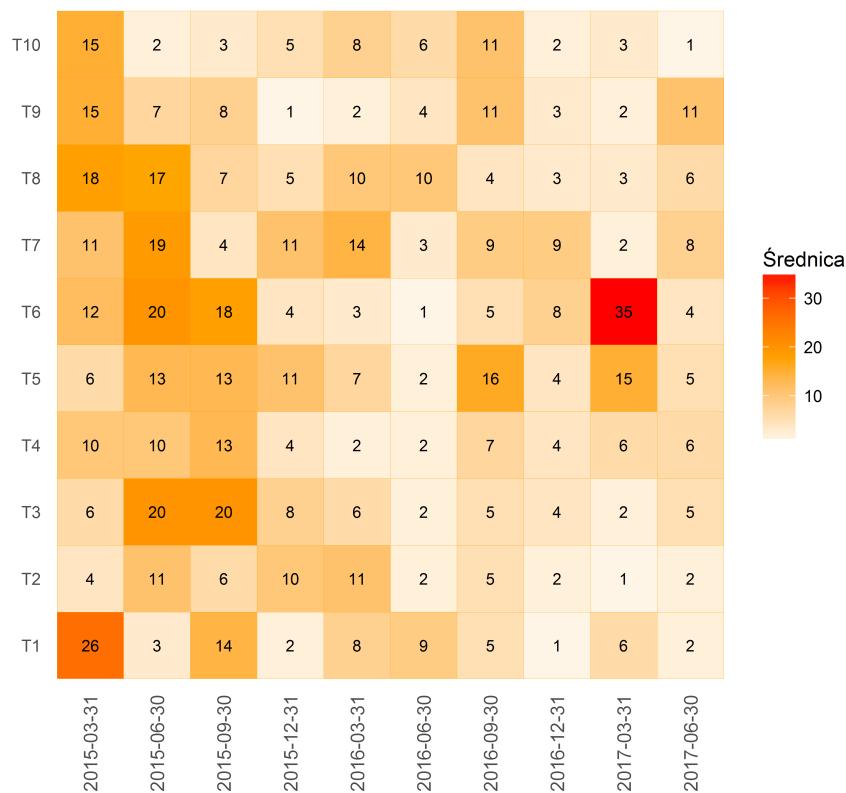
najdłuższej ścieżki znalezionej wśród najkrótszych ścieżek pomiędzy dwoma którymkolwiek węzłami sieci (u, v) , gdzie $d(u, v)$ jest długością grafu. Długość grafu jest minimalną długością ścieżki (stworzonej z węzłów) potrzebnej do połączenia dwóch określonych węzłów. Inaczej mówiąc średnica grafu jest maksymalną ilością węzłów, jakie trzeba pokonać by przejść z jednego węzła do drugiego.

Dla badanej sieci oznacza to maksymalną ścieżkę połączonych transakcji, a co za tym idzie, maksymalną krotność przekazywania pomiędzy adresami publicznymi bitcoinów.

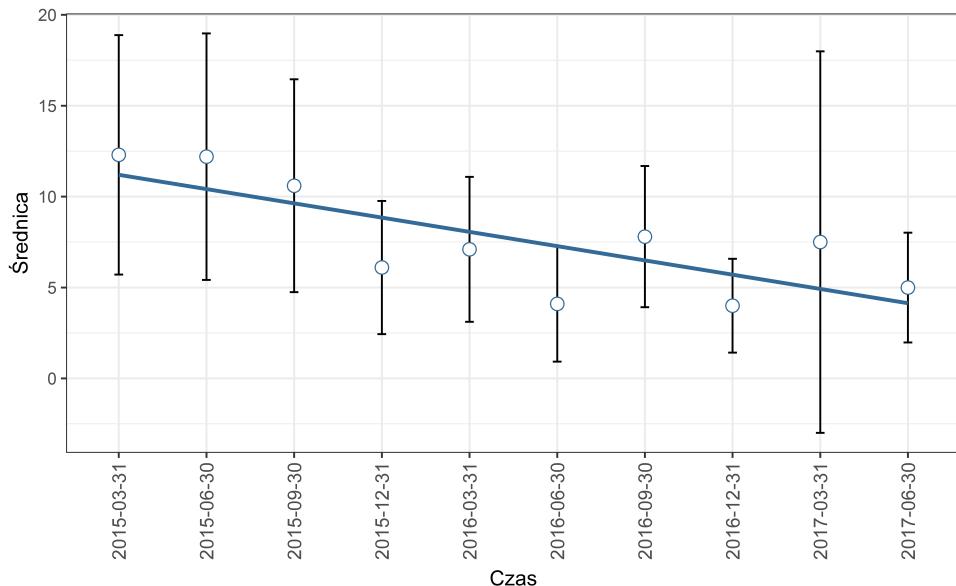
Na podstawie wartości długości ścieżki można wnioskować o własnościach sieci, takich jak na przykład jej gęstości.

Na rysunku 4.3. przedstawiono mapę ciepła średnicy sieci dla każdej z próbek. Na jej podstawie można wnioskować, że dla większości przypadków próbki w poszczególnych okresach są zbliżone. Zaobserwować można również, że w każdym z okresów istnieją pojedyncze sieci mocno odbiegające od reszty. Jedną z tych anomalii jest średnica sieci dla próbki zaczynającej się od transakcji szóstej, która została zrealizowana *2017-03-31*. Istnieje prawdopodobieństwo, że jest to transakcja realizowana w ramach wielokrotnego przesyłania bitcoinów przez jednego właściciela pomiędzy własnymi adresami w celu próby ukrycia ich źródła.

Na wykresie 4.4. przedstawiono regresję liniową na podstawie średniej średnicy dla okresu z odchyleniem standardowym. Z wykresu wynika, że gęstość sieci rośnie w czasie, o czym świadczy trend spadkowy średnicy sieci. Wzrost gęstości oznacza, że w sieci występuje coraz więcej krawędzi w stosunku do ilości węzłów. Dla sieci Bitcoin wzrost gęstości sieci oznacza wzrost ilość transakcji realizowanych na pojedyncze adresy publiczne, w krótszym czasie. Można zatem wnioskować, że bitcoinów coraz częściej gromadzone są na pojedynczych adresach. Odchylenia standardowe widoczne na 4.4. wynikają z pojedynczych próbek mocno odbiegających od średniej. Odrzucenie próbek (najwyżej oraz najniżej wartościowanych) z każdego okresu zniewelowałoby znacząco rozpiętość odchylenia standardowego. Przykładem może być wcześniej omawiana próbka zaczynająca się od transakcji *T6* z okresu *2017-03-31*.



Rysunek 4.3.: Mapa cieplna średnicy sieci dla 10 prób w 10 okresach.



Rysunek 4.4.: Regresja liniowa średniej średnicy sieci dla 10 okresów z odchyleniem standardowym.

