

## Analiza grafu World City Network

### 1. Opis plików

Graf został pobrany ze strony: <http://vlado.fmf.uni-lj.si/pub/networks/data/mix/mixed.htm>

Ściągnięte zostały dwa pliki:

- WorldCities.net – plik zawiera definicję wierzchołków oraz krawędzi grafu
- WorldCities.clu – plik zawiera informację, czy dany wierzchołek jest miastem czy firmą. Jeśli dany wierzchołek jest miastem to wtedy przyporządkowana zostaje wartość 0. Jeśli wierzchołek jest firmą wtedy przyporządkowywana jest liczba od 1 do 6 w zależności jaki jest to rodzaj firmy:
  - 1 – accountancy - księgowość
  - 2 – advertising - reklama
  - 3 - banking/finance – bankowość/finanse
  - 4 – insurance - ubezpieczenia
  - 5 – law - prawo
  - 6 – management - zarządzanie

Z pliku WorldCities.clu zostały usunięte pierwsze linijki zaczynające się od %, czyli zawierające komentarze, ponieważ powodowały błąd przy flagowaniu/oznaczaniu wierzchołków.

Dodatkowo stworzone zostały przez mnie trzy dodatkowe pliki znakujące wierzchołki grafu. Dzięki tym plikom, będzie możliwa analiza we względu na kontynent, państwo oraz stolicę:

- WorldCitiesCountries.clu – plik zawiera przyporządkowanie dla każdego węzła, który jest miastem, nazwę państwa gdzie znajduje się to miasto. Dla węzłów, które są firmami wpisana została wartość null
- WorldCitiesContinents.clu – plik zawiera przyporządkowanie dla każdego węzła, który jest miastem, kontynenty na którym znajduje się to miasto. Dla węzłów, które są firmami wpisana została wartość null
- WorldCitiesCapitals.clu – plik zawiera przyporządkowanie dla każdego węzła, który czy jest stolicą. Jeśli miasto jest stolicą przyporządkowana została wartość 1, w przeciwnym wypadku 0. Dla węzłów, które są firmami wpisana została wartość -1

## 2. Opis grafu

Graf zawiera 315 wierzchołków, które są miastami oraz 100 wierzchołków, które są firmami. Wierzchołki, które są miastami nie są połączone ze sobą. Taka sama sytuacja jest z wierzchołkami, które są firmami. Graf przedstawia jakie firmy mają swoje biura/siedziby w różnych miastach na świecie. Relacja zawsze zachodzi pomiędzy węzłem, który jest miastem oraz węzłem, który jest firmą. Graf jest grafem nieskierowanym.

Do grafu, dla każdego wierzchołka dodane zostały cztery atrybuty na podstawie wyżej wymienionych plików, które umożliwią analizę z różnego punktu widzenia.

- description – jeśli wierzchołek jest miastem to 0, jeśli firmą to 1-6
- continents – jeśli wierzchołek jest miastem to: Europe, Asia, North America, South America, Africa lub Australia and Oceania, jeśli firmą to null
- countries – jeśli wierzchołek jest miastem to państwo, do którego należy, np. Germany, Spain, jeśli firmą, to null
- capitals – jeśli wierzchołek jest miastem to 1 jeśli jest to stolica lub 0 jeśli nie jest to stolica, jeśli firma to -1

## 3. Analiza grafu WorldCities

Stopień wężła jest to liczba relacji, z którym węzeł jest połączony z resztą grafu a więc analizując graf za pomocą parametru stopień możemy odpowiedzieć na następujące pytania:

1. Które miasto ma najwięcej a które najmniej siedzib firm?

Największy stopień posiada miasto Londyn – 99. Oznacza to, że w Londynie aż 99 ze 100 firm ma swoją siedzibę. Znajdują się w grafie również dwa wierzchołki, które są izolowane a więc dla nich stopień jest równy 0, czyli żadna z firm nie posiada w tych miastach swojej siedziby: Lucknow oraz Pyongyang.

2. Która firma ma najwięcej, a która ma najmniej swoich siedzib?

Spośród wszystkich firm najmniej siedzib ma firma Morgan Lewis, jest ich tylko 10. Jest to firma prawnicza. Natomiast najwięcej siedzib ma firma KPMG. Jest to firma zajmująca się księgowością i ma 269 siedzib.

3. Jaki jest rozkład liczby siedzib firm w poszczególnych miastach?

Rozkład jest lewostronnie skośny oznacza to, że dużo miast ma mało siedzib firm, a tylko niektóre miasta takie jak na przykład London, New York, Tokyo czy Hong Kong mają siedziby prawie wszystkich firm.

