# Railway system "PG express"

For our project we chose railway system which describes regional communications between cities in Pomorskie voivodeship. We chose only biggest and most visited train stations.

For relationships we used such attributes as: Travelling time, ticket price, distance, train type and maximum passengers' number. All of them were used to answer analytical questions listed below.

For nodes we used: maximum human throughput, amount of platforms, maximum train throughput, passengers opinion about station, information about ticket office and covid zone. Some of the attributes were specifically used for above analysis.
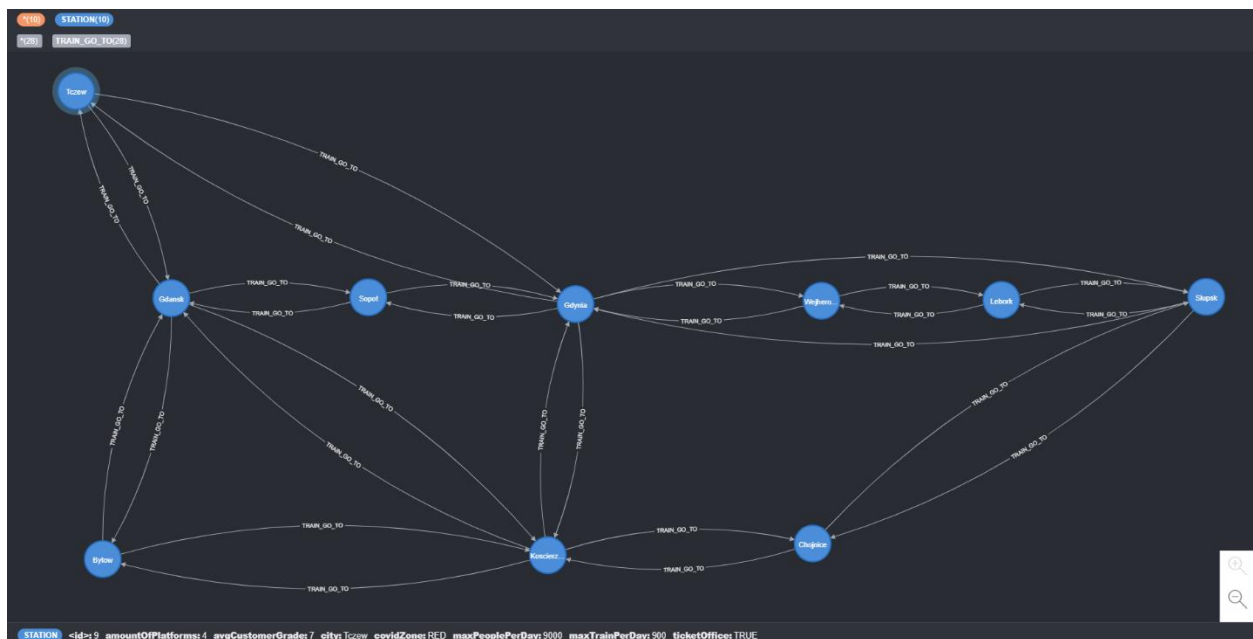
Competency questions

1. List all the stations to which you can reach from Gdynia.

city_related_Gdynia
1 "Slupsk"
2 "Wejherowo"
3 "Tczew"
4 "Koscierzyna"
5 "Sopot"

2. List cities that can be reached both from Gdynia and Gdansk.

station.city
1 "Sopot"
2 "Koscierzyna"
3 "Tczew"

3. List all stations from which you can get to Wejherowo with the possibility of one change.

	station.city
1	"Slupsk"
2	"Wejherowo"
3	"Lebork"
4	"Tczew"
5	"Koscierzyna"
6	"Sopot"
7	"Gdynia"

4. Show the route with the least amount of changes from Tczew to Gdańsk.
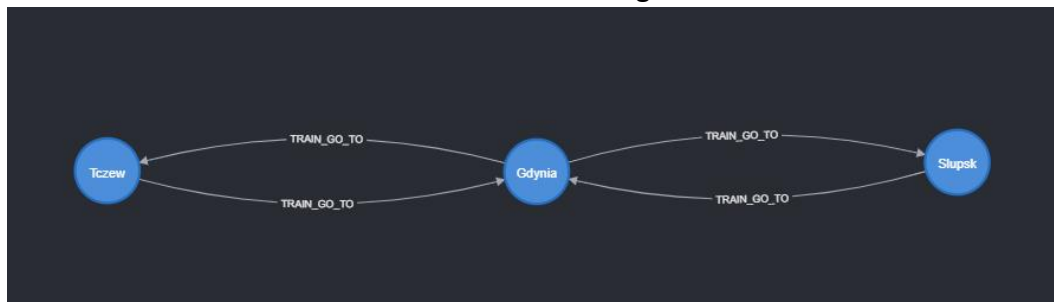
5. Check which route is the most optimal.