4.3.2. Badanie średniej długość ścieżki

Średnia długość ścieżki w sieci jest średnią wartością

$$\sum_{i \neq j} d(u_i, v_j) \frac{1}{n(n-1)} \quad (4.2.)$$

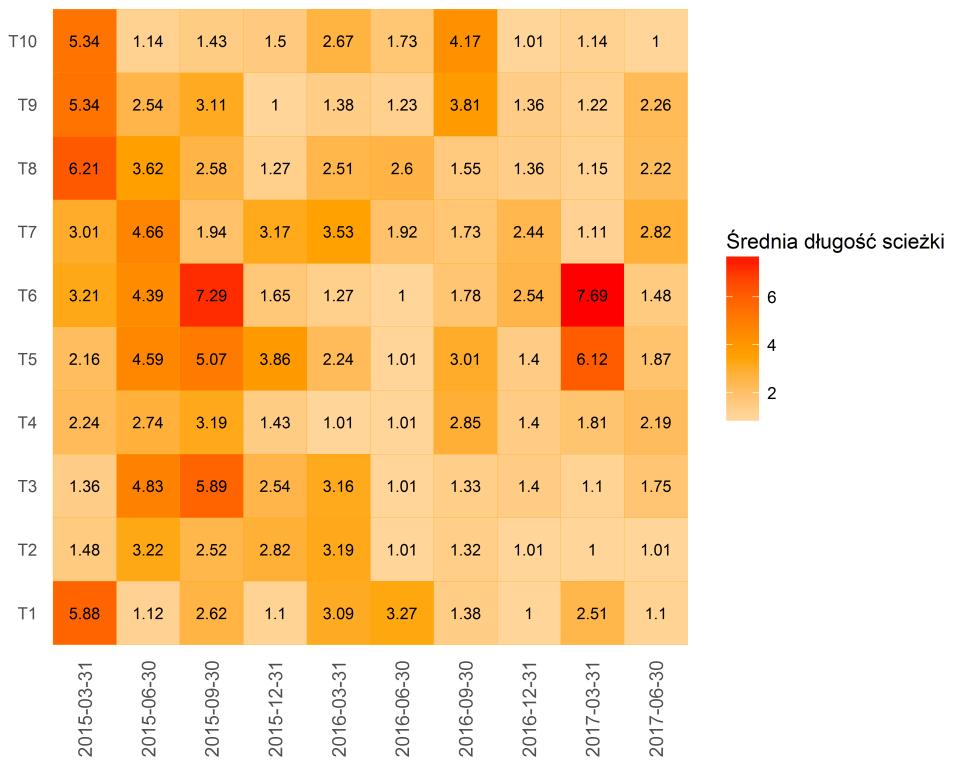
wszystkich możliwych par węzłów (u_i, v_j) w grafie, gdzie $d(u_i, v_j)$ jest najkrótszą długością ścieżki pomiędzy węzłami (u_i, v_j) , a n jest ilością węzłów. Długość ścieżki została zdefiniowana w 4.3.1..

Sieć Bitcoin jest siecią skierowaną, a każda z transakcji posiada niepowtarzalny adres. Eliminuje to możliwość pojawiения się cyklu w grafie, dlatego też ilość możliwych ścieżek w analizowanej sieci jest ograniczona. W badanej sieci średnia długość ścieżki oznacza średnią ilość transakcji, przy pomocy których przekazywano kolejno bitcoiny.

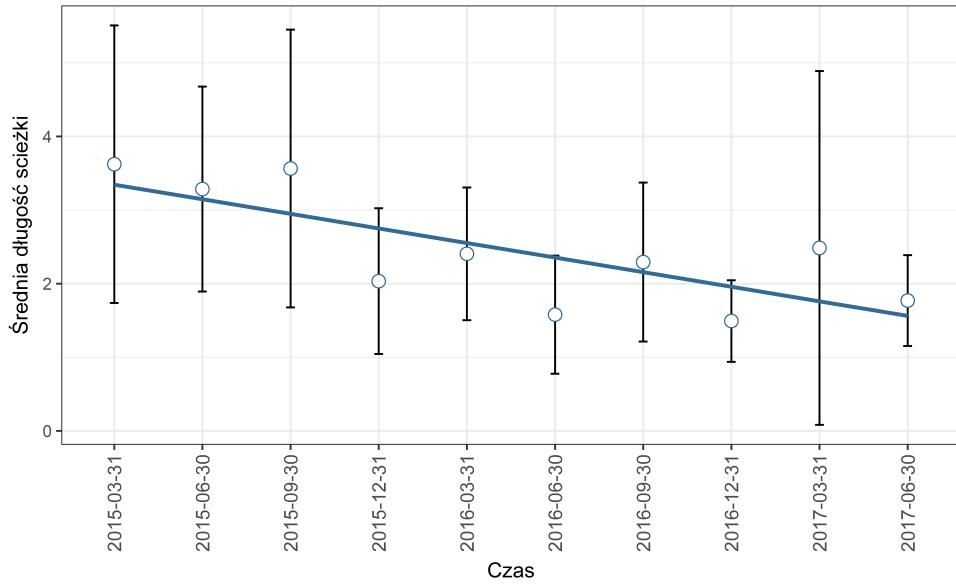
Na podstawie mapy cieplnej 4.5. reprezentująccej średnią długość ścieżki dla prób zaobserwowano pojedyncze przypadki odchyleń wartości badanej cechy. W przypadku transakcji szóstej zrealizowanej 2017-03-31, przyczyna prawdopodobnie jest analogiczna do przypadku analizy średnicy sieci, tzn. wielokrotne przesyłanie środków jednej osoby przez wiele adresów publicznych w sieci.

Na wykresie 4.6. reprezentującym regresję liniową średniej wartości długości sieci z okresów zaobserwowano trend spadkowy. W przypadku badanych prób średnia długość sieci zmienia się analogicznie do jej średnicy. Determinuje to zmniejszającą się średnią liczbę kolejnych transakcji w badanych próbach. Może to oznaczać coraz większą ilość transakcji rejestrowanych w sieci. Potwierdza to wzrastającą gęstość sieci.

Można przypuścić, że uszczegółowienie informacji dotyczących przypadków znaczających odchyleń standardowych poprzez wykonanie dodatkowych badań pozwoliłoby jednoznacznie stwierdzić ich przyczynę. Jednakże istnieje prawdopodobieństwo, że właściwości sieci nie są stałe w jednym okresie czasu, a zmieniają się dynamicznie w zależności od aktywności poszczególnych uczestników sieci.



Rysunek 4.5.: Mapa cieplna średniej długości scieżki w sieci dla 10 prób w 10 okresach.



Rysunek 4.6.: Regresja liniowa średniej długości scieżki sieci dla 10 okresów z odchyleniem standardowym.

4.3.3. Badanie średniego stopnia węzła

Stopień węzła v grafu oznacza liczbę krawędzi w grafie, które bezpośrednio dotykają węzła v . W grafach skierowanych wyróżnia się wejściowy stopień węzła (ilość krawędzi skierowanych w stronę węzła) oraz wyjściowy stopień węzła (ilość krawędzi skierowanych od strony węzła). Suma wejściowego stopnia węzła oraz wyjściowego stopnia węzła daje wynikowy stopień węzła. Średni stopień węzła obliczany jest jako