4. Jaki jest rozkład liczby miast w których znajduje się firma?

Rozkład jest lewostronnie skośny, czyli oznacza to, że mało jest takich firm które mają siedzibę w dużej liczbie miast. Można wskazać dwie czołowe firmy: PWC oraz KPMG. Większość firm ma siedziby w mniej niż 100 miastach na całym świecie.

5. Jaka firma w każdej kategorii (1-6) ma najwięcej siedzib?

Spośród firm zajmujących się księgowością najwięcej siedzib ma firma KPMG. Warto zauważyć, że każda z firm zajmujących się księgowością ma kontakt w 140 miastach lub więcej.

Firmy zajmujące się reklamą mają siedziby w znacznie mniejszej liczbie miast niż firmy zajmujące się księgowością. Firma z największą liczbą siedzib to firma McCann Erickson, która ma siedzibę w 154 miastach.

Duży rozstęp pomiędzy liczbą siedzib w różnych miastach można zaobserwować dla firm bankowych. Najwięcej siedzib ma firma Citibank bo aż 168, a najmniej Sakura ponieważ ma siedziby tylko w 32 miastach.

Kolejną grupą są firmy ubezpieczeniowe, które skupiają w sobie tylko 6 firm. Spośród nich najwięcej siedzib ma firma Winterhur – 210.

Firmy zajmujące się prawem mają najmniej siedzib w poszczególnych miastach, ponieważ firma, która zawiera najwięcej siedzib to Maker & MCKenzie i posiada ona tylko 58 siedzib.

Ostatnią grupę tworzą firmy zarządzające i wśród nich firma Deloitte ma najwięcej siedzib – 107.

6. Które miasto na poszczególnym kontynencie ma najwięcej oraz najmniej siedzib?

Kontynent	Miasto	Największa liczba siedzib	Miasto	Najmniejsza liczba siedzib
Europa	Londyn	99	Kraków, Liege	3
Azja	Tokio, Hong Kong	92	Lucknow, Pyongyang	0
Afryka	Johanesburg	55	Brazzaville	1
Ameryka Północna	Nowy Jork	99	Port-Au-Prince	3
Ameryka Południowa	Sao Paulo	72	Manaus	3
Australia i Oceania	Sydney	72	Suva	8

Jak można zauważyć na każdym kontynencie znajduje się miasto lub miasta, które mają siedziby ponad 70% firm, wyjątkiem jest Afryka. Dodatkowo można zauważyć, że niezależnie od kontynentu znajdują się miasta, które mają mniej niż 10% siedzib firm.

7. Które miasta w Stanach Zjednoczonych mają najwięcej siedzib firm?

Miasta, które są w czołówce miast z największą liczbą siedzib firm to Nowy Jork, Chicago oraz Los Angeles. Najwięcej miast posiada między 18 a 32 siedzib różnych firm.

8. Jaka jest średnia liczba siedzib firm dla stolic oraz dla pozostały miast?

W stolicach jest średnio prawie 26 siedzib różnych firm, a w pozostały miastach średnio 22. Różnica nie jest duża ale wynika to z tego, że bez problemu można wskazać miasta, które nie są stolicami a zawierają siedziby większości firm np. Hong Kong, Nowy Jork czy Frankfurt.

Kolejnym parametrem, który może być obliczony, jest centralność stopnia. Z racji tego, że w grafie miasta nie są połączone z innymi miastami oraz firmy nie są połączone z innymi firmami a jedyna możliwa relacja jest pomiędzy firmą a miastem, proponuję policzyć centralność stopnia jako stosunek liczby krawędzi dla danego węzła(firmy) do liczby wszystkich możliwych miast znajdujących się w danym podgrafie. Za pomocą tak zmodyfikowanego parametru centralności stopnia można odpowiedzieć na następujące pytania:

1. Jakie są udziały banków w Europie?

Bank	Udział w Europie
West LB	0.27
Dresdner	0.39
Commerzbank	0.37
Deutsche	0.53
Chase	0.20
BNP Paribas	0.33
ABN Amro	0.43
CSFB	0.15
Rabobank	0.17
UBS	0.13
ING	0.22
Barclays	0.38
Fuji	0.06
BHV	0.39
BLG	0.23
Sakura	0.05
Sumitomo	0.21
Sanwa	0.05
JP Morgan	0.13
BTM	0.17
DKB	0.07
HSBC	0.41
Citibank	0.41

Banki: ABN Amaro, HSBC oraz CitiBank mają siedziby w ponad 40% europejskich miast. Są one najbardziej finansowo połączone z Europą.

