

Analiza sieci lotów w USA z 2008 roku.

Projekt na przedmiot Grafy i Sieci.

Piotr Szyszka, Kacper Wójtowicz, Paweł Woźniak

Politechnika Lubelska

14 stycznia 2022



Spis treści

- 1 Wstęp
- 2 Podstawowe informacje
- 3 Analiza grafu
 - Spójność
 - O wierzchołkach i krawędziach
 - Analiza nieważonych stopni wierzchołków
 - Analiza ważonych stopni wierzchołków
 - Parametry
 - Średnica sieci
 - Centralność bliskości
 - Pośrednictwo bliskości
 - Gęstość grafu
 - Analiza klastrowa

Wprowadzenie

Prezentacja jest dokumentacją przeprowadzonej analizy w programie Gephi na sieci przedstawiającej loty pomiędzy 305-cioma lotniskami w Stanach Zjednoczonych w 2008 roku.

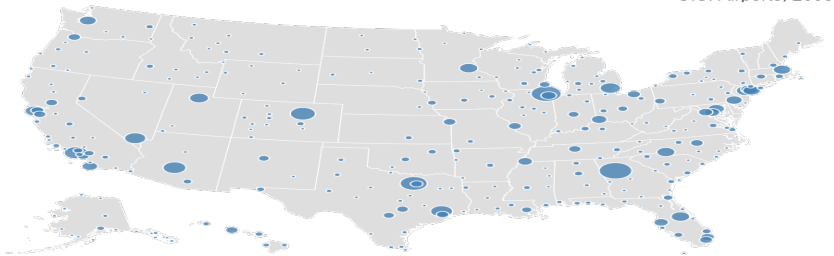
Sieć powstała na bazie jednego z [repozytoriów](#) użytkownika serwisu github o nazwie [jhree](#).

Pomiędzy dwoma wierzchołkami występuje krawędź jeżeli istnieje połączenie lotnicze pomiędzy nimi.

Waga krawędzi w grafie odzwierciedla ilość lotów, które odbyły się pomiędzy dwoma lotniskami.

Podstawowe informacje

Omawiany graf jest skierowany, ma **305** wierzchołków i **5366** krawędzi.



Analiza grafu

W dalszej części prezentacji omówimy takie zagadnienia jak:

1. *Spójność*
2. *Analiza wierzchołków i krawędzi*
3. *Parametry grafu*
4. *Analiza klastrowa*

Spójność

Istnieje wierzchołek, którego stopień wejściowy jest równy 0, przez co analizowany graf. **nie jest spójny**.

Wierzchołek ten ma identyfikator **PUB** i reprezentuje *Port lotniczy Pueblo Memorial*.

Id	Label	Modularity Class	In-Degree	Out-Degree	Degree
PUB	PUB	3	0	1	1

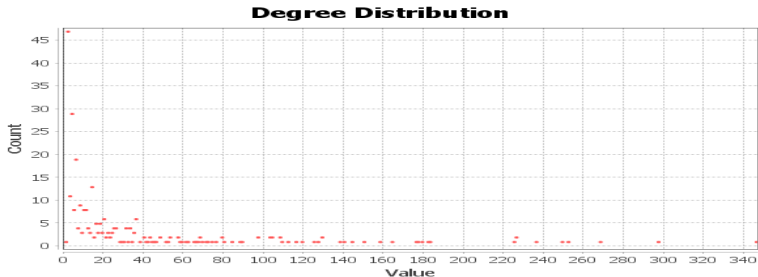
Analiza nieważonych stopni wierzchołków

Średni stopień wierzchołków - **17,593** (odzwierciedla średnią ilość połączeń lotniczych).

Degree Report

Results:

Average Degree: 17,593



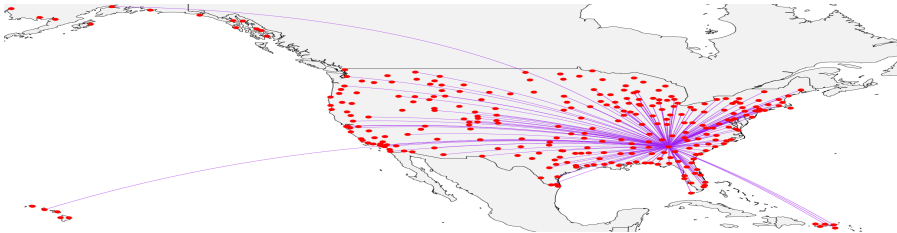
Rysunek: Raport mówiący o stopniach wierzchołków.

Największe lotniska

Wierzchołkiem o największym stopniu (**346**) jest **ATL** reprezentujący *Atlanta - Hartsfield-Jackson*.

Oznacza to, że ma najwięcej połączeń lotniczych - można do niego odlecieć i przylecieć z **173** lotnisk.

