

Banco de Dados

Igor Cortes Junqueira
Igor Patrício Michels

2020.2

1 Introdução

O presente documento tem por objetivo relatar o desenvolvimento da elaboração da visualização em grafo de um banco de dados para a segunda avaliação da disciplina de Banco de Dados da FGV-EMAp, ministrada pelo professor Renato Rocha Souza. O objetivo é realizar a visualização em grafo do banco de dados desenvolvido na primeira avaliação com informações sobre os voos domésticos dos EUA entre janeiro de 2019 e maio de 2020, cujos scripts podem ser encontrados em [3] e o banco pode ser encontrado **nesse link**. Já o desenvolvimento do presente trabalho pode ser encontrado em [2].

2 Elaboração

2.1 Fontes

Como estamos utilizando o banco desenvolvido no trabalho anterior, as fontes dos dados dos voos utilizados se mantém, ou seja, os dados dos voos foram obtidos por meio dos registros oficiais da BTS - Bureau of Transportation Statistics [5], já os dados das aeronaves foram obtidos pelos registros da FAA - Federal Aviation Administration [1].

2.2 Modelo Lógico

Para realizar a manipulação dos dados é interessante relembrar a estrutura lógica do banco para que facilite no momento em que formos realizar query's. Dessa forma, conforme relatório passado, o modelo lógico final é o que está ilustrado na Figura 1.

2.3 Desenvolvimento

Nosso primeiro passo foi elaborar um grafo básico com as companhias aéreas apontando para as rota que possuem, por exemplo, a Hawaiian Airlines INC possui, entre suas rotas, as rotas de Honolulu a Kuhului (rota HNL-OGG) e de Kuhului a Honolulu (rota OGG-HNL). Visto isso, teremos um nó que representa a companhia havaiana com uma aresta apontando para um nó que representa a rota HNL-OGG e também uma outra aresta que aponta para um nó representando a rota OGG-HNL¹. No fim, obtivemos o grafo que se encontra na Figura 2.

Conforme podemos notar na Figura 2, não conseguimos extrair informação alguma de um grafo com todas informações disponíveis ao mesmo tempo pois o mesmo acaba ficando poluído. Dessa forma elaboramos algumas restrições quanto aos dados analisados, limitando a quantidade de aeroportos e companhias. Além disso, mudamos a análise dos dados, ao invés de cruzarmos companhias com rotas, decidimos relacionar as companhias com os aeroportos em que elas operam e os aeroportos entre si caso eles possuem uma rota entre si. Dessa forma, obtemos o grafo que pode ser observado na Figura 3. O layout do grafo foi inspirado em um dos grafos vistos durante o curso e que pode ser encontrado em [4].

¹Existem as duas rotas, logo são duas arestas.