$$\frac{\sum_i^n \deg(v_i)}{n} \quad (4.3.)$$

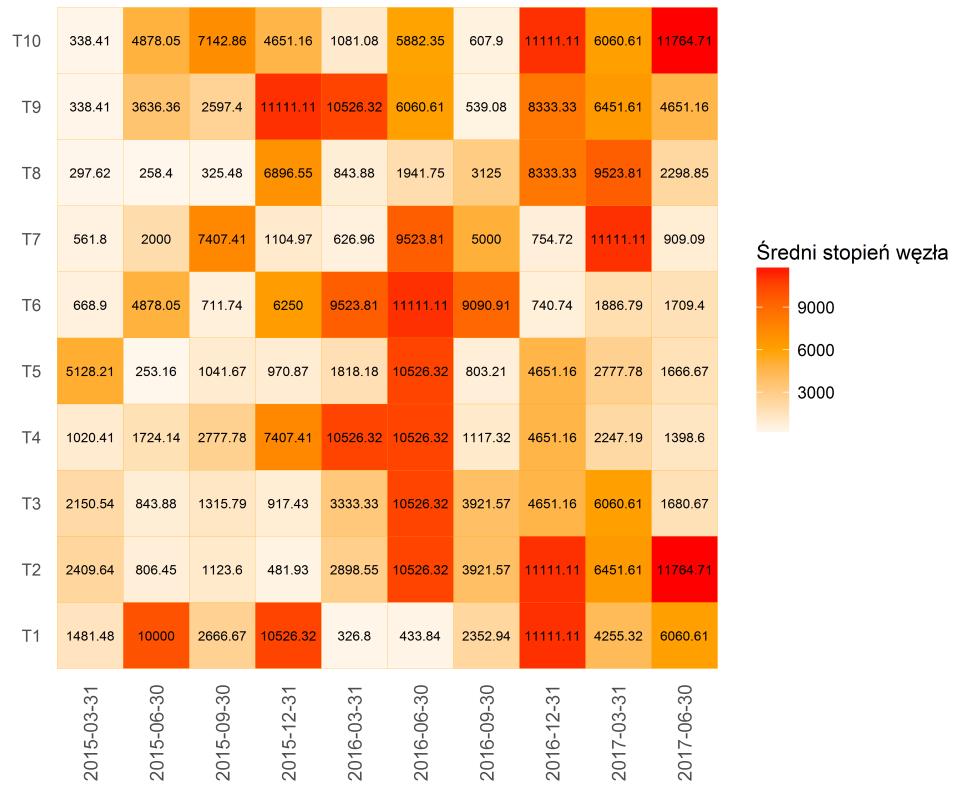
gdzie $\deg(v_i)$ to stopień i-tego węzła, a n to ilość węzłów. W celu zminimalizowania ryzyka popełnienia błędu, podczas interpretacji obliczanej średniej, badana próba powinna mieć rozkład normalny. W przeciwnym przypadku należy przedstawić dodatkowe źródło informacji o próbie, które zobrazuje częstotliwość występowania poszczególnych wartości, np. histogram.

W trakcie budowania sieci procesowanie każdej pojedynczej transakcji polega na sprawdzeniu jej adresów wyjściowych z adresami wejściowymi każdej transakcji do tej pory dołączonej do sieci. Wykrycie tego samego adresu w tych dwóch zbiorach adresów transakcji powoduje powstanie krawędzi, która je łączy. W przypadku badania stopnia węzła transakcji w łańcuchu transakcji sieci Bitcoin określana jest ilość transakcji, z którymi dana transakcja dzieli adresy w danej próbce. Średni stopień węzła w próbce oznacza średnią ilość transakcji, z którymi połączona jest każda transakcja w próbce.

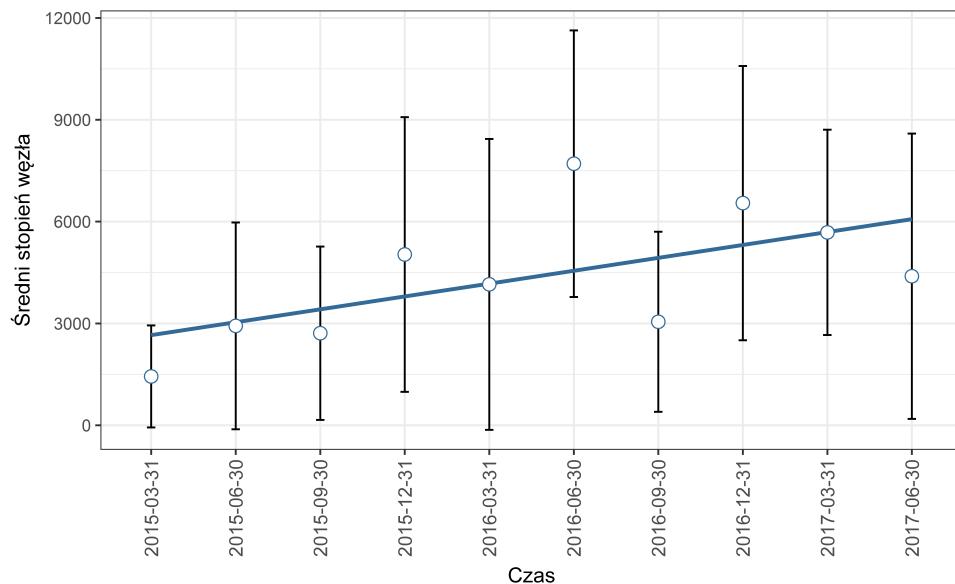
Na podstawie mapy cieplnej 4.7. badanej właściwości wywnioskowano, że średnia ilość transakcji, z którymi połączona jest każda transakcja jest bardzo zróżnicowana. Wynika to z różnego poziomu aktywności właścicieli adresów biorących udział w transakcjach w danych okresach czasu oraz ilości nowych adresów generowanych przez uczestników sieci. Prezentowane wyniki można interpretować na trzy różne sposoby. Wysoka wartość średniego stopnia węzła może świadczyć o wzmożonej działalność właściciela adresu jako odbiorcy, czy jako nadawcy zleceń. Niska wartość średniego stopnia węzła nie wyklucza dużej aktywności pojedynczych klientów sieci, a oznaczać może dużą ilość nowo generowanych adresów, na które przekazywane były środki. Trzecią prawdopodobną sytuacją jest mało dynamiczny fragment sieci, w którym utworzono niewiele nowych adresów, a środki przekazywane były kolejno pomiędzy ograniczoną ilością uczestników sieci.

Znaczące zróżnicowania badanej cechy zaobserwowano również na wykresie 4.8., który przedstawia regresję liniową stworzoną na podstawie wartości średnich z odchyleniami standardowymi dla okresu. Wykres ten wskazuje na tendencję do wzrostu analizowanej własności, jednakże z powodów wskazanych w akapicie pierwszym oraz dużych odchyleń standardowych, zaobserwowanych na prezentowanych wykresach, postanowiono przeprowadzić dodatkową analizę.

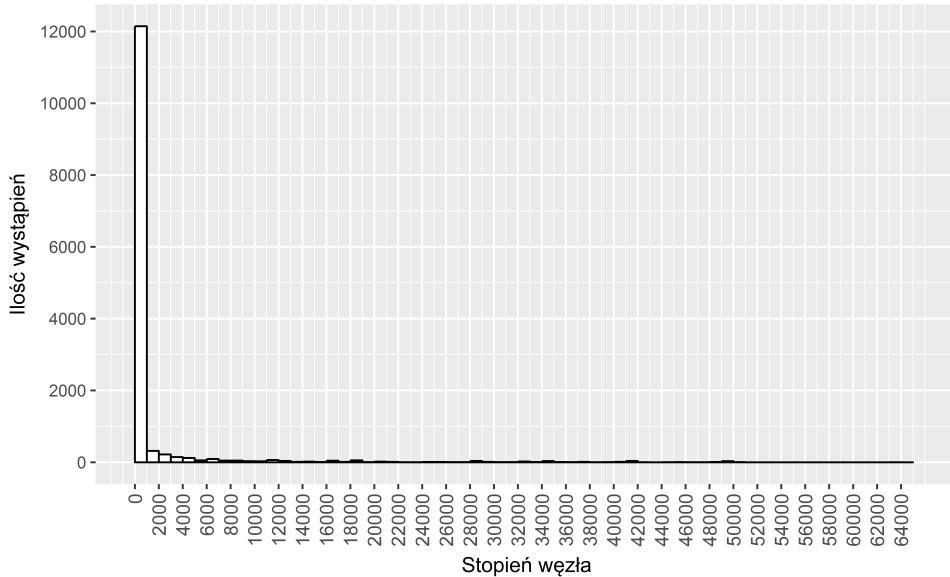
Na histogramie 4.9. przedstawiono ilość wystąpień poszczególnych wartości stopni węzłów dla wszystkich próbek razem. Zauważono zdecydowaną dominację wartości na poziomie do tysiąca wchodzących i wychodzących transakcji. Wcześniej zauważone odchylenia standardowe wynikają z anomalii sięgających do 50 tysięcy połączonych transakcji, z którymi dana transakcja dzieli adresy w danej próbce. Istnieje prawdopodobieństwo, że transakcje, których stopień węzła mocno odbiega od dominanty są jednymi z kluczowych węzłów w sieci.