ATL



Rysunek: Połączenia lotniska ATL

```
lotniska <- airports[airports$iata %in% c(grafairports$origin, grafairports$destination),]  
atl <- filter(airports, airports$iata=="ATL")  
lotniska1 <- merge(lotniska, grafairports, by.x=c("iata"), by.y=c("origin"))  
colnames(lotniska1)[6] <- "lat_origin_airport"  
colnames(lotniska1)[7] <- "lon_origin_airport"  
lotniska2 <- merge(lotniska, lotniska1, by.x="iata", by.y="destination")  
colnames(lotniska2)[6] <- "lat_dest_airport"  
colnames(lotniska2)[7] <- "lon_dest_airport"  
colnames(lotniska2)[8] <- "origin_airport"  
lotyzatlanty <- filter(lotniska2, lotniska2$origin_airport=="ATL")  
  
xlim <- c(-171.738291, -56.601563)  
ylim <- c(12.039321, 71.856229)  
map("world", col="#f2f2f2", fill=TRUE, bg="white", lwd=0.1, xlim=xlim, ylim=ylim)  
for (i in (1:dim(lotyzatlanty)[1])) {  
  inter <- gcIntermediate(c(atl$longitude[1], atl$latitude[1]), c(lotyzatlanty$lon_dest_airport[i], lotyzatlanty  
$lat_dest_airport[i]), n=305)  
  lines(inter, lwd=0.1, col="purple")  
}  
points(lotniska$longitude, lotniska$latitude, pch=16, cex=0.5, col="red")
```



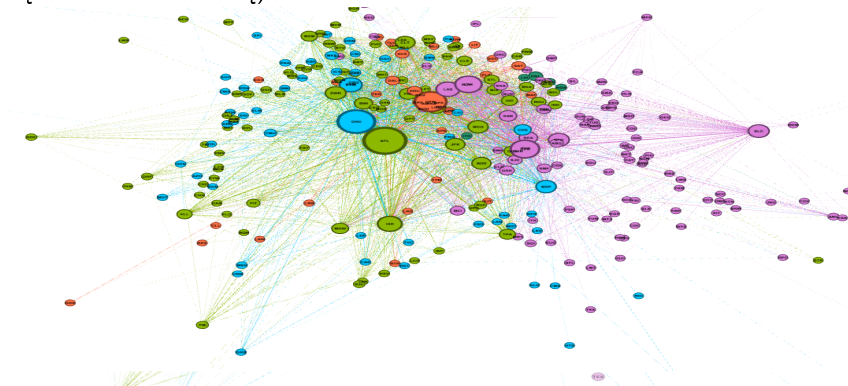
Istnieje jedno lotnisko do którego nie przylatuje żaden samolot. Jest to *Port Lotniczy Pueblo Memorial*(**PUB**). Stopień wejściowy tego wierzchołka(In-Degree) wynosi 0.

Są aż 2 lotniska z których nie ma połączenia lotniczego - są to *Port Lotniczy Cheyenne*(**CYS**) oraz *Ogden-Hinckley Airport*(**OGD**)- ich stopnie wyjściowe (Out-Degree) są równe 0.

Label	In-Degree	Out-Degree
CYS	2	0
OGD	2	0
PUB	0	1
ABY	1	1
ACT	1	1
ADK	1	1
ADQ	1	1
ALO	1	1
BET	1	1
BQK	1	1
CSG	1	1
DHN	1	1
EWN	1	1
FLO	1	1

Analiza ważonych stopni wierzchołków

Średni ważony stopień wierzchołków - 22983,715 (odzwierciedla średnią ilość lotów na trasę).



Najbardziej ruchliwe lotniska

Największy ruch odnotowano na lotnisku *Port lotniczy Atlanta - Hartsfield-Jackson*(**ATL**) przez który przewinęło się **829 034** samolotów, następne są *Port lotniczy Chicago-O'Hare* oraz *Port lotniczy Dallas-Fort Worth*.

Label	Weighted In-Degree	Weighted Out-Degree	Weighted Degree
ATL	414521.0	414513.0	829034.0
ORD	350452.0	350380.0	700832.0
DFW	281401.0	281281.0	562682.0
DEN	241470.0	241443.0	482913.0
LAX	215685.0	215608.0	431293.0
PHX	199416.0	199408.0	398824.0
IAH	185160.0	185172.0	370332.0
LAS	172871.0	172876.0	345747.0
DTW	162000.0	161989.0	323989.0
SFO	140579.0	140587.0	281166.0
SLC	139077.0	139088.0	278165.0
EWK	138491.0	138506.0	276997.0

Najwięcej lotów

Pomiędzy lotniskami *Port lotniczy San Francisco (SFO)*, a *Port lotniczy Los Angeles (LAX)* wykonano najwięcej lotów - aż 13788!