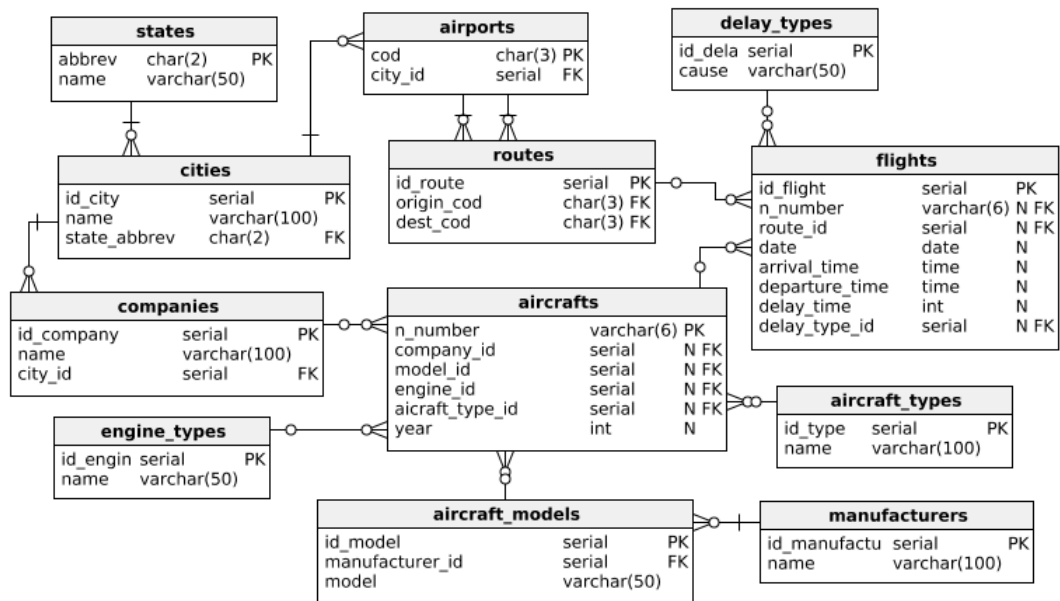


Figura 1: Modelo Lógico

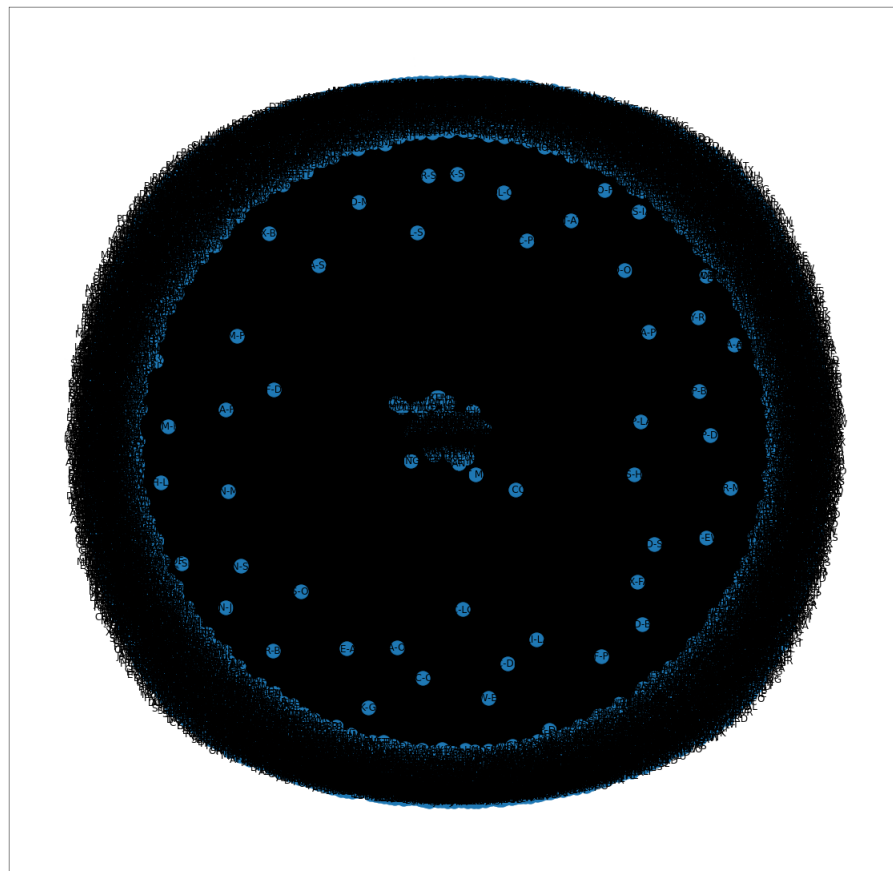


Figura 2: Grafo com todas as companhias e rotas

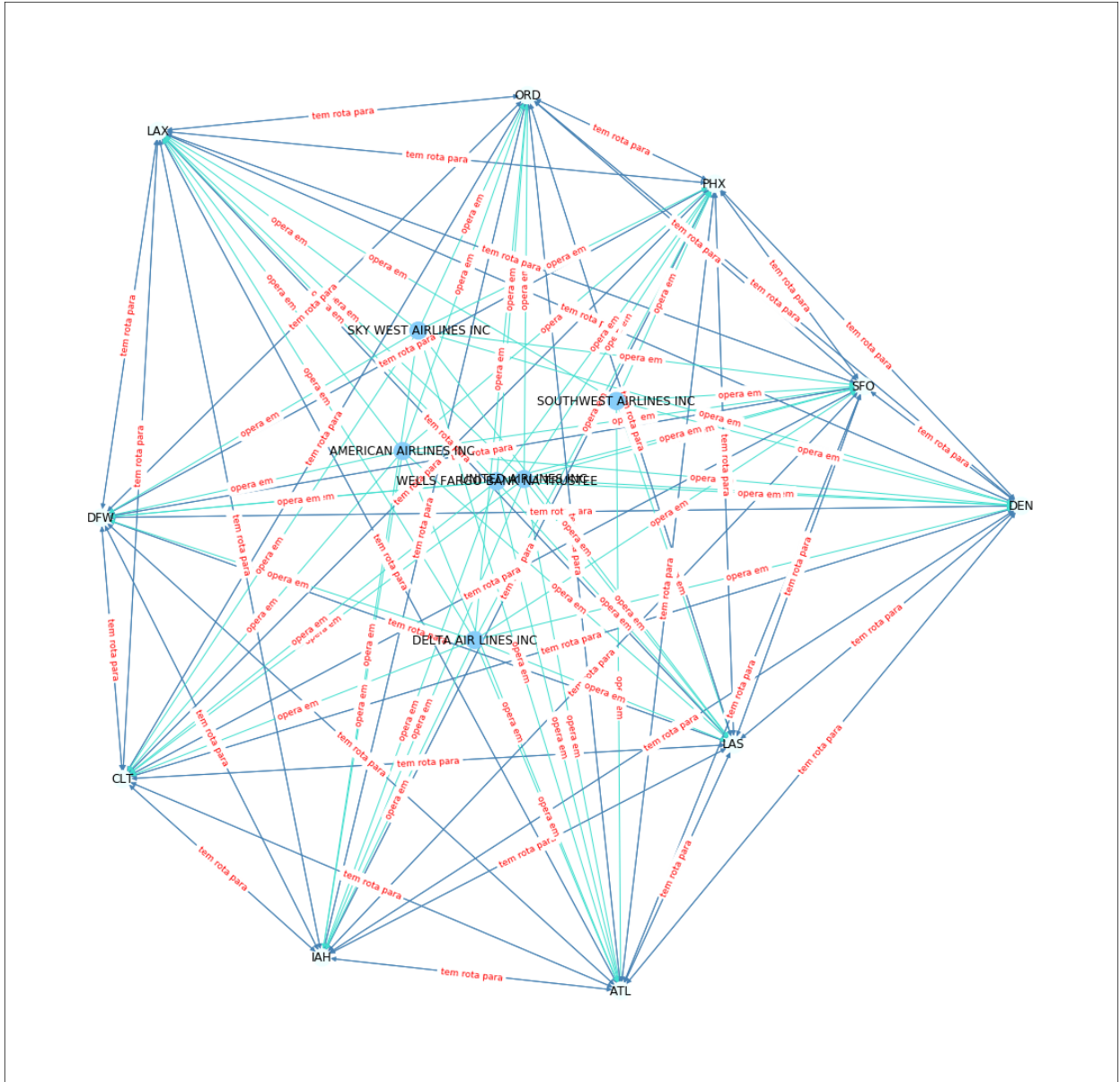


Figura 3: Grafo mais limpo das companhias e rotas

Feitas tais modificações, a eficiência do código estava abaixo do satisfatório², dessa forma incluímos uma nova restrição, mas dessa vez no banco, onde passamos a considerar apenas os dados de maio de 2020³. Além disso, cruzamos os dados de fabricantes com as companhias também, fazendo uma visualização de que fabricantes de aeronave fabricam para cada companhia. No final criamos um grafo interativo onde, onde o usuário pode escolher quantas companhias, aeroportos e fabricantes deseja plotar, limitados a 12, 30 e 5, respectivamente. Os grafos finais poderão ser vistos na próxima seção.

Por fim, vale a pena relatar um problema que ocorreu durante a execução de uma query no Ubuntu. Uma das configurações padrão do MySQL Server estava gerando um erro no `GROUP BY`. Para resolver esse problema foi adicionada uma linha com o texto `sql_mode = ""` no final do arquivo localizado em `/etc/mysql/mysql.conf.d/mysqld.cnf`.

3 Resultados

Nos grafos plotados, temos em “azure” os aeroportos, em “lightskyblue” as companhias e em “turquoise” as empresas fabricantes de aeronaves. Conforme a seção anterior, a interação entre aeroportos se dá através de uma aresta que significa que existe uma rota entre dois aeroportos. Já a interação entre companhias e aeroportos se dá por uma arestas que representa que a companhia opera naquele aeroporto e a interação entre companhias e fabricantes se dá por uma aresta que representa que um fabricante fabricou aeronaves para aquela companhia.

Nas Figuras 4, 5, 6, 7 e 8 temos algumas das principais visualizações geradas pelo código elaborado. Temos combinações que trazem as três informações como um todo. Também, há grafos com quantidade de aeroportos ou de fabricantes iguais a 0, o que possibilita fazer análises de quem fabrica aeronaves para cada companhia ou que companhia opera em que aeroporto.