2. Jaki jest największy oraz najmniejszy procentowy udział banków w Azji?

W Azji siedziby banku Sumitomo znajdują się w ponad 78% azjatyckich miast, czyli można powiedzieć, że ten bank jest bardzo rozpowszechniony w Azji. Dla porównania bank Sumitomo w Europie ma znacznie mniej siedzib.

Bank UBS ma najmniej siedzib w azjatyckich miastach spośród wszystkich banków.

3. Które banki nie mają siedzib w Afryce?

Banki, które nie mają ani jednej siedziby w Afryce to: Fuji, Skura, Sanwa oraz DKB.

4. Które banki mają najwięcej siedzib w Ameryce Północnej?

Banki, które mają najwięcej siedzib w Ameryce Północnej to Chase oraz Citibank.

Porównując centralność stopnia dla banków na kontynentach: Europa, Azja, Afryka oraz Ameryka Północna można powiedzieć, że najmniej udziałów finansowych banki mają w Afryce. W Afryce aż 4 banki nie mają żadnej siedziby na całym kontynencie. Dodatkowo można powiedzieć, że pozostałe banki w większości, mają mało siedzib. Tylko 6 banków z 23 ma swoje siedziby w ponad 20% miast stąd rozkład centralności stopnia dla banków na kontynencie Afryka nie ma rozkładu normalnego ( $pvalue < 0.0005$ , test Shapiro).

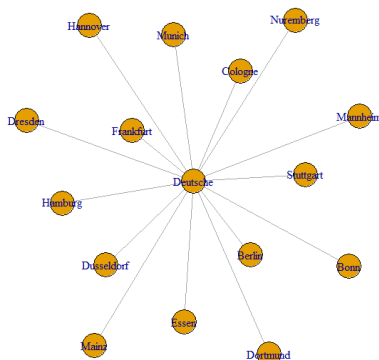
W Europie nie ma banku-lidera, który miałby dużo więcej siedzib niż pozostałe banki.

Natomiast w Azji oraz Ameryce Północnej można wskazać banki-lidery.

Kolejne pytania, na które można odpowiedzieć to:

5. Czy są banki w Niemczech, które mają siedzibę w każdym mieście?

W Niemczech 2 banki mają siedzibę w każdym mieście, jednym z nich jest bank Deutsche



Fragment grafu dla banku „Deutsche” i niemieckich miast.

6. Jak wygląda sytuacja z liczbą siedzib w Stanach Zjednoczonych?

W Stanach Zjednoczonych sytuacja z liczbą siedzib w miastach jest inna niż w przypadku Niemiec. W USA nie ma banku, który miał by siedzibę w każdym mieście. Dodatkowo znajdują się banki, które mają siedzibę w kilku miastach w USA. Być może jest to spowodowane tym, że Niemcy są małym krajem w stosunku do USA, wielkością przypominające jeden stan USA, stąd w USA jest większe zróżnicowanie.

7. Które stolice są centrami finansowymi (które mają najwięcej siedzib)?

Do stolic, w których mieści się największa liczba różnych banków należą: Pekin, Jakarta, Paryż, Seul, Londyn oraz Tokio. W Londynie oraz w Tokio wszystkie banki mają swoją siedzibę.

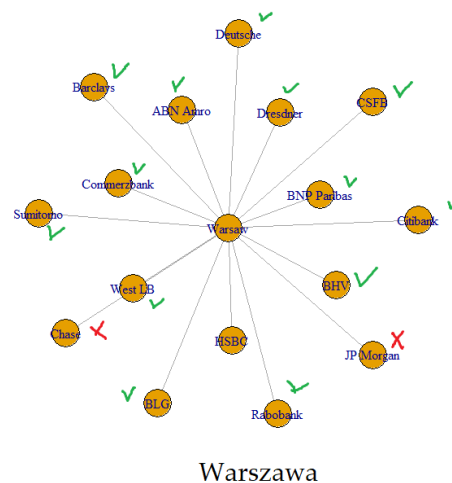
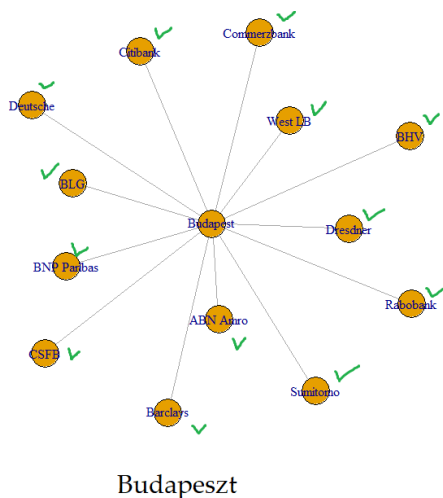
8. Które stolice nie mają ani jednej siedziby banku?