Source	Target	Type	Id	Weight
SFO	LAX	Directed	4746	13788.0
LAX	SFO	Directed	2786	13390.0
OGG	HNL	Directed	3682	12383.0
LGA	BOS	Directed	2829	12035.0
BOS	LGA	Directed	588	12029.0
HNL	OGG	Directed	2117	12014.0
LAX	LAS	Directed	2753	11773.0
LAS	LAX	Directed	2659	11729.0
LAX	SAN	Directed	2781	11257.0
SAN	LAX	Directed	4513	11224.0

Najwięcej i najmniej przylotów

Lotniskiem z największą sumaryczną liczbą przylotów jest *Port lotniczy Atlanta - Hartsfield-Jackson*(**ATL**) do którego przyleciało aż **414 521** samolotów.

Label	Weighted In-Degree
ATL	414521.0
ORD	350452.0
DFW	281401.0
DEN	241470.0
LAX	215685.0
PHX	199416.0
IAH	185160.0
LAS	172871.0
DTW	162000.0
SFO	140579.0
SLC	139077.0
EWR	138491.0
MCO	130859.0

Lotniskiem z najmniejszą sumaryczną liczbą przylotów jest *Port lotniczy Pueblo Memorial*(**PUB**) do którego nie przyleciał żaden samolot.

Następnym jest *Port lotniczy Cheyenne*(**CYS**) do którego przyleciały 2 samoloty.

Label	Weighted In-Degree
PUB	0.0
CYS	2.0
OGD	2.0
PIR	5.0
TUP	9.0
INL	72.0
BJI	74.0
GST	85.0
ADK	102.0

Najwięcej i najmniej odlotów

Lotniskiem z największą sumaryczną liczbą odlotów jest ponownie *Port lotniczy Atlanta - Hartsfield-Jackson*(**ATL**) do którego przyleciało aż **414 513** samolotów.

Label	Weighted In-Degree	Weighted Out-Degree
ATL	414521.0	414513.0
ORD	350452.0	350380.0
DFW	281401.0	281281.0
DEN	241470.0	241443.0
LAX	215685.0	215608.0
PHX	199416.0	199408.0
IAH	185160.0	185172.0
LAS	172871.0	172876.0
DTW	162000.0	161989.0
SFO	140579.0	140587.0
SLC	139077.0	139088.0
EWB	138491.0	138506.0

Lotniskami z najmniejszą sumaryczną liczbą odlotów są *Port lotniczy Cheyenne*(**CYS**) oraz *Ogden-Hinckley Airport*(**OGD**) z którego nie odleciał żaden samolot, Następnym jest *Port lotniczy Pueblo Memorial*(**PUB**) z którego odleciały 2 samoloty.

Label	Weighted In-Degree	Weighted Out-Degree
CYS	2.0	0.0
OGD	2.0	0.0
PUB	0.0	2.0
PIR	5.0	5.0
TUP	9.0	10.0
BNL	72.0	72.0
BJI	74.0	74.0
GST	85.0	85.0
ADK	102.0	102.0
ACY	113.0	116.0
AKN	116.0	116.0
DLG	116.0	116.0

Średnica sieci

Parametr ten (z ang. *Network diameter*) odzwierciedla największą możliwą liczbę lotów potrzebną do dostania się z dowolnego lotniska na inne.

Największą liczbę lotów żeby dostać się na najbardziej oddalone lotnisko potrzebuje *Port lotniczy Pueblo Memorial*(**PUB**) który potrzebuje aż 5 lotów."

Label	Eccentricity
PUB	5.0
TPA	4.0
FLL	4.0
LGA	4.0
SAT	4.0
RDU	4.0
MSY	4.0
PIT	4.0

Dla porównania *Port lotniczy Seattle-Tacoma* (**SEA**) potrzebuje dwóch przesiadek, aby tego dokonać. Dwie pierwsze wiersze mają wartość 0, ponieważ nie ma możliwości z nich wylecieć

Label	Eccentricity
CYS	0.0
OGD	0.0
SEA	2.0
ATL	3.0
ORD	3.0
DFW	3.0
DEN	3.0
MSP	3.0
DTW	3.0
SLC	3.0
IAH	3.0

Centralność bliskości

Parametr ten (z ang. *Closeness centrality*) mierzy średnią odległość od jednego wężła do dowolnego innego wężła. Im bardziej centralny jest węzeł, tym bliższy jest dla wszystkich innych węzłów.

W ten sposób dostajemy informację, że najbardziej scentralizowanym lotniskiem jest *Port lotniczy Atlanta-Hartsfield-Jackson* (**ATL**).