Rysunek 4.7.: Mapa cieplna średniego stopnia węzła sieci dla 10 prób w 10 okresach.



Rysunek 4.8.: Regresja liniowa średniego stopnia węzła sieci dla 10 okresów z odchyleniem standardowym.



Rysunek 4.9.: Histogram średniego stopnia węzła łączny dla wszystkich próbek.

4.3.4. Badanie średniej centralności węzła

Centralność węzła v jest miarą bazującą na własności grafu polegającą na znalezieniu najkrótszych ścieżek w grafie. Centralność węzła opiera się na znalezieniu wszystkich możliwych najkrótszych ścieżek w grafie pomiędzy połączonymi węzłami, które następnie grupowane są na ścieżki przechodzące przez węzeł v . Średnia centralność węzła dana jest wzorem

$$\frac{\sum_i^n \sum_{s \neq v_i \neq t} \frac{\sigma_{st}(v_i)}{\sigma_{st}}}{n} \quad (4.4.)$$

gdzie σ_{st} jest liczbą wszystkich najkrótszych ścieżek przechodzących od węzła s do węzła t , $\sigma_{st}(v_i)$ jest liczbą wszystkich najkrótszych ścieżek przechodzących od węzła s do węzła t przez węzeł v_i , a n jest liczbą wszystkich węzłów w grafie.

W badanej sieci centralność pojedynczej transakcji stanowi informację o jej znaczeniu dla całej sieci. Taka transakcja może zawierać adresy wejściowe jak i wyjściowe, które są mniej lub bardziej znaczące dla istnienia sieci, jednakże można stwierdzić, że im mniej adresów użytych w transakcji, a większa centralność transakcji, tym większe ich znaczenie. Zbudowana sieć bazuje na nieokreślonej wielkości zbiorach adresów w transakcji, więc pozwala zaobserwować ich znaczenie tylko jako jednostki. Duża centralność pojedynczej transakcji oznacza, że przekazywane przy jej pomocy środki są źródłem dla wielu innych transakcji. Średnia centralność wszystkich transakcji w próbce wyznacza zatem średnią istotność każdej transakcji w próbce.

Na podstawie mapy cieplnej 4.10. stworzonej dla omawianej właściwości sieci wywnioskowano, przy założeniu zignorowania granicznych wyników, że w badanych okresach średnia centralność transakcji w próbach jest zbliżona. Zaobserwowano również, że dla większości badanych próbek w kolejnych okresach istotność transakcji w sieci spada. Potwierdza to wykres 4.11., który przedstawia trend spadkowy średniej centralności transakcji w sieci, ciągłe zagęszczanie sieci oraz wzrost ilości wykonywanych transakcji w czasie, co zostało już omówione w poprzednich analizach.

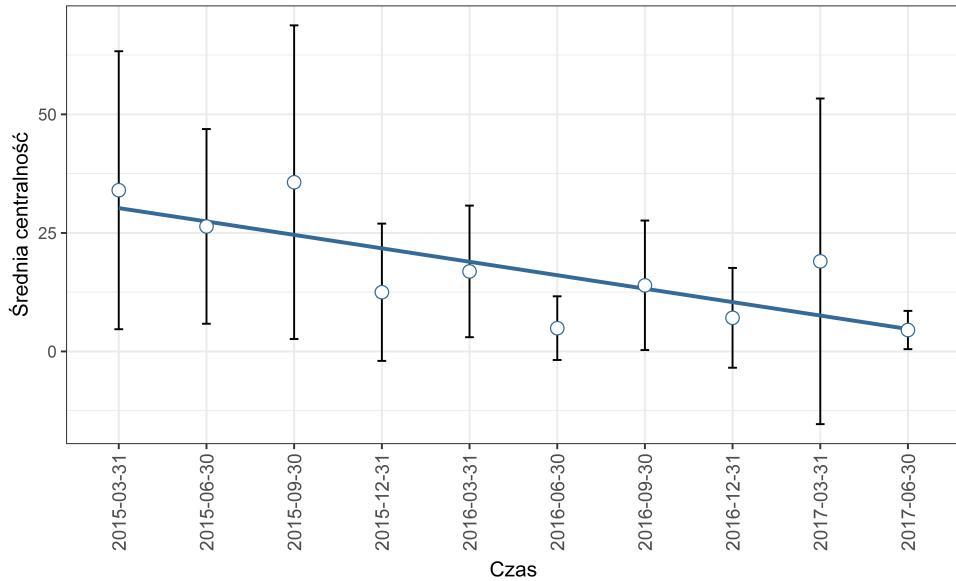
Na histogramie 4.12. prezentującym ilość wystąpień poszczególnych wartości centralności węzłów dla wszystkich próbek razem zaobserwowano zdecydowaną przewagę węzłów

o małej centralności. Determinuje to zdecydowany spadek znaczenia pojedynczej transakcji w sieci.

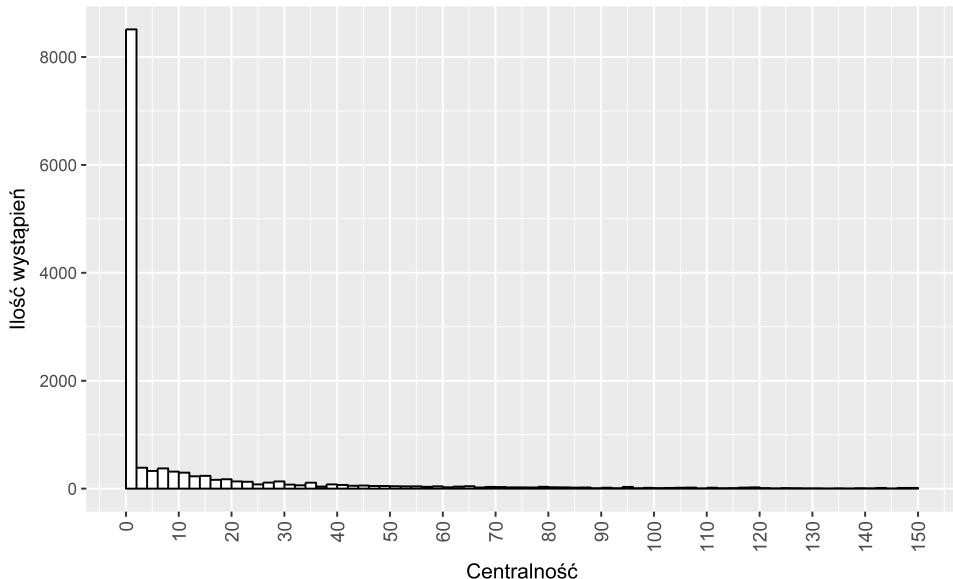
Dodatkowo zauważono widoczną zależność pomiędzy średnią centralnością węzłów, a średnim stopniem węzła. Im mniejsza centralność węzłów tym większy stopień węzła, co oznacza, że sieć wciąż ewoluje i zmienia swoją topologię w topologię siatki.



Rysunek 4.10.: Mapa cieplna średniej centralności węzłów w sieci dla 10 prób w 10 okresach.