Stolice takie jak: Bagdad, Belgrade, Brazzaville, Freetwon, Jerusalem, Kabul, Khartoun, Managua, Port Moresby, Phonyang, Reykjavik, Suva, Tallin, Tirana, Tripoli, Yaonde. Jak widać większość z tych stolic to stolice biednych Państw, albo takich w których jest ustrój totalitarny jak na przykład stolica Korei Północnej. Niektóre z tych państw przez ostatnie 20 lat bardzo się rozwinęły, na przykład: Islandia czy Estonia.

Kolejny parametr grafu, który został użyty do analizy to ekwiwalencja strukturalna, która pozwala określić jak bardzo podobne są otoczenia dwóch wierzchołków. W kontekście analizowanego grafu, użyta zostanie ekwiwalencja strukturalna aby wyznaczyć miasta, których pozycja finansowa/ubezpieczeniowa jest podobna (mają wszystkie albo prawie wszystkie takie same siedziby firm). Należy pamiętać, że jeśli miara podobieństwa jest równa 1 to oznacza to, że wierzchołki są takie same, nie oznacza to, że muszą zawierać dużo firm. Do poniższej analizy wzięto pod uwagę tylko te pary miast, które mają współczynnik podobieństwa jest większy niż 0.8. Za pomocą tego parametru można odpowiedzieć na następujące pytania:

1. Jakie państwa w Europie mają podobną pozycję finansową w znaczeniu posiadają siedziby większości tych samych firm?

Jako przykład może pokazać Budapeszt i Warszawę. Za pomocą grafu sąsiedztwa możliwe było narysowanie dla obydwu miast ich sąsiadów, czyli firm, które mają siedzibę w tych miastach. Kolorem zielonym zaznaczone zostały firmy, które mają siedzibę zarówno w Budapeszcie jak i w Warszawie. Wszystkie firmy, które mają siedzibę w Budapeszcie mają również siedzibę w Warszawie. Dodatkowo w Warszawie mieszczą się dwie siedziby firm, których nie ma w Budapeszcie.



Podobna sytuacja zachodzi dla miast pary miast: Warszawa i Praga, Warszawa i Mediolan, Warszawa i Moskwa, Londyn i Frankfurt, Londyn i Paryż, Madryt i Paryż. Dużo miast w Niemczech posiada siedziby tych samych firm. Podobna sytuacja może być zaobserwowana w Wielkiej Brytanii.

Można znaleźć również takie miasta, których współczynnik jest równy dokładnie 1, ale w większości zachodzi tam sytuacja, że mają one siedzibę jednej czy dwóch firm.

2. Jakie państwa w Azji mają podobną pozycję finansową?

W Azji pary miast, które mają prawie takie same siedziby firm to na przykład: Bangkok i Pekin, Pekin i Hong Kong, Istanbul i Mumbai, Singapur i Tokio, Shanghai i Tokio, Seul i Tokio, Jakarta i Tokio, Singapur i Pekin. Tutaj zostały wymienione takie miasta, które mają dużo firm. Można również znaleźć pary miast dla których współczynnik podobieństwa wyniósł 1 a mają tylko siedzibę jednego banku na przykład: Ułan Bator i Sanaa czy Malacca i Yerevan.

Jak można zauważyć, miasta, które są duże i bogate, są podobne do siebie pod kątem tego, że posiadają dużo siedziby tych samych firm.

3. Jakie państwa w Europie mają podobną pozycję związaną z liczbą siedzib firm ubezpieczeniowych?

Większość państw europejskich ma tylko niektóre siedziby firm ubezpieczeniowych, przez co łatwiej pary miast osiągają współczynnik podobieństwa równy 1. Jednakże można wyróżnić kilka par miast, które są podobne do siebie oraz mają więcej niż przeciętnie siedzib firm, na przykład: Madryt i Sztokholm, Londyn i Paryż, Dublin i Londyn. Można powiedzieć, że większość państw europejskich ma podobną sytuację z firmami ubezpieczeniowymi.

Można zauważyć, że dla miast które mają dużo siedzib firm w Europie to więcej przypadków podobieństwa występuje dla firm finansowych niż dla ubezpieczeniowych.

