

# Network Analysis - Weltweiter Schiffsverkehr

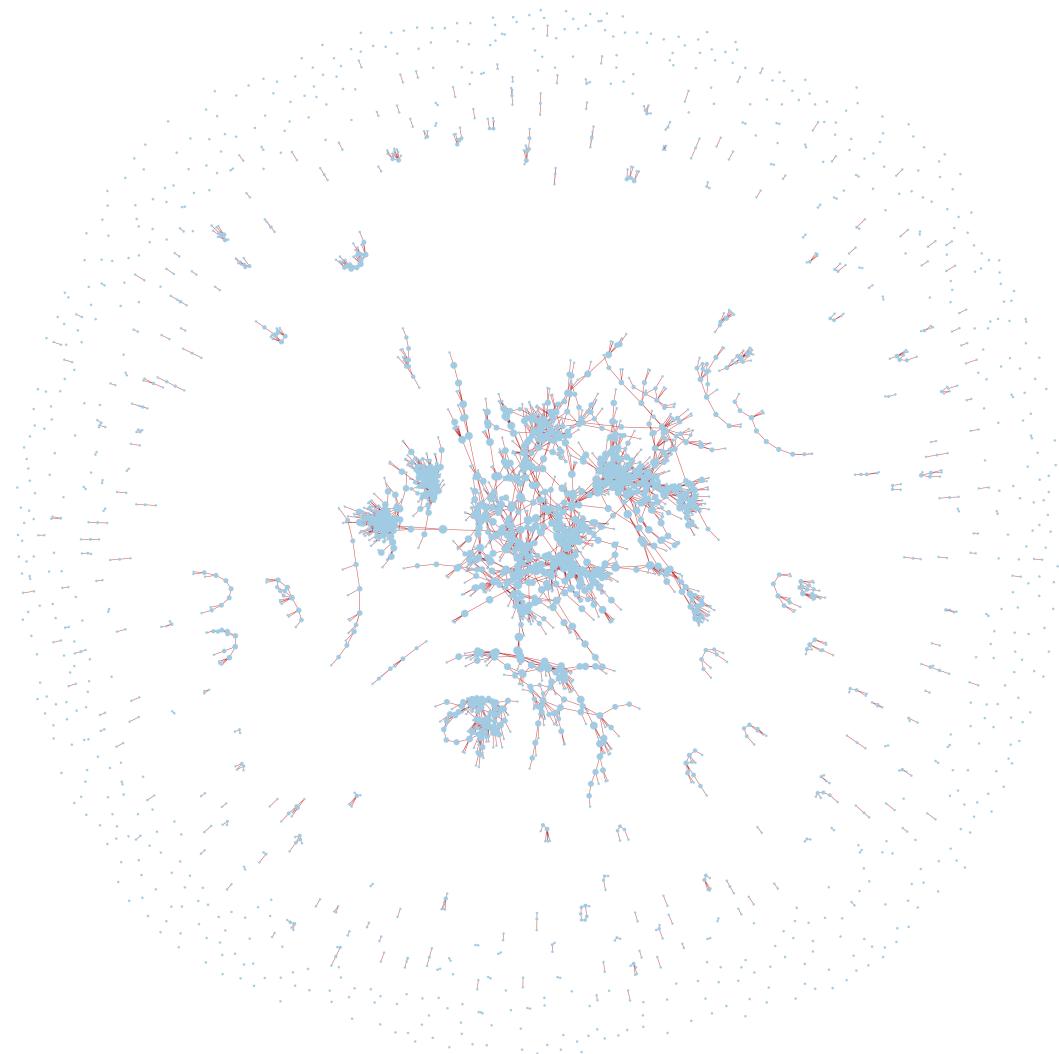
Github repository: [https://github.com/Beniiii/sna\\_shipTraffic](https://github.com/Beniiii/sna_shipTraffic) ([https://github.com/Beniiii/sna\\_shipTraffic](https://github.com/Beniiii/sna_shipTraffic))

## Ungerichteter Graph inkl. Betweenness Zentralität

Dieser Plot zeigt das ganze Netzwerk als ungerichteter Graph auf. Die Nodegrösse wird anhand der Betweenness Zentralität bestimmt.

**Analyse:** Alle Nodes, welche aussen platziert sind und keine Verbindung (roter Strich) zu einem anderen Node haben, visualisieren die Häfen, welche keine Fahrten zu anderen Häfen haben. Das können zum Beispiel Fischerhäfen sein. In der Mitte der Grafik sind die Häfen abgebildet, welche am meisten Verbindungen zu anderen Nodes haben. Grundlegend kann gesagt werden, dass rund die Hälfte aller Häfen nur mit sich oder mit einzelnen anderen Nodes verknüpft sind. Die andere Hälfte dagegen haben Verbindungen zu mehreren anderen Häfen, wobei auch viele der stark verbundenen Häfen eine hohe Betweenness-Zentralität haben.

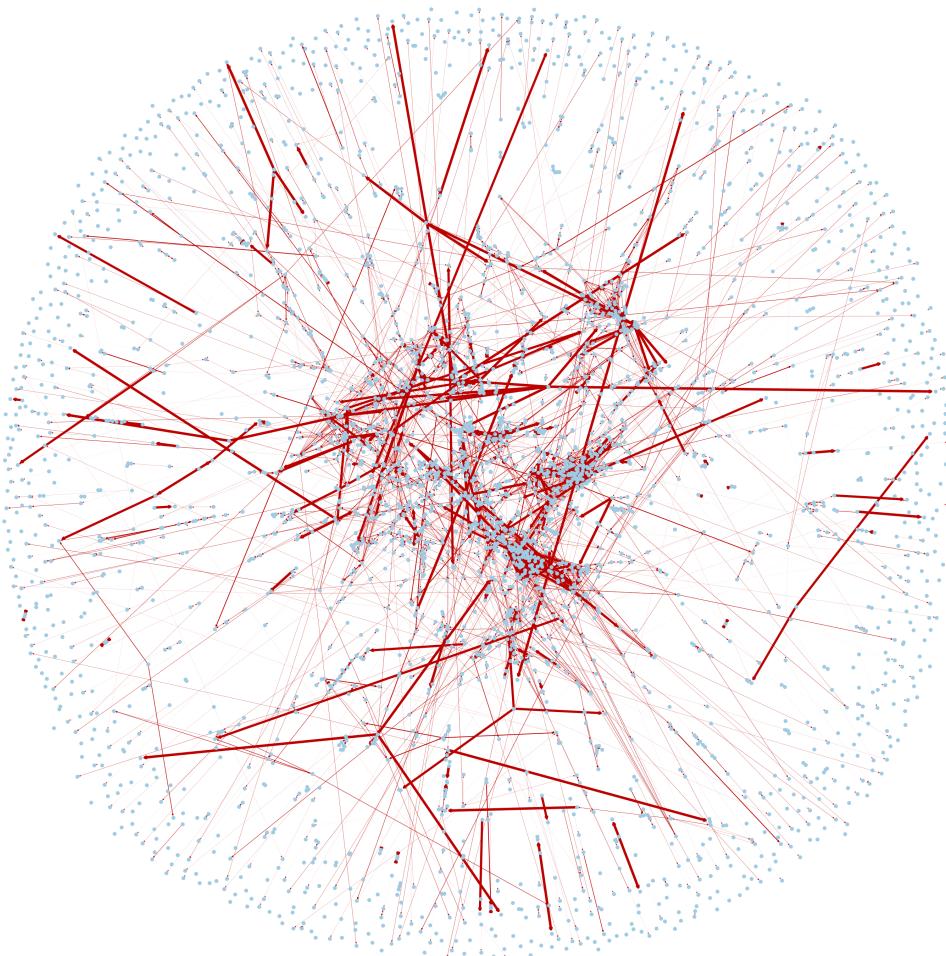
Das gesamte Netzwerk weist 3565 Nodes und 98292 Edges auf. Die Nodes repräsentieren die Häfen und Edges sind die Fahrten zwischen zwei Häfen.