Rysunek 4.11.: Regresja liniowa średniej centralności węzłów w sieci dla 10 okresów z odchyleniem standardowym.



Rysunek 4.12.: Histogram centralności węzłów w sieci łącznej dla wszystkich prób.

4.3.5. Badanie średniej wartości transakcji

Średnią wartości transakcji jest miarą bezpośrednio związaną z typem badanej sieci. Oznacza średnią wartości bitcoinów, która wiąże ze sobą węzły reprezentujące transakcje, tworząc graf ważony. Wartość bitcoinów pomiędzy dwoma węzłami stanowią wagę krawędzi przejścia pomiędzy nimi. Średnia wartość transakcji można zapisać jako

$$\frac{\sum_i^n \omega(e_i)}{n} \quad (4.5.)$$

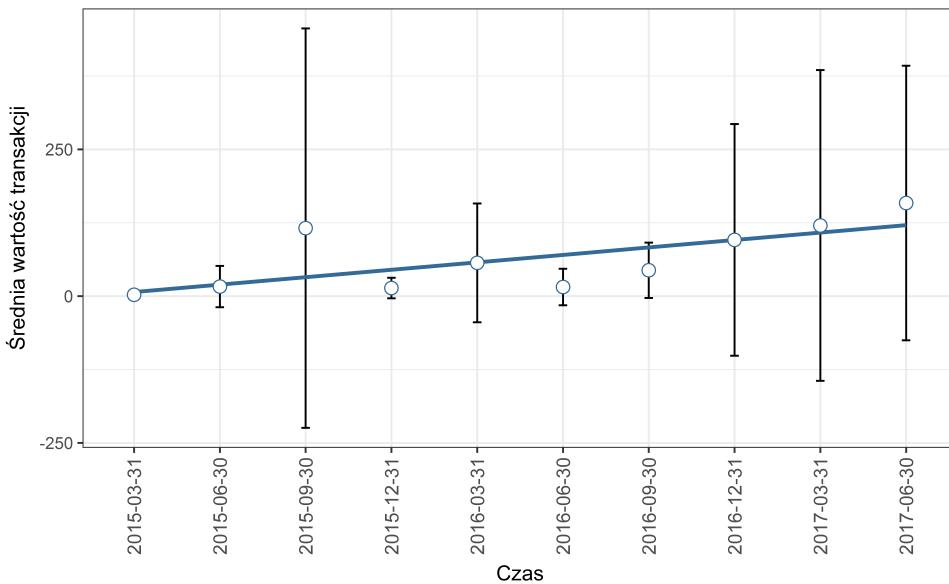
gdzie $\omega(e_i)$ oznacza wagę przejścia (ilość bitcoinów) przez krawędź łączącą dwa węzły (transakcje), a n oznacza ilość wszystkich krawędzi (połączeń pomiędzy transakcjami).

Przeprowadzone badanie wykazało, że w różnych okresach wartość mierzona pojedynczych próbek znaczaco odbiega od pozostałych, co może wpływać na ogólny wynik badań. Na mapie cieplnej 4.13. przedstawiono rezultat tych badań. Możliwe jest, że odbiegające próbki stanowią przekazy środków pomiędzy adresami giełd lub giełdą, a innymi uczestnikami sieci. Wskazują na to znaczaco wyższe sumy łączące transakcje. Chwilowe pominięcie tych wyników pozwala na stwierdzenie, że ilości przekazywanych środków we wszystkich okresach są bardzo porównywalne. Niedużą tendencję wzrostową można zauważać w ostatnich trzech okresach, co potwierdza wykres 4.14. przedstawiający rosnący trend średniej wartości transakcji.

Transakcje do badań wybierane było kolejno z bloków, a priorytety kolejno podpisywanych transakcji zależą między innymi od wniesionej opłaty przez zleceniodawców. Przy szybkim wzroście wartości bitcoinów oraz założeniu, że giełdy lub uczestnicy sieci prowadzący duże wymiany wnoszą większe opłaty (w celu zapewnienia jak najszybszego przekazu środków) można również zauważać wzrost zainteresowania, w stosunku do okresów poprzednich, bogatych instytucji prowadzących interesy w sieci Bitcoin. Niestety z powodu dużej anonimowości uczestników sieci wyciągnięty wniosek nie jest niepodważalny.



Rysunek 4.13.: Mapa cieplna średniej wartości transakcji w sieci dla 10 prób w 10 okresach.



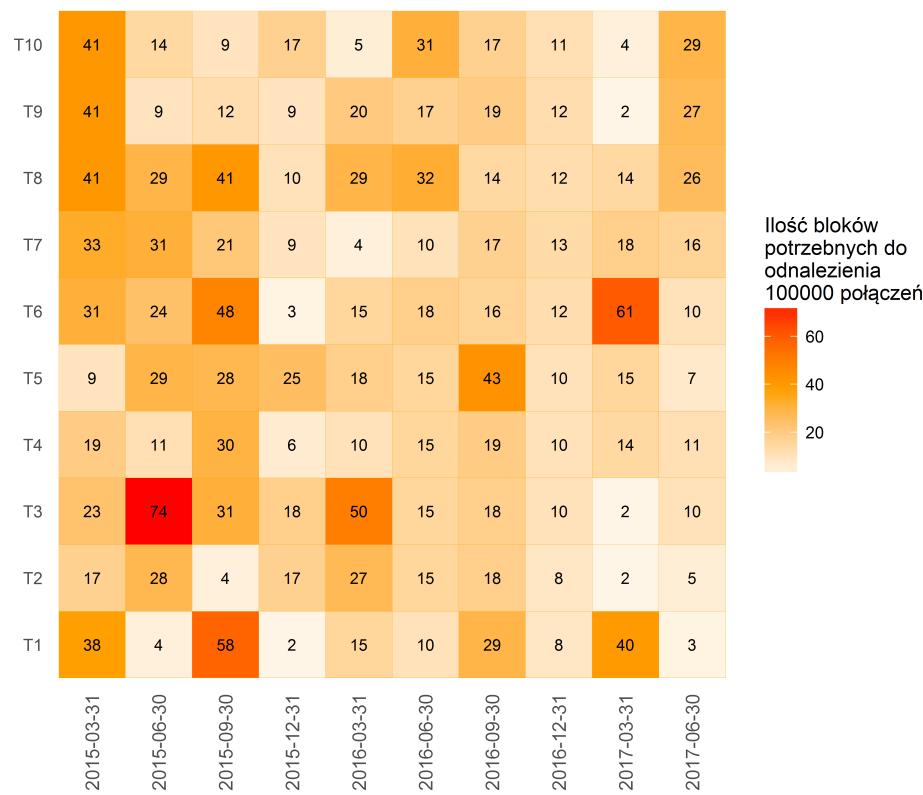
Rysunek 4.14.: Regresja liniowa średniej wartości transakcji w sieci dla 10 okresów z odchyleniem standardowym.