4. Jakie miasta w Ameryce Północnej mają prawie wszystkie siedziby firm bankowych takie same jak miasta w Europie?

Do takich miast należą: Londyn i Toronto, Nowy Jork i Paryż, Chicago i Paryż, Frankfurt i Nowy Jork, Londyn i Meksyk, Nowy Jork, Zurich. Miasta Londyn i Nowy Jork posiadają siedziby wszystkich firm bankowych i ich współczynnik podobieństwa jest równy 1. Można łatwo to wytłumaczyć, ponieważ Nowy Jork uważany jest za stolicę finansową w USA jak również w całej Ameryce Północnej a Londyn jest jego odpowiednikiem w Europie.

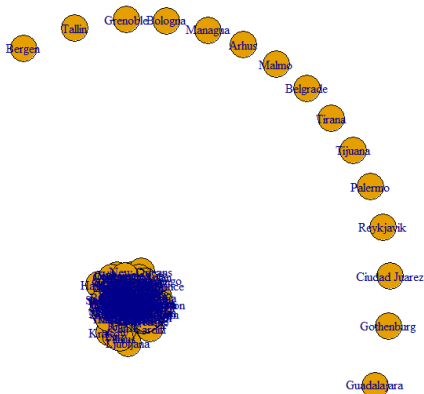
#### 4. Analiza grafu stworzonego w oparciu o graf WorldCities

Kolejną analizą, którą można zaproponować to stworzenie grafu, w którym, firmy „staną się” krawędziami, a więc uzyskamy graf połączeń między miastami. Jeżeli mamy miasto A oraz miasto B, które mają siedzibę tej samej firmy to możemy powiedzieć, że miasto A i B też są połączone między sobą, ponieważ firma ma siedziby w obydwu tych miastach. Dzięki tak stworzonemu grafowi będzie możliwa analiza, które miasta są na przykład finansowo powiązane ze sobą. Poniżej przedstawiony został algorytm w jaki sposób taki graf został stworzony:

1. Stworzone zostały dwa podgrafy: jeden zawierający tylko wybrane miasta ( $g_1$ ), a drugi zawierający wybrane miasta wraz z firmami z danej dziedziny ( $g_2$ )
2. Stworzona została kwadratowa macierz ( $m$ ) o wymiarach liczba wierzchołków grafu ( $g_1$ ), której indeksami są nazwy miast z grafu ( $g_1$ )
3. Dla każdego wierzchołka – miasta ( $w_1$ ) z grafu ( $g_2$ ) pobierani są sąsiedzi ( $s_1$ )
4. Dla każdego sąsiada ( $s_1$ ) pobierani są jego sąsiedzi ( $s_2$ )
5. Wśród wierzchołków ( $s_2$ ) wybierane są tylko wierzchołki ( $w_2$ ) – miasta a nie wierzchołki – firmy za pomocą operacji na zbiorze – część wspólna
6. Inkrementowana jest wartość w macierzy ( $m$ ) o współrzędnych [ $w_1, w_2$ ]
7. Kroki 3-6 powtarzane są dla każdego wierzchołka
8. Zerowanie wartości leżących na prawej przekątnej macierzy ( $m$ ), aby zredukować relację miasta z samym sobą.

Po zastosowaniu tego algorytmu stworzona została macierz sąsiedztwa. Wartości w macierzy sąsiedztwa są wagami relacji, czyli mówią ile firm łączy daną parę miast. Im większa jest wartość w macierzy sąsiedztwa tym więcej banków ma siedziby w danej parze miast. W języku R można przekształcić macierz sąsiedztwa na obiekt graf, a więc analizy zostaną przeprowadzone na obiekcie graf.

Na podstawie opisanego algorytmu został stworzony graf połączeń między miastami z Europy i z Ameryki Północnej, które mają wspólne siedziby firm bankowych. Graf zawiera 167 miast. Na przedstawionym poniżej grafie widać, że na pewno są miasta, które nie są połączone z innymi a oznacza to, że nie mają ani jednej siedziby firmy bankowej. Do takich miast należą między innymi: Bergen, Tallin, Bologna, Malmo, Tijuana, Guadalajara.



W celu zbadania jak bardzo miasto jest wpływowe finansowo można zastosować miarę obliczającą wagi dla każdego wierzchołka (ang. strength). Im większa wartość tej miary dla



danego węzła tym miasto jest bardziej wpływowo. Do takich miast należą Londyn oraz Nowy Jork, czyli jedno miasto w Europie oraz drugiej w Ameryce Północnej. Dodatkowo aż 15 miast w ogóle nie jest powiązanych z niczym a więc oznacza to, że nie mają ani jednej siedziby banku u siebie.