| express_cost	expressTime	viaCity	cityTime
12	40	12	47

6. List those stations that have ticket offices, the number of maxPeoplePerDay more than 5000 and rank according to the ratings of the people (from best to lowest).

n.city	n.avgCustomerGrade
1	"Sopot"	8
2	"Gdynia"	7
3	"Tczew"	7
4	"Gdansk"	6
5	"Wejherowo"	5

7. List the city to which we can go from Gdynia with 1 change, not being in the red zone no closer than 100 km but the route should be shorter than 150 minutes.

transferCity	targetCity	distance	time
"Koscierzyna"	"Bytow"	113	106

8. Show all the routes that we can get from Tczew to Gdynia in up to three changes.
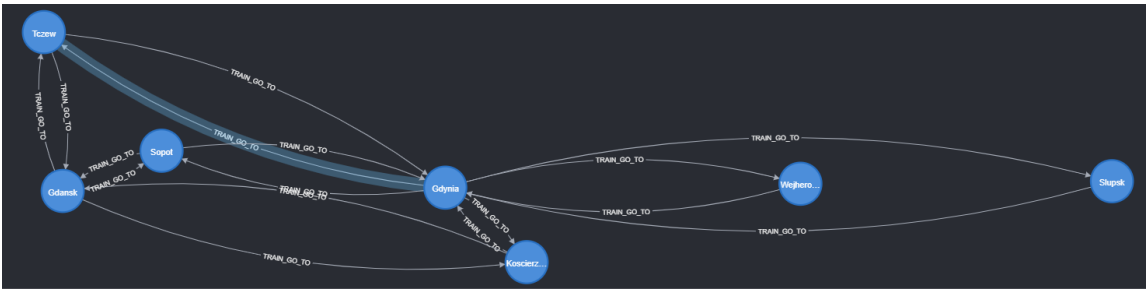
9. List the 5 most expensive sections of the route(per km).

startStation	finishStation	costOneKm
1	"Slupsk"	"Gdynia"	84
2	"Gdynia"	"Slupsk"	84
3	"Chojnice"	"Slupsk"	75
4	"Slupsk"	"Chojnice"	75
5	"Chojnice"	"Koscierzyna"	62

10. List the three stations with cash registers the least number of connections.

n.city	amountOfTrack
1 "Wejherowo"	2
2 "Sopot"	2
3 "Chojnice"	2

Analysis

1. The Random Walk Theory

According to the latest government guidelines, we should track people with covid and try to predict which stations they could reach. In order to fulfill these requirements, we decided to use this algorithm to detect three cities that could be visited by a sick person with covid. We chose the city by random.

city
1 "Gdynia"
2 "Wejherowo"
3 "Gdynia"
4 "Tczew"

2. Number of triangles and clustering coefficient

We use triangles and clustering coefficient algorithm to check traffic density to reduce costs. According to clustering coefficient our stations are not strongly so we don't have to reduce any traffic.

- Count of triangles

name	triangleCount
1 "Gdansk"	1
2 "Koscierzyna"	1
3 "Bytow"	1
4 "Gdynia"	0
5 "Sopot"	0
6 "Wejherowo"	0
7 "Lebork"	0
8 "Slupsk"	0
9 "Chojnice"	0
10 "Tczew"	0

- Local clustering coefficient

	name	localClusteringCoefficient
1	"Bytow"	1.0
2	"Gdansk"	0.16666666666666666
3	"Koscierzyna"	0.16666666666666666
4	"Gdynia"	0.0
5	"Sopot"	0.0
6	"Wejherowo"	0.0
7	"Lebork"	0.0
8	"Slupsk"	0.0
9	"Chojnice"	0.0
10	"Tczew"	0.0

- Average clustering coefficient

averageClusteringCoefficient
0.13333333333333336

3. Closeness centrality

In our company we decided to build a special warehouse for trains from which we can send the train to all stations in the shortest time. After brainstorming with our managers, we decided to build it in the town which has the shortest path to another city so we used it for its closeness centrality. Below you can find the values of centrality.

	city	centrality
1	"Gdynia"	0.6923076923076923
2	"Koscierzyna"	0.6
3	"Slupsk"	0.5294117647058824
4	"Gdansk"	0.5
5	"Sopot"	0.5
6	"Tczew"	0.5
7	"Wejherowo"	0.47368421052631576
8	"Chojnice"	0.47368421052631576
9	"Bytow"	0.45
10	"Lebork"	0.391304347826087

After analyzes we decided to build our warehouse in Gdynia.