4.3.6. Badanie ilości bloków

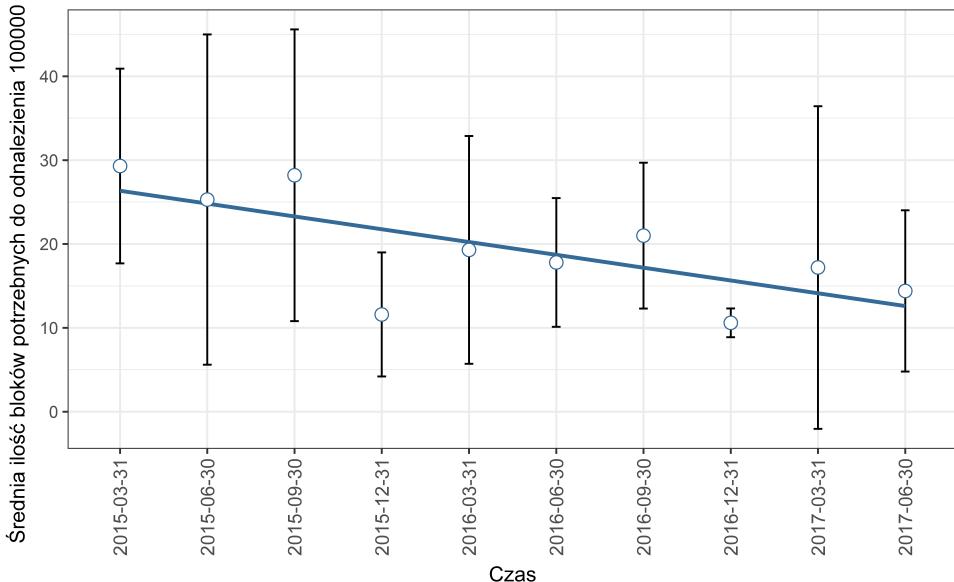
Ilość bloków jest kolejną miarą związaną z rodzajem badanej sieci. Polega na zliczeniu wszystkich bloków, w których znajdują się transakcje dołączone do próbki sieci. Każda transakcja może być zawarta jedynie w jednym bloku, natomiast jeden blok może zawierać wiele transakcji, które zostały połączone w procesie budowania sieci. Informacje o przynależności węzłów (transakcji) do poszczególnych bloku zostały wyekstrahowane podczas budowania próbek sieci i stanowią w niniejszym badaniu metadane węzła.

Zliczenie ilość bloków, w których znajdują się wszystkie powiązane ze sobą węzły zostało zrealizowane poprzez znalezienie wszystkich unikalnych adresów bloków z transakcji.

Na mapie cieplnej 4.15. zaprezentowano ilość bloków potrzebną na znalezienie stu tysięcy połączeń pomiędzy transakcjami. Na podstawie przebadanych prób zaobserwowano zróżnicowanie badanej wartości w okresach. Duża ilość bloków oznacza małą aktywność uczestników sieci rozciągniętą w czasie, natomiast mała ilość bloków oznacza dużą aktywność uczestników sieci w określonych momentach czasu. Należy zauważać, że prezentowana ilość bloków nie oznacza, że bloki te generowane były kolejno po sobie. Aktywność uczestników, zlecających transakcje, pozwalająca na budowę poszczególnych prób sieci mogła być nagła w różnych momentach czasu. Zakładając jednak sytuację, w której bloki budujące każdą z prób występowali kolejno po sobie, a ilość bloków generowana w sieci dziennie wała się w okolicy 170 bloków, stwierdzić można, że sieć składającą się ze stu tysięcy połączeń, która zaczyna się od pojedynczej transakcji, można zbudować w nie dłużej niż w przeciągu pół dnia. Prezentuje to ogrom badanej sieci, a zaprezentowany na wykresie 4.16. trend spadkowy wskazuje na coraz szybszy jej rozwój.



Rysunek 4.15.: Mapa cieplna ilości bloków sieci potrzebnych na znalezienie 100000 połączeń dla 10 prób w 10 okresach.



Rysunek 4.16.: Regresja liniowa ilości bloków sieci potrzebnych na znalezienie 100000 połączeń dla 10 okresów z odchyleniem standardowym.

4.3.7. Badanie średniej różnicy czasów kolejnych transakcji

Badanie różnicy czasów kolejnych transakcji, to własność badanej sieci ścisłe związana z jej fachową stroną. Polega na zliczeniu różnic czasów realizowania kolejnych transakcji wstecz. Oparta jest na przeglądzie wszystkich możliwych ścieżek w sieci z jednoczesnym ekstrahowaniem czasu $t(v_i)$ ich wykonania. W konsekwencji pozwala to na wyliczenie różnic czasu pomiędzy transakcjami v_i i v_{i-1} w ścieżkach. Średnia wartość różnic wykonywania kolejnych transakcji w ścieżce jest ilorazem sumy wszystkich różnic w ścieżce i ich ilości, co można zapisać jako

$$M(\sigma) = \frac{\sum_{i=1}^{n-1} t(\sigma(v_{i+1})) - t(\sigma(v_i))}{n - 1} \quad (4.6.)$$

gdzie M to średnia wartość różnic czasu kolejnych transakcji w ścieżce σ , $t(\sigma(v_i))$ to czas wykonania i -tej transakcji w ścieżce σ , $t(\sigma(v_{i+1}))$ to czas t wykonania transakcji v_{i+1} w ścieżce σ , a n to ilość połączeń pomiędzy transakcjami w ścieżce.

Średnią wartości różnic wykonania kolejnych transakcji w całej sieci można zapisać zatem jako

$$\frac{\sum_{s \neq t} M(\sigma_{st})}{m} \quad (4.7.)$$

gdzie σ_{st} to ścieżka prowadząca od węzła s do węzła t , a m to liczba wszystkich ścieżek.

Różnica czasów kolejnych transakcji pozwala określić częstotliwość ich wykonywania. Im mniejsza wartość, tym transakcje wykonywane były częściej, co w konsekwencji pozwala określić dynamikę sieci.

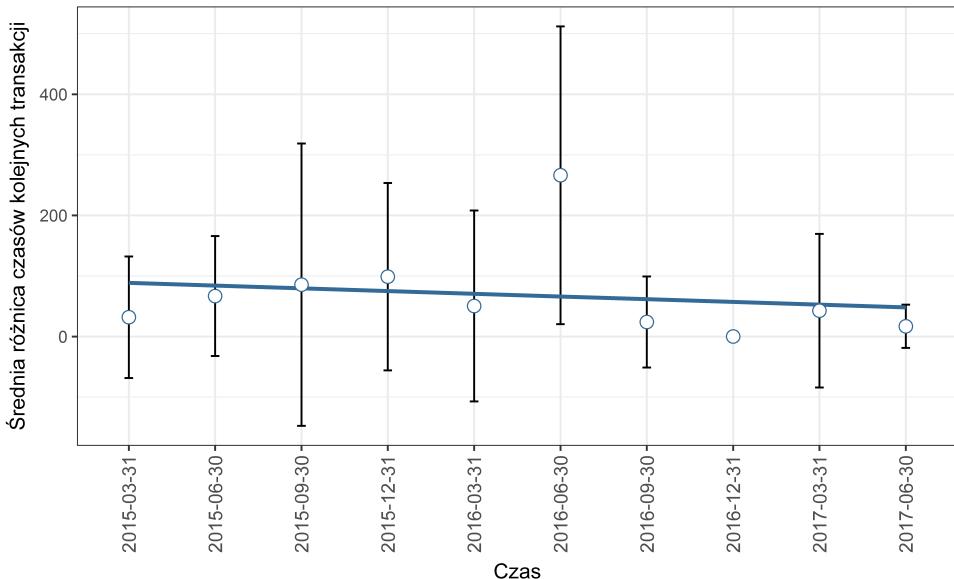
Na podstawie mapy cieplnej badanej właściwości 4.17. wywnioskowano, że połączenia pomiędzy transakcjami dla większości prób zostały odnalezione już w kilku poprzedzających blokach. Wskazują na to bardzo małe wartości badanej cechy. Sądzić można zatem, że sieć Bitcoin jest bardzo dynamiczna, a jej uczestnicy przekazują środki pomiędzy adresami w bardzo krótkich odstępach czasu. Może to również oznaczać ciągłą, wzmożoną aktywność posiadaczy badanych adresów.