## Gerichteter Graph

Dieser Plot zeigt das ganze Netzwerk als gerichteter Graph auf. Die Dicke der roten Linie zeigt das Gewicht der Kante an.

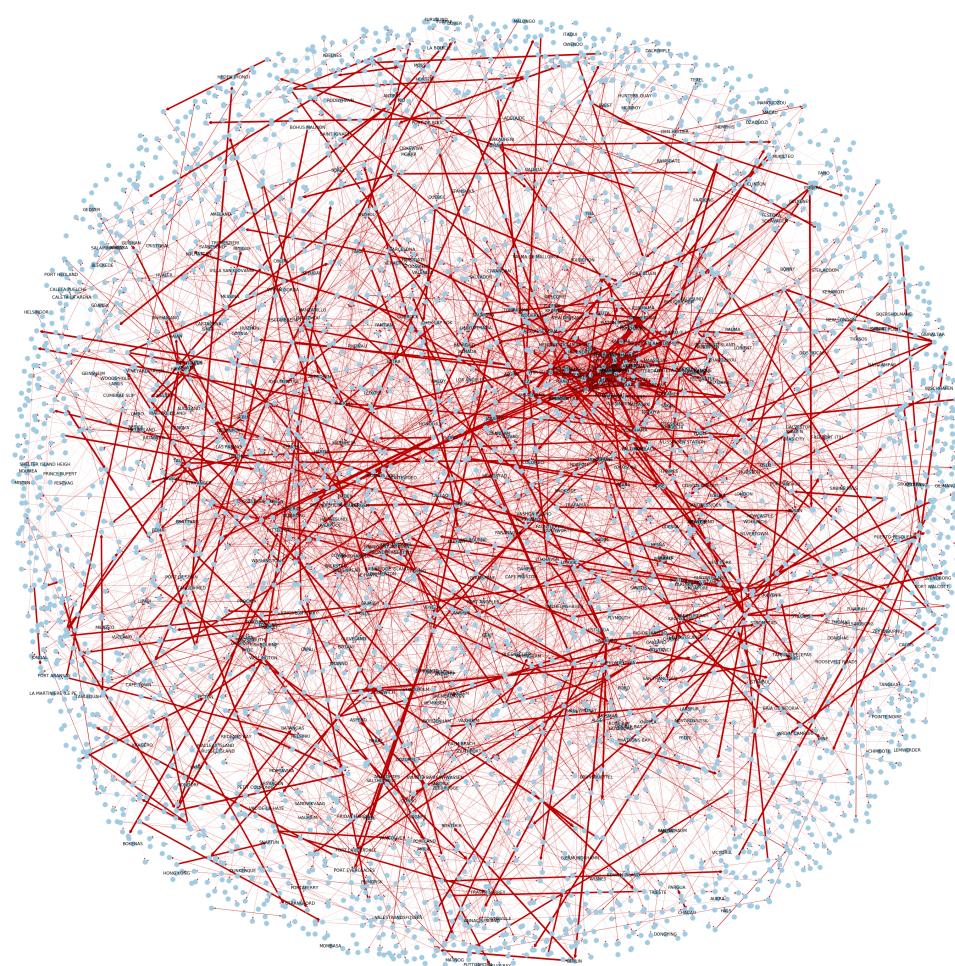
**Analyse:** Im Gegensatz zum ungerichteten Graphen werden Pfeile angezeigt, welche die Fahrtrichtung der Schiffe visualisieren. Je zentraler ein Hafen ist, desto höher ist die Wahrscheinlichkeit, dass die Schiffe in beide Richtungen verkehren. Das gilt auch für das Gewicht der Kanten. Es ist auch erkennbar, dass Schiffe von eher zentralen Häfen nur in eine Richtung zu Häfen fahren, welche aussen positioniert sind. Das können zum Beispiel Insel-Häfen sein, welche hauptsächlich Waren importieren und nichts exportieren, oder solche Häfen, welche nur als Zwischenstopp dienen. Bei Verbindungs-Häfen mit einem hohen Gewicht kann es sich um Passagierschiffe oder Fähren handeln, welche mehrere Mal pro Stunde Hin- und zurück fahren.



## Gerichteter Graph mit Labels

Dieser Plot zeigt das ganze Netzwerk als gerichteter Graph auf. Zusätzlich werden die Labels der Häfen angezeigt, welche einen höheren gewichteten Degree als 50 haben.

**Analyse:** Es wurden nur die Labels der Häfen angezeigt, welche einen hohen gewichteten Degree aufweisen. Sonst wäre der Graph schlicht zu überfüllt und unübersichtlich. Sehr auffällig ist nun, dass die Nodes der stark gewichtete Kanten oft nicht beschriftet sind, was bedeutet, dass diese nicht sehr viele Verbindungen zu anderen Nodes haben. So wird die oben gestellte Hypothese bestätigt, dass solche Verbindungen hauptsächlich von Passagierschiffen oder Fähren befahren werden, welche in einem kleinen Intervall hin- und zurückfahren. Im Zentrum des Graphs ist ein Cluster von holländischen Häfen wie Rotterdam und Amsterdam zu sehen. Diese Verbindungen führen wahrscheinlich alle durch den Ärmelkanal.