Kolejną miarą, którą można zastosować to obliczyć stopień dla każdego wierzchołka czyli miasta. Za pomocą tej miary możemy określić z iloma miastami połączone jest finansowo miasto, czyli ma chociaż jeden wspólny bank, który ma siedzibę w obydwu tych miastach. Do miast które są najbardziej połączone finansowo ( mają chociaż jeden wspólny bank) należą: Nowy Jork, Londyn, Paryż, Frankfurt oraz Mediolan. Dodatkowo można zauważyć, że większość państw, które są w czołówce finansowych połączeń to kraje europejskie. Średnio każde miasto jest połączone z prawie 98 innymi miastami.

Na podstawie dwóch wyżej opisanych miar można stworzyć miarę, która pozwoli obliczyć z iloma średnio bankami jest powiązane dane miasto z drugim. Dla każdego wierzchołka mamy obliczony stopień i moc (strength) a więc aby obliczyć średnią liczbę banków należy podzielić moc przez stopień. Miasta: Londyn, Nowy Jork oraz Paryż są połączone średnio z innymi miastami za pomocą ponad 6 banków a mniej niż 7. Kolejnych 14 miast jest średnio połączone z innymi miastami przez ponad 5 banków a mniej niż 6.

Kolejny parametr to gęstość grafu, który mówi o stosunku liczby krawędzi występujących w grafie do maksymalnej liczby krawędzi, która wynika z kombinatoryki. W przypadku omawianego grafu gęstość jest równa 0.588. Można powiedzieć, że graf jest dość gęsty wierzchołki (miasta) mają połączenie z wieloma innymi wierzchołkami (miastami).

Przeprowadzony został test statystyczny, który polegał na wygenerowaniu 30 grafów losowych o podanej liczbie wierzchołków oraz gęstości grafu. Następnie dla tak wygenerowanego grafu została policzona średnia stopnia. Otrzymane średnie zostały porównane ze średnią grafu, który jest analizowany. Test statystyczny t Studenta ( $p < 0.005$ ) wykazał, że wynik empiryczny jest istotnie różny od wyniku losowego. Średni stopień jest o wiele wyższy niż w grafie losowym, więc można powiedzieć, że mamy sytuację w której część państw jest bardziej powiązana finansowo ze sobą poprzez posiadanie siedzib tych samych banków niż reszta.

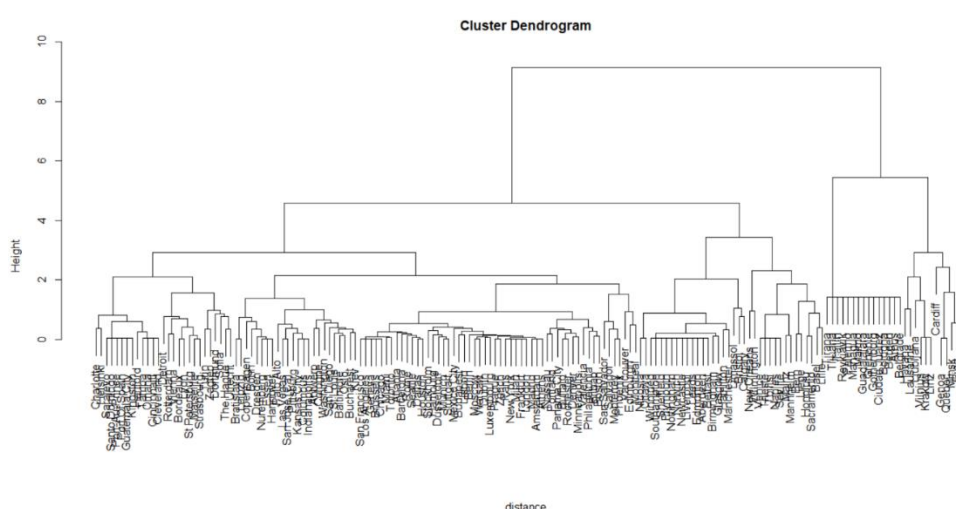
Analiza centralności stopnia wskazała również te same wierzchołki – miasta, które są najbardziej centralne, posiadają najwięcej połączeń, czyli mają największe wpływy finansowe.

Wyznaczona została centralność globalna, która jest równa 0.33, można więc twierdzić, że jest spora grupa miast, które są połączone ze sobą. Jeśli centralność globalna byłaby równa 0 to oznaczało by to że graf jest pełny, każde miasto jest połączone z każdym innym.