Istnieje również ograniczona liczba prób, w których różnica czasu wykonania połączonych transakcji bliska jest kilkunastu minut, co przy wielkości badanej sieci sumarycznie oznaczać może kilkanaście miesięcy. Mogą to być transakcje wykonywane przez klientów sporadycznie korzystających z sieci Bitcoin, którzy rzadko przekazują środki na inne adresy lub przekazują je pomiędzy ich ograniczoną liczbą. Utrudnia to budowę sieci, gdyż w celu odnalezienia połączeń wymagany jest długi przegląd łańcucha bloków. Podstawą niniejszego wnioskowania mogą być próby z okresu *2016-06-30*, dla których wartość zdecydowanie odbiega od reszty okresów.

Na wykresie 4.18. zauważać można okresy, w których wartości średniej dla próbek lokalnie odbiegają od średniej. W kontekście całego badania nie mają one jednak dużego wpływu na wynik końcowy tworzący linię trendu badanej właściwości. Na jej podstawie można wnioskować o zmniejszającym się średnim czasie rejestrowania kolejnych transakcji, co w konsekwencji oznacza wzrastającą ilość wykonywanych transakcji i większe zaangażowanie użytkowników sieci.



Rysunek 4.17.: Mapa cieplna średniej różnicy czasu kolejnych transakcji sieci dla 10 prób w 10 okresach.



Rysunek 4.18.: Regresja liniowa średniej różnicy czasu kolejnych transakcji sieci dla 10 okresów z odchyleniem standardowym.

4.3.8. Badanie różnicy czasów granicznych transakcji

Badanie różnicy czasów granicznych transakcji jest właściwością sieci bezpośrednio związaną z jej specyfiką. Polega na odnalezieniu wśród wszystkich możliwych ścieżek takiej ścieżki, której różnica czasów pierwszej oraz ostatniej transakcji jest największa. Jako że próby przygotowane w celu przeprowadzenia serii badań powstały od określonych transakcji t , czasem wykonania pierwszej transakcji będzie zawsze czas wykonania transakcji początkowej t . Różnica czasu wykonania transakcji początkowej t oraz ostatniej transakcji dołączonej do danej ścieżki stanowi podstawę omawianego badania i może zostać przedstawiona jako

$$D(\sigma) = t(\sigma(v_1)) - t(\sigma(v_l)) \quad (4.8.)$$

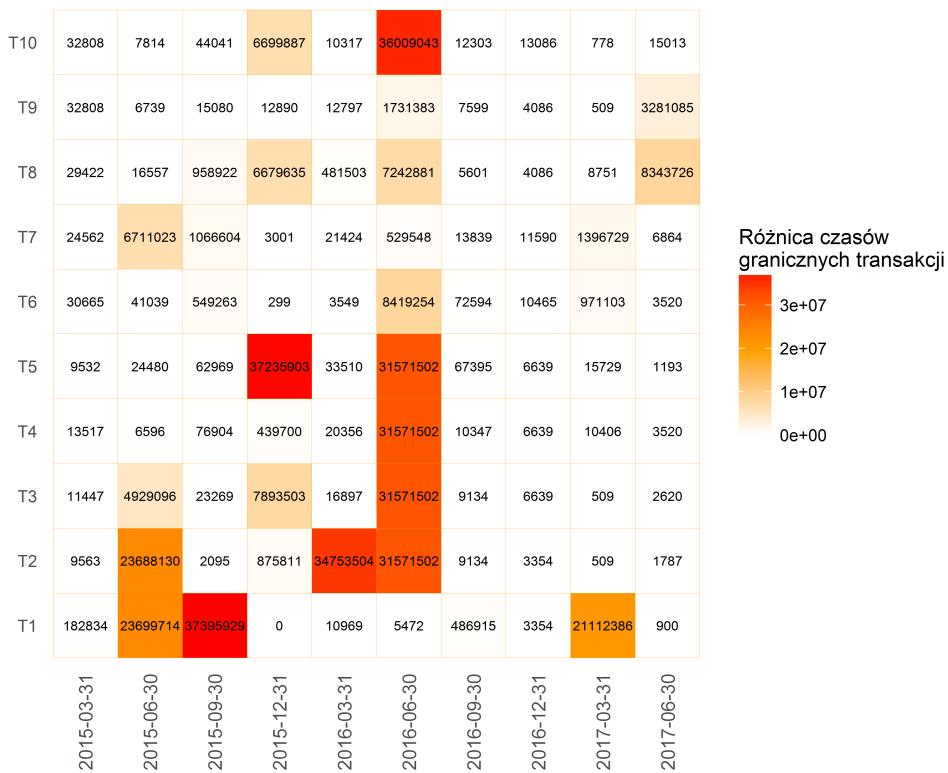
gdzie $D(\sigma)$ to różnica czasów wykonania transakcji, $t(\sigma(v_1))$ to czas wykonania pierwszej transakcji ścieżki, a $t(\sigma(v_l))$ to czas wykonania ostatniej transakcji ścieżki.

Różnicę czasów granicznych transakcji można zatem zapisać jako

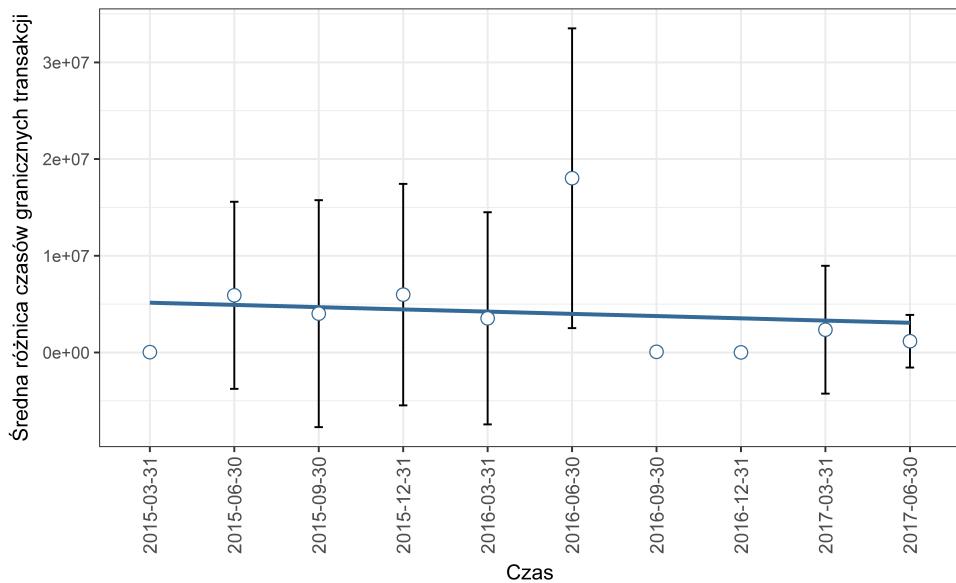
$$\max_{\sigma_{st}, s \neq t} D(\sigma_{st}) \quad (4.9.)$$

gdzie σ_{st} to ścieżka prowadząca od węzła s do węzła t .