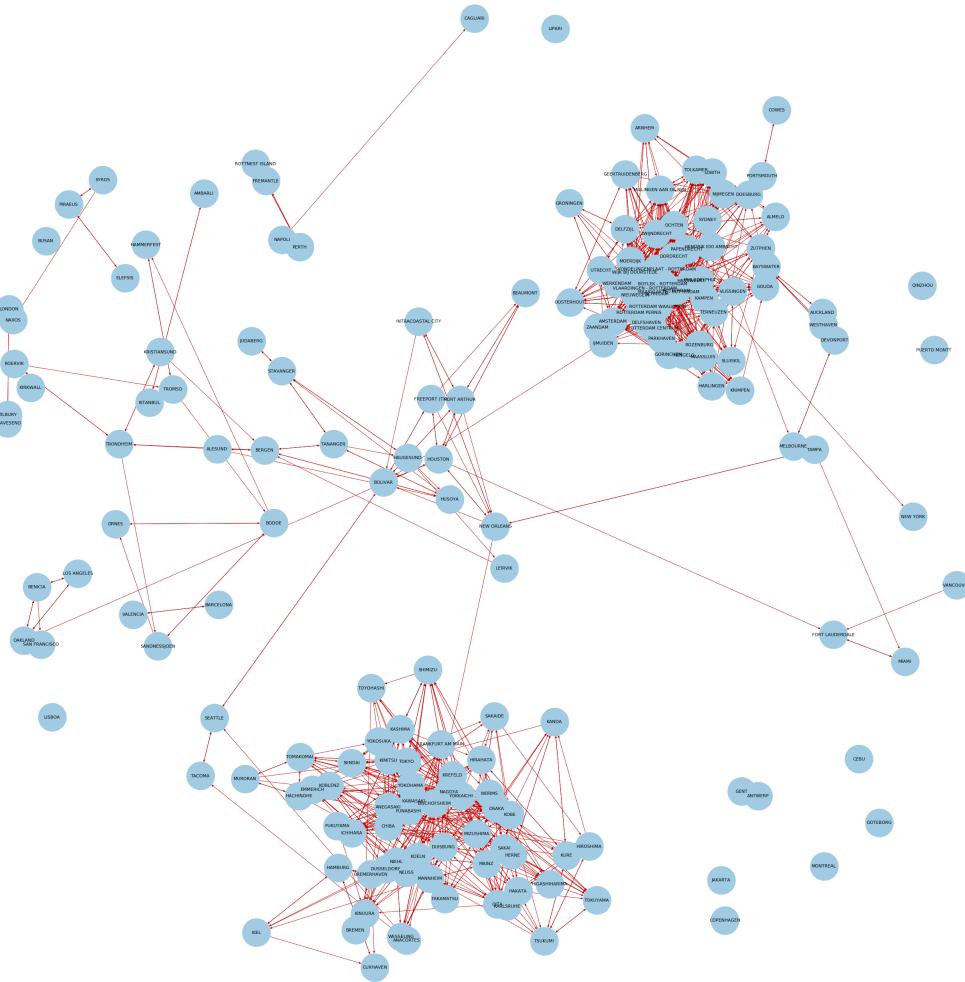


# Degree Zentralität

## Häfen mit den höchsten Degree Zentralitäten

Dieser Plot zeigt die Verbindungen der Häfen, welche einen höheren Degree als 15 haben.

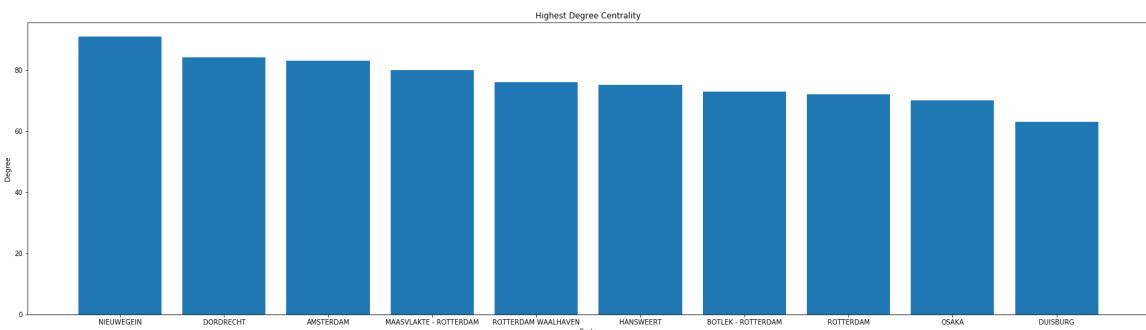
**Analyse:** Was beim gerichteten Graph mit Labels bereits erkannt wurde, wird hier nochmals verstärkt, denn der Cluster der holländischen Häfen ist nun besser ersichtlich. Weitere Clusters der Häfen mit hohen Degrees gibt es bei den Häfen in Japan und Deutschland.



## Top 10 Häfen mit den höchsten Degree Zentralitäten

Dieser Bar-Plot zeigt die 10 Häfen auf, welche die höchsten Degree Zentralitäten aufweisen.

**Analyse:** Bei den Top 10 Häfen, sind 8 von 10 Häfen in Holland platziert. Dies zeigt auf, dass Holland die meisten Verbindungen zu anderen Häfen haben. Spannend ist auch, dass 4 verschiedene Häfen in Rotterdam in den Top 10 Häfen sind. So ist Rotterdam die mächtigste Hafenstadt, gefolgt von Nieuwegein.