Dodatkowo wyznaczona została centralność pośrednicząca, która pokazała, że najbardziej kluczowymi wierzchołkami – miastami w grafie są: Nowy Jork, Paryż, Londyn, Frankfurt oraz Mediolan. Centralność została policzona zarówno dla grafu z uwzględnieniem wag jak i bez uwzględniania wag. Wagi są używane do obliczania ważonych najkrótszych ścieżek, dlatego są interpretowane jako odległości. Jeśli wartości wag są duże, to w tym sensie, najkrótsza ścieżka będzie długa. W tym przypadku należy odwrócić wagi dla krawędzi, i dopiero wtedy obliczyć centralność pośredniczącą, ponieważ, wtedy będziemy mogli wyznaczyć najbardziej centralne węzły. Wyniki dla centralności pośredniczącej zarówno dla pierwszego i drugiego podejścia wskazały te same wierzchołki – miasta jako centralne, ale w

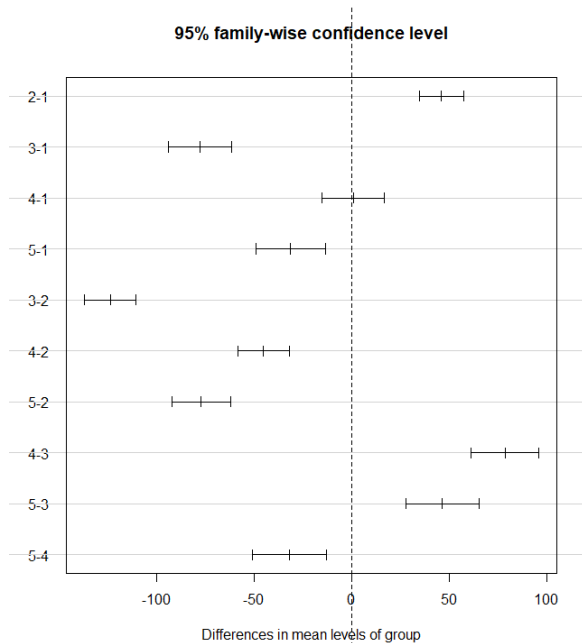
przypadku podejścia z uwzględnieniem wag jest to jeszcze bardziej zauważalne. Liderem jest Londyn z centralnością pośredniczącą – liczbą ścieżek równą ok. 928, następnie jest Nowy Jork – 842, potem dopiero Paryż ok. 557, Mediolan 508, Madryd – 350, Chicago – 300. Kolejnych 12 miast uzyskało wynik ponad 100, natomiast reszta miast znacząco od nich odbiega. Ich usunięcie może rozbić strukturę grafu, ponieważ są łącznikami szlaków finansowych.

Kolejną analizę, która można przeprowadzić to wyznaczyć ekwiwalencję strukturalną i na jej podstawie przeprowadzić klastrowanie hierarchiczne. Za pomocą ekwiwalencji strukturalnej możemy znaleźć węzły, które pełnią tę samą rolę (są podobne do siebie). Przeszukane zostały wszystkie podobieństwa, których współczynnik wynosił więcej niż 0.9. Większość dużych europejskich miast takich jak: Madryt, Warszawa, Frankfurt, Zurich, Praga czy Praga są bardzo podobne do siebie pod względem kontaktów finansowych. Można wskazać więcej par miast które są podobne na europejskim kontynencie niż na amerykańskim. Są również pary miast, które są podobne między kontynentami np. Londyn i Nowy Jork czy Miami i Zurich. Algorytm klastrowania hierarchicznego pozwala na łączenie w grupy wierzchołków podobnych do siebie. Na podstawie takiego algorytmu został wygenerowany poniższy wykres. Klastry zawierają grupę wierzchołków, które są podobne do siebie, pełnią podobną rolę.



Miasta podzielone zostały na 5 klastrow do dalszej analizy. Dla każdego klastra została obliczona gęstość grafu oraz średni stopień. Pierwszy klaster zawiera w dużej większości miasta Wielkiej Brytanii, drugi klaster jest największy i zawiera miasta które mają duże powiązanie finansowe między sobą, trzeci klaster – zawiera miasta, które nie są powiązane z żadnymi innymi miastami. Dodatkowo dla każdego miasta w danym klastrze policzony został współczynnik mówiący z iloma miastami ze swojego klastra połączone miasto w stosunku do połączeń w całym grafie. Policzone zostały również średnie takich współczynników dla każdego klastra. Z przeprowadzonej analizy wynika, że miasta w klastrze 1 oraz 2 silnie są powiązane ze sobą finansowo, w przypadku klastra 1 mamy do czynienia z grafem pełnym. Miasta klastrow 4 oraz 5 mają połączenie finansowe raczej z miastami z innych klastrow, ich gęstość jest niska. Jako przykład miasta które ma kontakty finansowe z miastami z innych klastrow można podać Ljubljane, która w ogóle nie ma powiązania z miastami w swoich klastrze a ogólnie w całym analizowany grafie ma połączenie z 26 miastami. Natomiast miasta z klastra 2 są silnie połączone z miastami ze swojego klastra na przykład: Madryt ok.