Label	Eccentricity	Closeness Centrality
ATL	3.0	0.691781
ORD	3.0	0.655844
DFW	3.0	0.63522
DEN	3.0	0.626033
MSP	3.0	0.626033
DTW	3.0	0.61336
SLC	3.0	0.610887
IAH	3.0	0.609658
CVG	3.0	0.607214
LAS	3.0	0.582692
LAX	3.0	0.581574
PHX	3.0	0.57935

Pośrednictwo bliskości

Pośrednictwo (z ang. *Betweenness centrality*) mierzy liczbę najkrótszych ścieżek, na których znajduje się węzeł. W tej sytuacji również *Port lotniczy Atlanta - Hartsfield-Jackson (ATL)* znajduje się na samym szczycie listy z ponad **18937** ścieżkami. Blisko dwukrotnie mniejszą wartość posiada Port lotniczy Salt Lake City.

Label	Eccentricity	Closeness Centrality	Harmonic Closeness Centrality	Betweenness Centrality
ATL	3.0	0.691781	0.782728	18937.277612
SLC	3.0	0.610887	0.685919	10398.296844
MSP	3.0	0.626033	0.705721	9661.426559
DFW	3.0	0.63522	0.718372	9575.040146
ORD	3.0	0.655844	0.743124	7746.750695
ANC	3.0	0.519726	0.543454	6593.838165
DEN	3.0	0.626033	0.706821	6112.032336
DTW	3.0	0.61336	0.691419	5867.652973
IAH	3.0	0.609658	0.685369	5421.970544
LAX	3.0	0.581574	0.645765	4170.73262
SFO	3.0	0.564246	0.619362	4146.359635

Gęstość grafu

Gęstość grafu (z ang. *Graph density*)- jest miarą tego, ile istnieje powiązań między węzłami w porównaniu do tego, ile powiązań między węzłami jest możliwych. Ilość krawędzi przez możliwe krawędzie między wszystkimi wierzchołkami.

W badanym grafie - **Graph density = 0,058**.

Modułowość

Modułowość jest miarą struktury sieci lub wykresów, która mierzy siłę podziału sieci na moduły (zwane również grupami, klastrami lub społecznościami). Sieci o dużej modułowości mają gęste połączenia między węzłami w modułach, ale rzadkie połączenia między węzłami w różnych modułach.

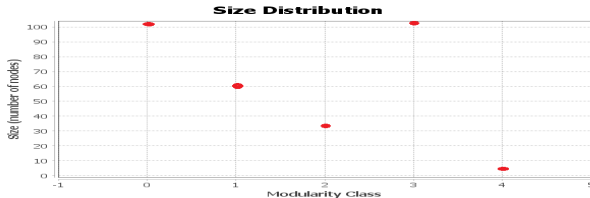
Modularity Report

Parameters:

Randomize: On
Use edge weights: On
Resolution: 1.0

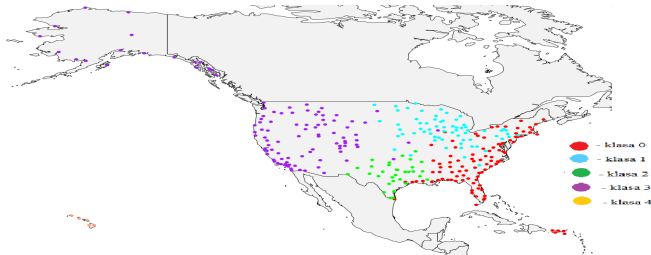
Results:

Modularity: 0,296
Modularity with resolution: 0,296
Number of Communities: 5



Opis społeczności

Wierzchołki badanej sieci można podzielić na 5 społeczności ze względu na regiony:



```
modularity <- read.csv("tabela.csv")
lotniskamodularity <- merge(lotniska, modularity, by.x="lata", by.y="Id")
modularity0 <- filter(lotniskamodularity, lotniskamodularity$modularity_class=="0")
modularity1 <- filter(lotniskamodularity, lotniskamodularity$modularity_class=="1")
modularity2 <- filter(lotniskamodularity, lotniskamodularity$modularity_class=="2")
modularity3 <- filter(lotniskamodularity, lotniskamodularity$modularity_class=="3")
modularity4 <- filter(lotniskamodularity, lotniskamodularity$modularity_class=="4")

map("world", col="#f2f2f2", fill=TRUE, bg="white", lwd=0.1, xlim=xlim, ylim=ylim)
points(modularity0$longitude, modularity0$latitude, pch=16, cex=0.5, col="red")
points(modularity1$longitude, modularity1$latitude, pch=16, cex=0.5, col="cyan")
points(modularity2$longitude, modularity2$latitude, pch=16, cex=0.5, col="green")
points(modularity3$longitude, modularity3$latitude, pch=16, cex=0.5, col="purple")
points(modularity4$longitude, modularity4$latitude, pch=16, cex=0.5, col="lightsalmon")
```

Rysunek: Kod w R generujący powyższą mapę

Dziękujemy za uwagę!