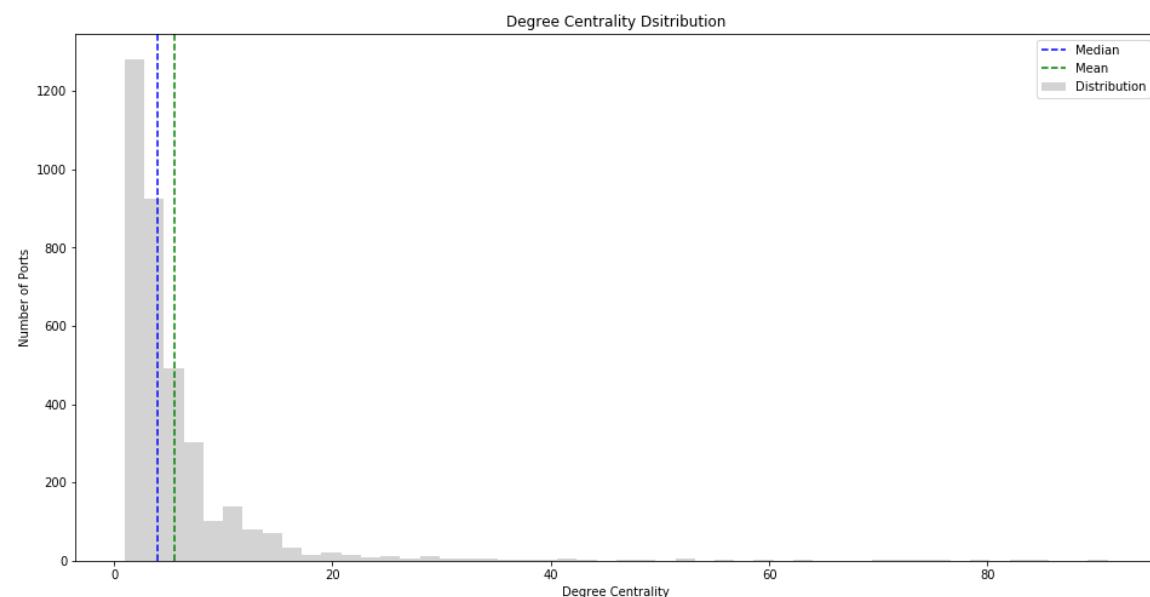


## Degree Zentralitäts-Verteilung

Dieser Bar-Plot zeigt die Verteilung der Degree-Zentralitäten auf.

**Analyse:** Der Durchschnitt aller Degree-Zentralitäten liegt bei 5.59. Die meiste Degree-Zentralität liegt zwischen 2-4, wo auch der Median liegt. Mehr als 20 Verbindungen haben nur noch wenige Häfen, welche sich aber bis zu einer Degree-Zentralität von 100 ausdehnen.

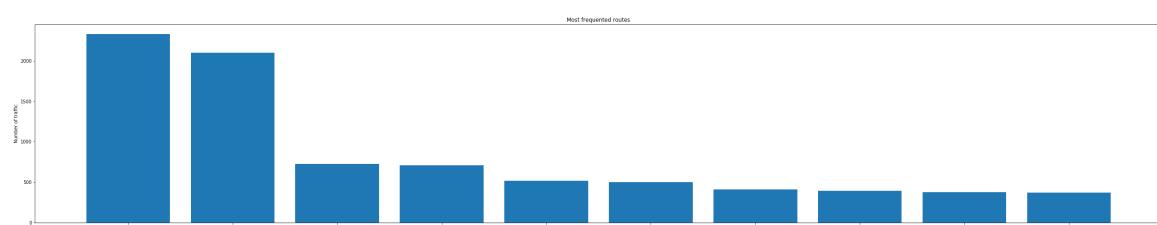
median 4.0  
mean 5.595511921458625



## Meist befahrene Strecken

Dieser Bar-Plot zeigt die meist befahrenen Verbindungen zwischen zwei Häfen.

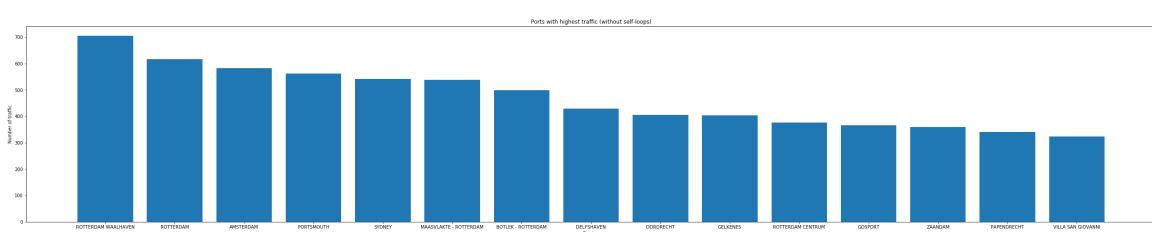
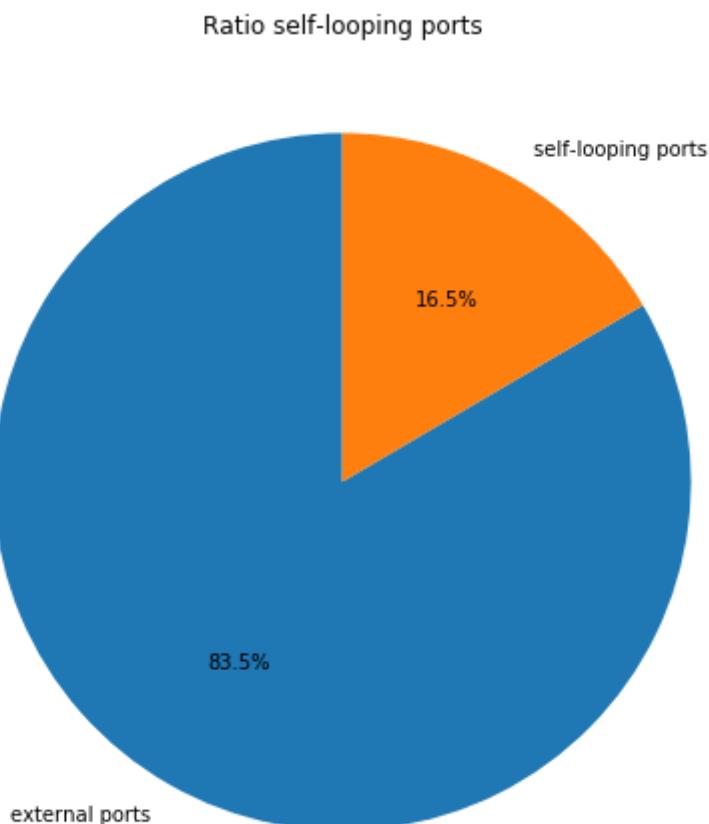
**Analyse:** Die meist befahrenen Strecken haben den selben Start- und Endhafen, was auch wieder auf Personenschiffe und Fähren zu führen ist.



# Häfen mit den meisten externen Verbindungen

Der Pie-Chart zeigt das Verhältnis zwischen Häfen, welche externe Verbindungen haben mit den Häfen, welche keine externen Verbindungen haben. Der Bar-Plot zeigt die Häfen an, welche am meisten externe Verbindungen haben (ohne self-loops).

**Analyse:** Mehr als 16% der Häfen sind nicht mit anderen Häfen verbunden und stellen ein eigenen Component dar. Beim Bar-Plot wurden die Top 10 Häfen dargestellt, welche am meisten externe Verbindungen haben, was bedeutet, dass die Self-loops ausgeschlossen wurden. So werden Strecken wie Singapore-Singapore nicht mehr einberechnet, sondern nur solche welche zu einem anderen Hafen führen. Auch hier erkennt man auch wieder die Wichtigkeit der holländischen Häfen.