70% połączeń stanowią połączenia z miastami z klastra w którym znajduje się (klaster 2). Dodatkowo wykonany został test statystyczny ANOVA, aby sprawdzić czy są różnice między klastrami jeśli chodzi o stopień węzła. Zastosowany został również test statystyczny Tukeya, który wskazał między którymi grupami są różnice. Wszystkie klastry różnicują się ze sobą.



Wyznaczone zostały również współczynniki klasteryzacji dla kilku miast. Dla miast które są centralnymi miastami – są pośrednikami bo mają kontakty finansowe w prawie każdym mieście: Madryt, Londyn, Nowy Jork czy Paryż współczynnik klasteryzacji jest równy około 0.7. Natomiast dla takich miast jak Ljubljana czy Nikozja współczynnik jest równy 1. Te węzły są bardzo dobrze zintegrowane z otoczeniem. Usunięcie węzłów, które są pośrednikami może spowodować, że graf zostanie rozbity. W tym przypadku mamy kilka węzłów, które pełnią rolę centralnych więc usunięcie jednego z nich nie zaburzy grafu.

Jako kolejny parametr grafu wyznaczona została największa klika, która łączy 88 wierzchołków – miast. Tworzą ją duże miasta Europy i Ameryki Północnej. W skład tej największej kliki wchodzi prawie połowa wszystkich wierzchołków – miast analizowanego grafu. Kraje wchodzące do kliki są wszystkie ze wszystkim połączone przynajmniej jedną wspólną siedzibą banku.

W przypadku grafu nieskierowanego możliwa jest analiza komponentów słabych, ponieważ tam nieuwzględniany jest kierunek relacji, brane pod uwagę są półścieżki. Z analizy komponentów słabych dla całego analizowanego grafu można zaobserwować, że mamy jeden duży komponent a pozostałe komponenty są jedno elementowe. W przypadku analizowanego grafu analiza ścieżek, znajdowanie najkrótszej ścieżki nie ma większego uzasadnienia, ponieważ jeśli miasto A jest połączone finansowo z miastem B, a miasto B jest połączone z miastem C, to oznacza, że miasta A oraz C nie są powiązane finansowo, ponieważ gdyby były to na grafie miasto A byłoby połączone z miastem C.

## Podsumowanie

Graf WorldCities jest bardzo dużym grafem, którego wierzchołki są miastami lub firmami. Każdemu wierzchołkowi, który jest firmą przyporządkowany jest rodzaj firmy, dzięki temu możliwa jest analiza grafu tylko dla danej firmy. Dzięki stworzonemu przeze mnie plikowi z nazwami państw, kontynentów do których należą miasta jak również oznaczenie czy dane miasto jest stolicą możliwa jest analiza z podziałem na kontynenty, państwo oraz stolice. Sposobów analizy takiego grafu jest bardzo dużo, na przykład można wziąć pod uwagę kontynent i konkretny rodzaj firmy, albo kontynent i dokładnie jedną firmę. W pracy przedstawione zostały przykłady analiz, dla całego grafu lub jego fragmentów. Należy zauważyć, że jedyna relacja, która jest możliwa w tym grafie to relacja w której jeden wierzchołek jest miastem a drugi firmą.

Zaproponowane zostało stworzenie kolejnego grafu na podstawie grafu WorldCities, w którym wierzchołki – firmy stały się krawędziami i dzięki temu możliwe jest stworzenie mapy połączeń firmowych (np. finansowych), między miastami. Jeśli dany bank ma siedziby w dwóch miastach to można powiedzieć, że w jakiś sposób te miasta są połączone między sobą bo w jednym i w drugim mieście firma na jakieś wpływy. Z racji tego, że w danych miastach może być kilka firm, to można mówić o grafie ważonym. Waga określa jak bardzo miasta są połączone ze sobą, czyli ile firm ma siedziby w jednym i drugim mieście. Taki rodzaj grafu można stworzyć dla miast na danym kontynencie, w danym państwie, lub tych które są stolicami biorąc pod uwagę firmy z danej kategorii albo tylko jedną wybraną firmę. Ze względu na wielkość grafu została przeprowadzona analiza finansowa dla miast znajdujących się w Europie i Ameryce Północnej, ale ciekawe było by przeprowadzenie takiej analizy również dla innych kombinacji. Wszystkie obliczenia zostały przeprowadzone za pomocą języka R.