Wyniki przeprowadzonego badania zaprezentowano na mapie cieplnej 4.19.. Potwierdza ona wnioski opisane w podrozdziale 4.3.7. opisującym badanie różnic czasów kolejnych transakcji. Dla większości prób we wszystkich okresach zaobserwowano wykrycie wszystkich stu tysięcy połączeń pomiędzy transakcjami w bardzo krótkim czasie, determinując przegląd tylko kilku poprzedzających bloków. Różnica czasów granicznych transakcji dla prób w okresach, których średni czas kolejnych transakcji znacznie odbiegał od średniej, jest największa. Potwierdza to sporadyczne realizowanie transakcji właścicielami adresów w danych próbkach w okresie 2016-06-30. Na podstawie ostatnich czterech okresów wnioskować można, że aktualnie podobne sytuacje są rzadkością i większość uczestników sieci realizuje znacznie więcej transakcji niż wcześniej, co potwierdza również wykres 4.20.. Pozwala to na stwierdzenie, że sieć Bitcoin w przeciągu ostatnich czterech badanych okresów bardzo urosła i zyskała zaufanie większej liczby użytkowników.



Rysunek 4.19.: Mapa cieplna różnic czasu granicznych transakcji dla 10 prób w 10 okresach.



Rysunek 4.20.: Regresja liniowa średniego różnic czasów granicznych transakcji dla 10 okresów z odchyleniem standardowym.

4.4. Wnioski

Sieć Bitcoin jest bardzo złożoną, dynamiczną i szybko rozwijającą się siecią. Na podstawie przeprowadzonych badań wywnioskowano ciągle rosnącą gęstość sieci, co w odniesieniu do jej specyfiki oznacza coraz większą ilość zlecanych transakcji, powstawanie dużej ilości nowych adresów oraz wzrost realizowanych transakcji pomiędzy różnymi uczestnikami sieci. W podrozdziale 4.3.1. zjawisko to zostało przedstawione bardziej szczegółowo. Badanie centralności węzłów opisane w podrozdziale 4.3.4. pozwoliło na zarejestrowanie znaczącego spadku istotności pojedynczej transakcji w całej sieci, co ponownie potwierdza fakt rosnącej ilości zlecanych transakcji. Zaobserwowano również, że właściwości sieci nie są stałe w jednym okresie, a zależeć mogą od aktywności poszczególnych uczestników, dlatego też ilość połączeń pomiędzy transakcjami może być bardzo zróżnicowana, co zaprezentowano w podrozdziałach 4.3.2. i 4.3.3..

W trakcie badania poszczególnych próbek zauważono wzmożoną działalność pojedynczych właścicieli adresów, którzy zlecali lub odbierali znaczne sumy bitcoinów. Z powodu dużej anonimowości w sieci nie da się wprost określić właścicieli użytych adresów w tych transakcjach. Przypuszczać można jednak, że mogą to być adresy należące do giełd lub dużych instytucji. Dodatkowo na podstawie analizy średniej wartości transakcji zauważono, że wartość większości transakcji nie przekracza 100 bitcoinów, a zazwyczaj są to małe przekazy środków pomiędzy klientami sieci. Porównanie wyników średnich wartości transakcji dla okresów w podrozdziale 4.3.5. pozwala stwierdzić, że ilość przekazywanych bitcoinów w transakcjach cały czas rośnie. Badanie ilości bloków potrzebnych do odnalezienia 100 tysięcy połączeń, w powiązaniu z analizą średniej różnicy czasów kolejnych transakcji oraz różnicy czasów transakcji granicznych (opisane kolejno w podrozdziałach 4.3.6., 4.3.7., 4.3.8.), ilustruje wielkość oraz tempo rozwoju badanej sieci. W ostatnich okresach transakcje do prób zostały odnalezione w kilku blokach, a różnice czasów granicznych transakcji zawartych w próbach nie przekraczają kilkudziesięciu minut. W rezultacie pozwala to na wywnioskowanie o kolejności występowania po sobie bloków, w których odnaleziono połączenia pomiędzy transakcjami. Odnalezienie tak olbrzymiej ilości powiązanych transakcji w paru kolejnych blokach potwierdza rosnącą ilość realizowanych transakcji pomiędzy różnymi uczestnikami sieci w bardzo ograniczonym czasie.

Reasumując, sieć Bitcoin jest siecią bardzo dynamicznie się rozwijającą. Można przypuszczać, że coraz więcej użytkowników używa jej jako codziennego medium transmisyjnego środków, a rosnącą gęstość sieci świadczy o rosnącej ilości samych użytkowników.

Podsumowanie

Bibliografia

- [1] Sebastian Bala Witold Srokosz. *Kryptowaluty jako elektroniczne instrumenty płatnicze bez emitenta*. Wydawnictwo Uniwersytetu Wrocławskiego, 2016.
- [2] Michael J. Casey Paul Vigna. *Cryptocurrency*. Random House UK Ltd, 2016.
- [3] un. coinmarketcap.com. <https://coinmarketcap.com/>, November 2017.
- [4] Andreas M. Antonopoulos. *Mastering Bitcoin: Unlocking Digital Cryptocurrencies*. O'Reilly Media, 2014.
- [5] un. blockchain.info. <https://blockchain.info/>, 2017.
- [6] Arvind Narayanan, Joseph Bonneau, Edward Felten, Andrew Miller, and Steven Goldfeder. *Bitcoin and Cryptocurrency Technologies*. Princeton Univers. Press, 2016.

Spis rysunków

1.1. Krzywa eliptyczna zastosowana do wyznaczenia wartości G w protokole Bitcoin	6
2.1. Przykładowy łańcuch 3 bloków sieci Bitcoin.	9
2.2. Przykładowa zawartość jednego bloku sieci Bitcoin.	10
4.1. Przykładowy wycinek próbki sieci.	17
4.2. Wizualizacja grafu składającego się ze stu tysięcy połączeń.	18
4.3. Mapa cieplna średnicy sieci dla 10 prób w 10 okresach.	20
4.4. Regresja liniowa średniej średnicy sieci dla 10 okresów z odchyleniem standardowym.	21
4.5. Mapa cieplna średniej długości scieżki w sieci dla 10 prób w 10 okresach.	22
4.6. Regresja liniowa średniej długości scieżki sieci dla 10 okresów z odchyleniem standardowym.	22
4.7. Mapa cieplna średniego stopnia węzła sieci dla 10 prób w 10 okresach.	24
4.8. Regresja liniowa średniego stopnia węzła sieci dla 10 okresów z odchyleniem standardowym.	24
4.9. Histogram średniego stopnia węzła łączny dla wszystkich prób.	25
4.10. Mapa cieplna średniej centralności węzłów w sieci dla 10 prób w 10 okresach.	26
4.11. Regresja liniowa średniej centralności węzłów w sieci dla 10 okresów z odchyleniem standardowym.	26
4.12. Histogram centralności węzłów w sieci łączny dla wszystkich prób.	27
4.13. Mapa cieplna średniej wartości transakcji w sieci dla 10 prób w 10 okresach.	28
4.14. Regresja liniowa średniej wartości transakcji w sieci dla 10 okresów z odchyleniem standardowym.	28
4.15. Mapa cieplna ilości bloków sieci potrzebnych na znalezienie 100000 połączeń dla 10 prób w 10 okresach.	29
4.16. Regresja liniowa ilości bloków sieci potrzebnych na znalezienie 100000 połączeń dla 10 okresów z odchyleniem standardowym.	30
4.17. Mapa cieplna średniej różnicy czasu kolejnych transakcji sieci dla 10 prób w 10 okresach.	31
4.18. Regresja liniowa średniej różnicy czasu kolejnych transakcji sieci dla 10 okresów z odchyleniem standardowym.	32
4.19. Mapa cieplna różnic czasu granicznych transakcji dla 10 prób w 10 okresach.	33
4.20. Regresja liniowa średniego różnic czasu granicznych transakcji dla 10 okresów z odchyleniem standardowym.	33

Spis tabel.

2.1. Struktura transakcji.	12
2.2. Struktura wyjścia transakcji.	12
2.3. Struktura wprowadzania transakcji.	13
4.1. Związek między transakcjami.	17