# Ärmelkanal

Geplant war die Analyse der Häfen im Ärmelkanal. Jedoch haben wir ziemlich schnell gemerkt, dass diese keine gute Analyse-Grundlagen mit sich bringen. Die Häfen im Ärmelkanal selber sind gar nicht stark befahren. Wir haben uns aus diesem Grund für die Analyse aller Häfen entschieden und sind dieser Fragenstellung nicht mehr genauer nachgegangen. Unten werden die Anzahl Häfen, zu welchen eine Verbindung besteht, gezählt (Degree-Zentralität). Es ist schnell ersichtlich, dass die Häfen im Ärmelkanal nicht stark verbunden sind.

```
ROSCOFF 0
DOVER 0
TEIGNMOUTH 0
FALMOUTH 1
FECAMP 1
CHERBOURG 1
CAEN 2
PORTLAND UK 4
OUISTREHAM 5
PORTSMOUTH ANCH 5
LE HAVRE 10
SOUTHAMPTON 12
PORTSMOUTH 17
```

Es gibt 17 ,in unserem Datensatz nicht gefundenen Häfen, welche einen anderen Namen haben oder keine Verbindung besteht. Man müsste die Daten genauer analysieren.

Auch wurde festgestellt, dass es mehrere Häfen in unterschiedlichen Ländern gibt, welche im Datensatz ähnlich benannt wurden: PORTSMOUTH (NH) != PORTSMOUTH (UK) != PORTSMOUTH (FR). Auch diese müsste man separat analysieren.

# Clusters

Durchschnittliches Clustering des Netzwerks (gerichtet): 0.163857136357137

5

Durchschnittliches Clustering des Netzwerks (ungerichtet): 0.1924660196085

192

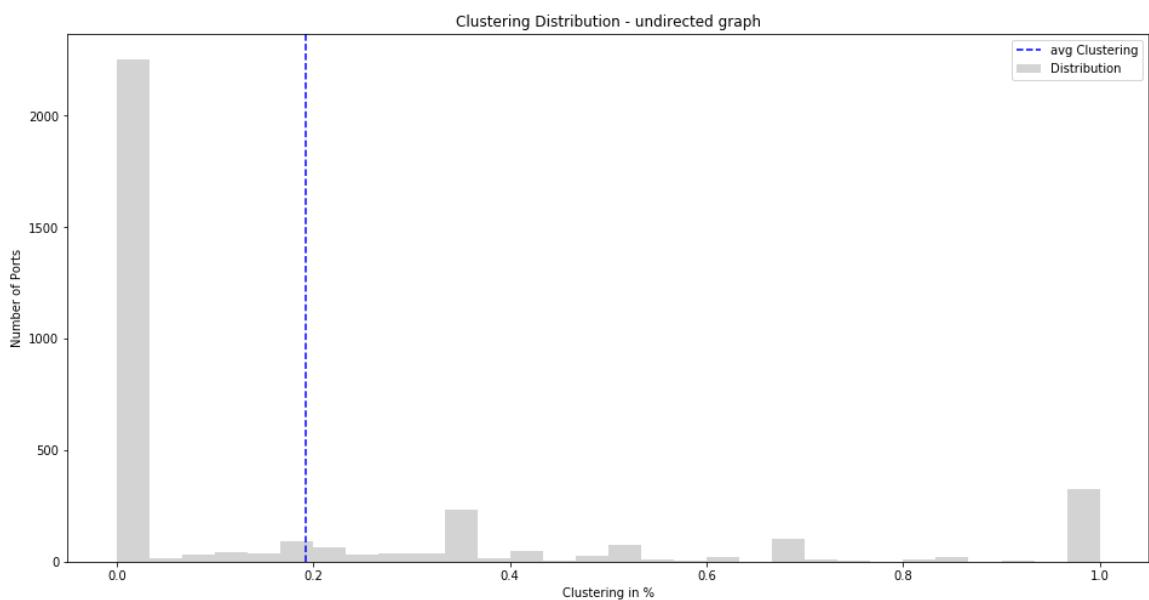
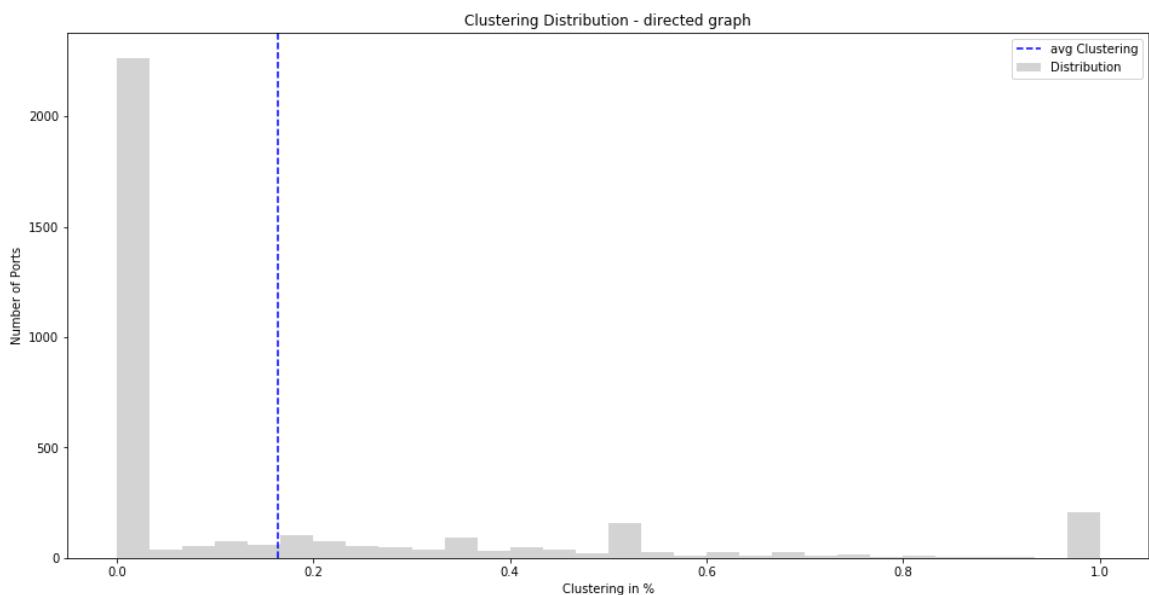
Diese tiefen Clustering-Werte bedeuten, dass das Netzwerk mehrere Sub-Netzwerke hat, welche nicht miteinander verbunden sind oder Häfen, welche nicht in Verbindung mit anderen stehen.

## Clustering Verteilung

Die Histogramme zeigen die Verteilung der Clusterings von gerichteten und ungerichteten Graphen auf.

**Analyse:** Ein grosser Anteil der Häfen gehört zu keinen oder sehr wenigen Clusters. Diese Verteilung variiert ziemlich stark. Es ist auch ersichtlich, dass es doch ca. 200 Häfen gibt, welche sehr stark vernetzt sind. Die beiden Graphen unterscheiden sich nur minim.

Durchschnitts-Clustering: 0.1638571363571375



## Components

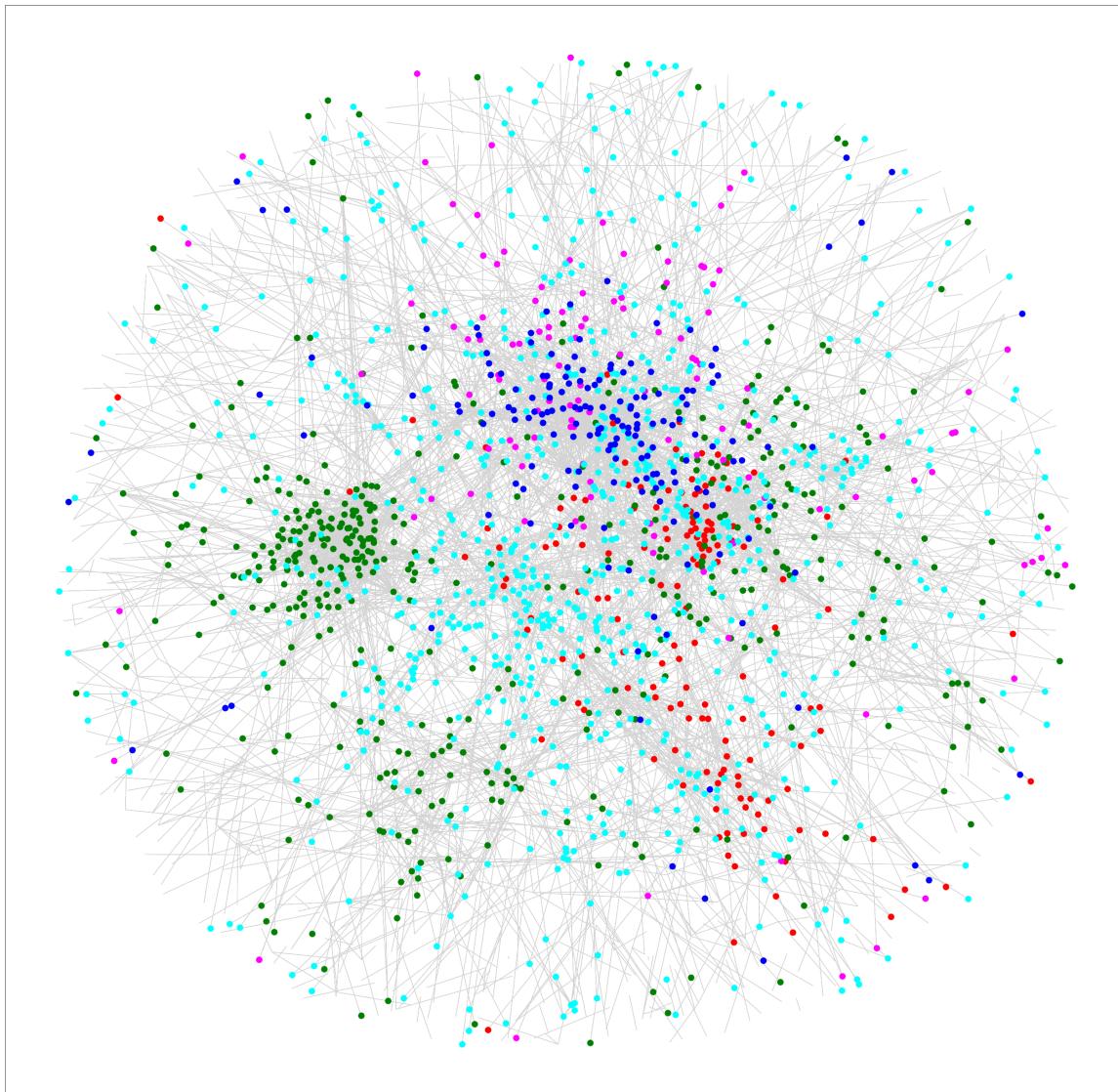
Anzahl der unterschiedlichen Components: 876

### Grössten Components

Das Netzwerk visualisiert die Components, welche mehr als 100 verbundene Nodes besitzen.

**Analyse:** Da es eine grosse Anzahl Components gibt, sind viele Häfen nur von gewissen Häfen erreichbar und völlig abgeschottet von anderen Components. Das heisst, dass Häfen von anderen Components nicht erreichbar sind. Der grösste Component beinhaltet 759 Häfen, der 5. grösste jedoch nur noch 121. Alle anderen umfassen weniger als 100 Nodes, was bedeutet, dass es viele Häfen in sehr kleinen Components gibt. Diese Aussage passt auch zu deren vom ersten Graph.

```
size red component: 150
size green component: 448
size magenta component: 121
size cyan component: 759
size blue component: 174
```



Nun ist auch ziemlich klar, dass der Graph weder stark noch schwach verbunden ist:

Ist der Graph Strongly Connected? False  
Ist der Graph Weakly Connected? False

## Kürzeste Wege

Die kürzesten Wege kann mit NetworkX einfach herausgefunden werden. Dabei werden hier zufällige Häfen ausgewählt und den Weg sowie die Anzahl der Hops gezeigt. Er ist erstaunlich, dass es keine direkteren Wege von Miami nach Melbourne gibt. Mit den vorhandenen Daten (von 3 Tagen) wäre ein Zwischenstop in England nötig. Auch kann man sehen, dass es bei gewissen Ländern einen "Haupthafen" gibt, welche Schiffe oft zuerst anfahren. Das wäre nun in diesem Beispiel von Australien der Hafen Brisbane.

Weg: ['KETAPANG', 'GILIMANUK']

Anzahl Hops: 1

Weg: ['MIAMI', 'FORT LAUDERDALE', 'WILMINGTON', 'SOUTHPORT', 'BRISBANE',  
'BOTANY', 'MELBOURNE']

Anzahl Hops: 6

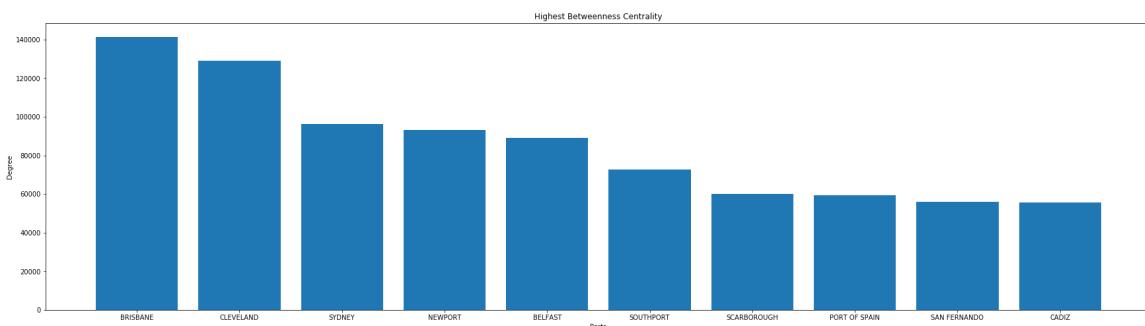
Weg: ['MIAMI', 'FORT LAUDERDALE', 'WILMINGTON', 'SOUTHPORT', 'BRISBANE',  
'SYDNEY']

Anzahl Hops: 5

## Betweenness Zentralität

Die Betweenness-Zentralität wurde bereits im ersten Graph visualisiert. Hier wird zusätzlich ein Bar-Plot angezeigt, welche die grössten Betweenness Zentralität haben.

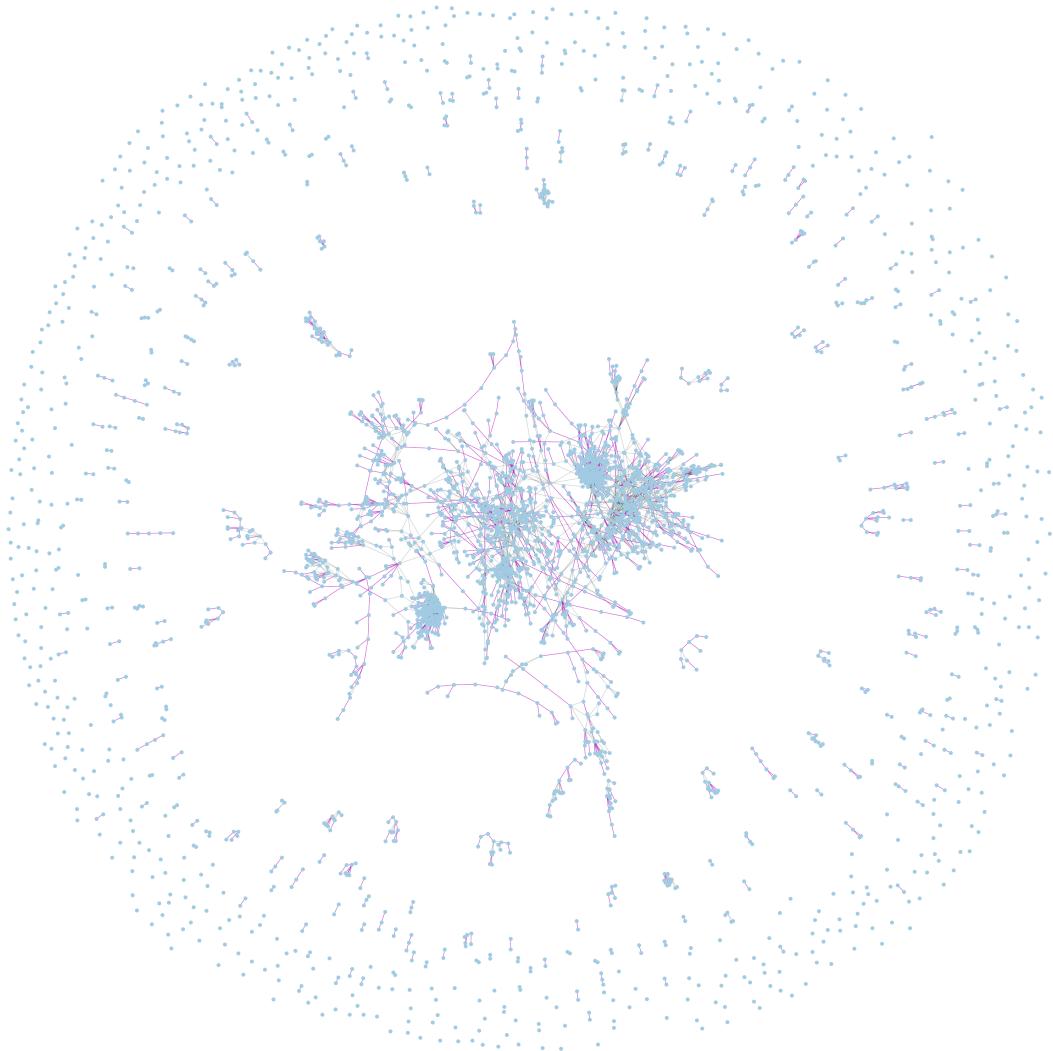
**Analyse:** Ein hoher Betweenness-Zentralitätswert bedeutet, dass viele Häfen von einem Port mit dem hohen Betweenness-Wert abhängig sind. Wenn der Hafen nicht mehr angefahren wird, sind viele andere auch nicht mehr erreichbar. Wir haben den Hafen Port of Spain, welcher in Trinidad and Tobago liegt genauer untersucht. Es handelt sich hier um einen Inselstaat in der Karibik. Obwohl die Insel nur etwa 1/8 der Fläche von der Schweiz umfasst, verkehren viele Schiffe über diesen Hafen. Die Güter werden dann intern auf die Inseln der Karibik verteilt.



## Bridges

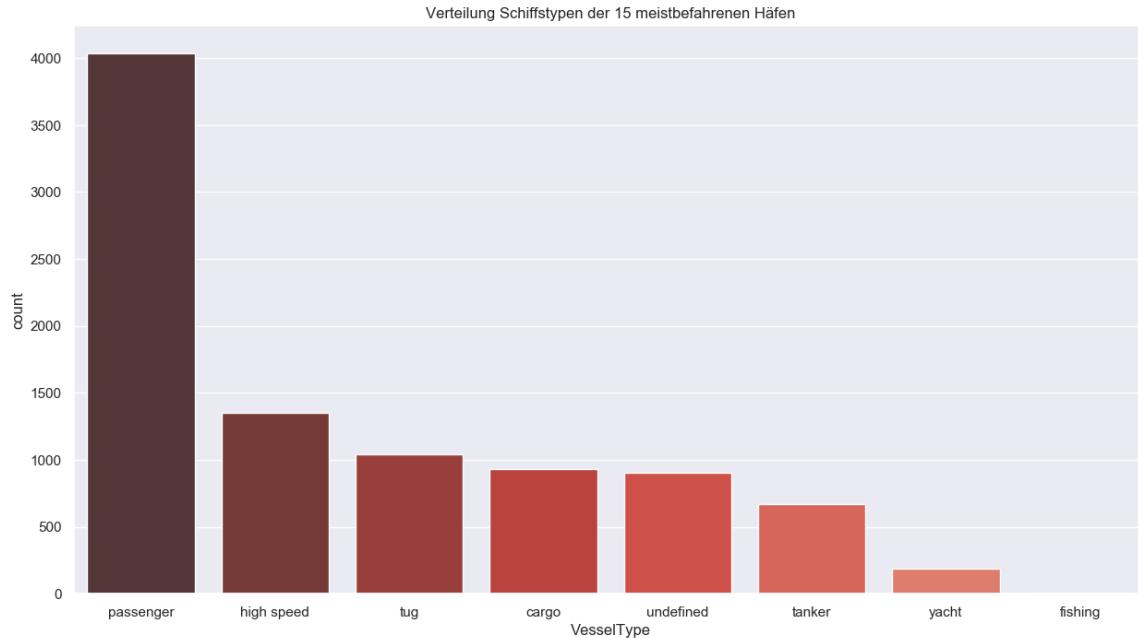
Das Netzwerk visualisiert die verschiedenen Bridges, welche violet eingefärbt sind. Edges, welche keine Bridge werden schwarz angezeigt.

**Analyse:** Erstaunlich ist, dass es viele Components gibt, welche nur aus Bridges bestehen. Das heisst, dass diese Components untereinander nicht stark vernetzt sind, sondern jede weitere Verbindung von nur einem einzigen Hafen abhängt. Es ist ca. jede 6. Edge eine Bridge, was uns als eher hoch erscheint.

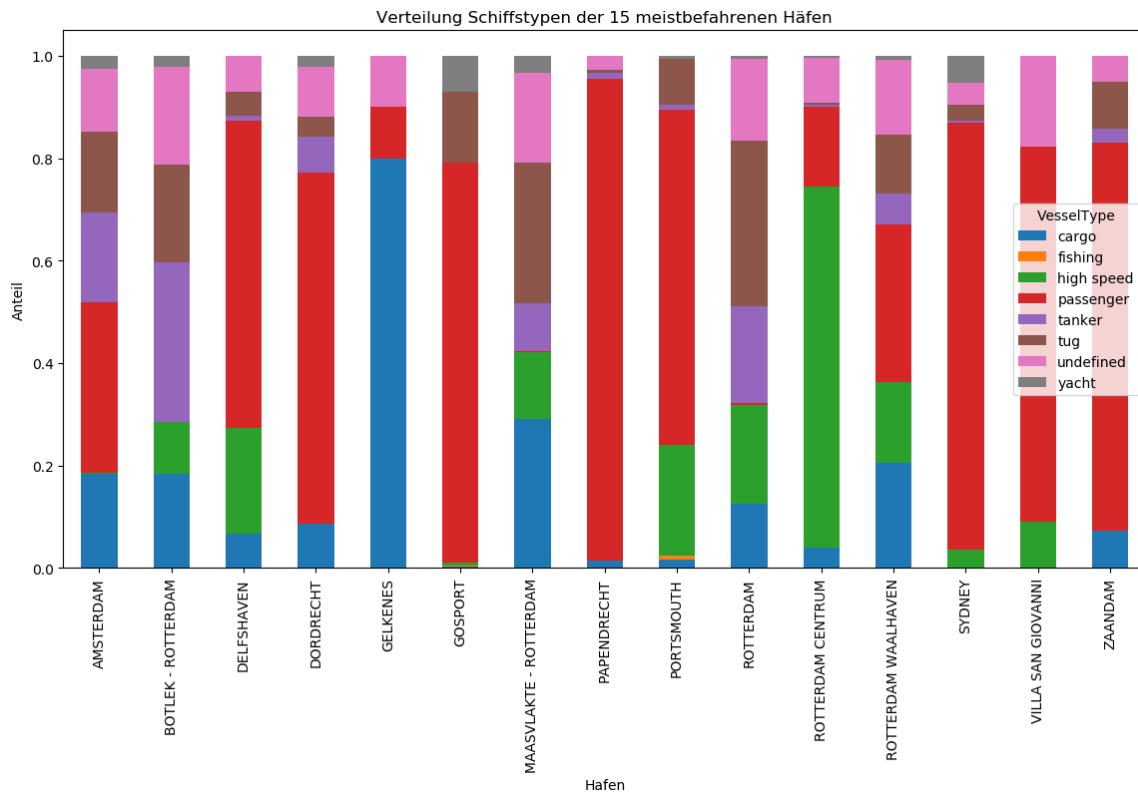


Bridges-Anteil in Prozent: 15.58559729624334 %

## Schiffstypen



Der Plot zeigt, die Verteilung der Schiffstypen in den 15 meist befahrenen Häfen auf. Es ist auffällig, dass fast die Hälfte dabei Personen transportieren.



Der Plot zeigt, wie die Verteilung der Ankünfte der Schiffstypen auf die verschiedenen der 15 meist befahrenen Häfen ist. Cargo Schiffe gibt es vor allem in Gelkenes, aber auch viele in den verschiedenen Häfen von Rotterdam und Amsterdam wo ausserdem ein Grossteil der Tanker anlegen. Viele Ankünfte mit Passagieren gibt es in Delfshaven, Dordrech, Gosport, Papendrecht, Portsmouth, Sydney, Villa San Giovanni und Zaandam, wobei einige Häfen fast ausschliesslich von Passagierschiffen befahren werden. Rotterdam Centrum beherbergt vor allem High Speed Boote. Die Fischerboote sind vor allem in Portsmouth unterwegs.

## Lessons Learned & Ausblick

- Das Einarbeiten in NetworkX hat uns am Anfang viel Zeit gekostet. Nach den ersten Plots wurde uns aber klar, wie man die Library benutzt muss. Die Dokumentation ist sehr umfangreich und hat uns dank den Beispielen sehr geholfen.
- Da wir mit einer sehr grossen Datenmenge arbeiten, ist die Wartezeit der Erstellung eines Graphens zum Teil eher lange.
- Leider konnten wir die Daten nicht wirklich Filtern, da wir die Geodaten oder Länder der Häfen nicht zur Verfügung hatten. Im nächsten Schritt würden wir probieren die Länder mit den Häfen mit Hilfe eines zusätzlichen Datensatzes zu mappen, um länderspezifische Analysen zu machen.