Mais visualizações podem ser obtidas diretamente no **painel interativo em Jupyter Notebook** gerado, o qual está disponível no GitHub [2] e permite personalização dos três parâmetros utilizados.

²Algumas vezes demorava cerca de 30 segundos para gerar um grafo!

³A opção por criar um novo banco e não fazer a restrição via query se deu pelo fato do banco completo ter na ordem de 10 milhões de registros de voos, assim a eficiência do código poderia ser comprometida.

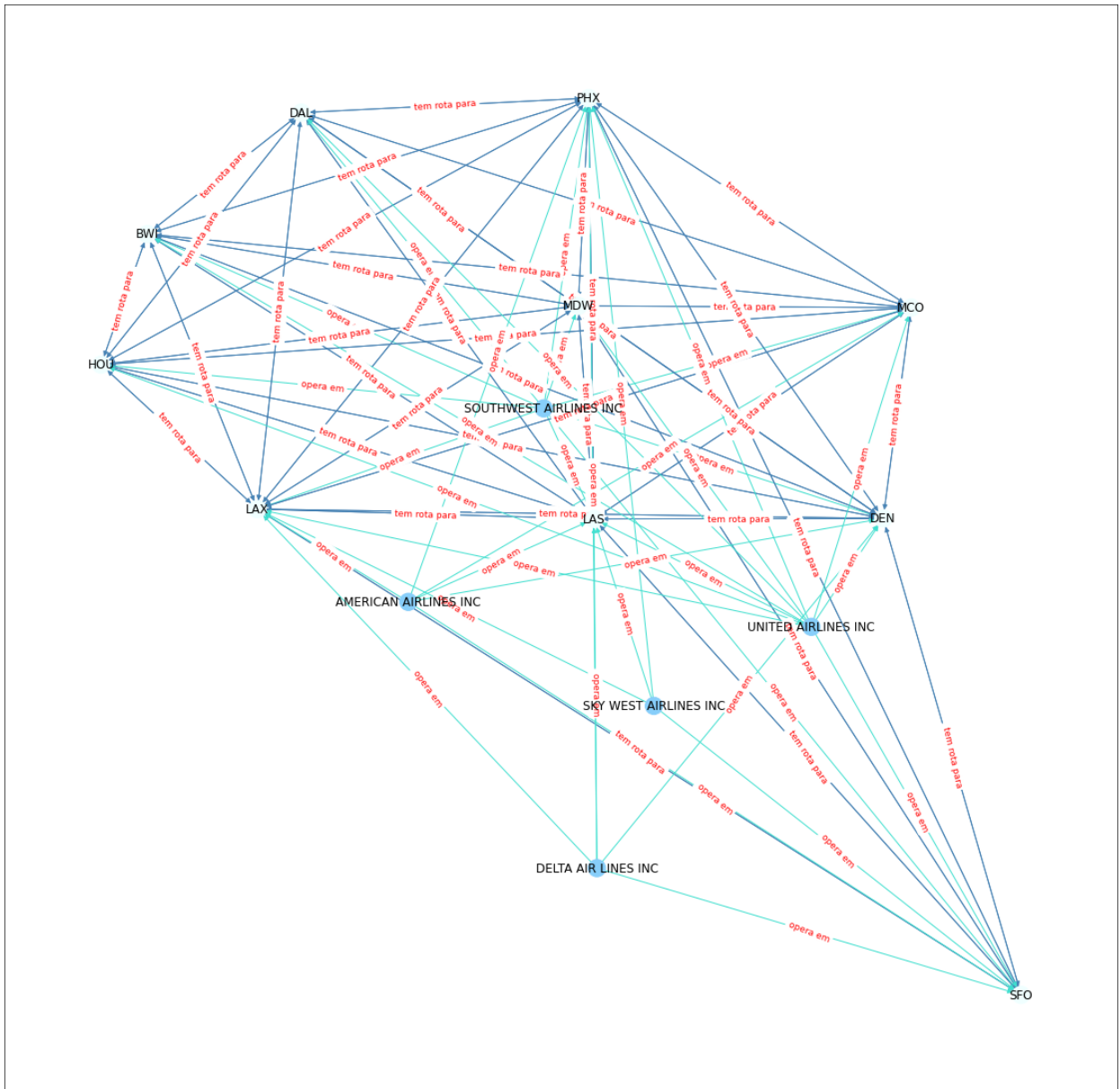


Figura 4: 5 maiores companhias e os 10 maiores aeroportos, sem fabricantes.

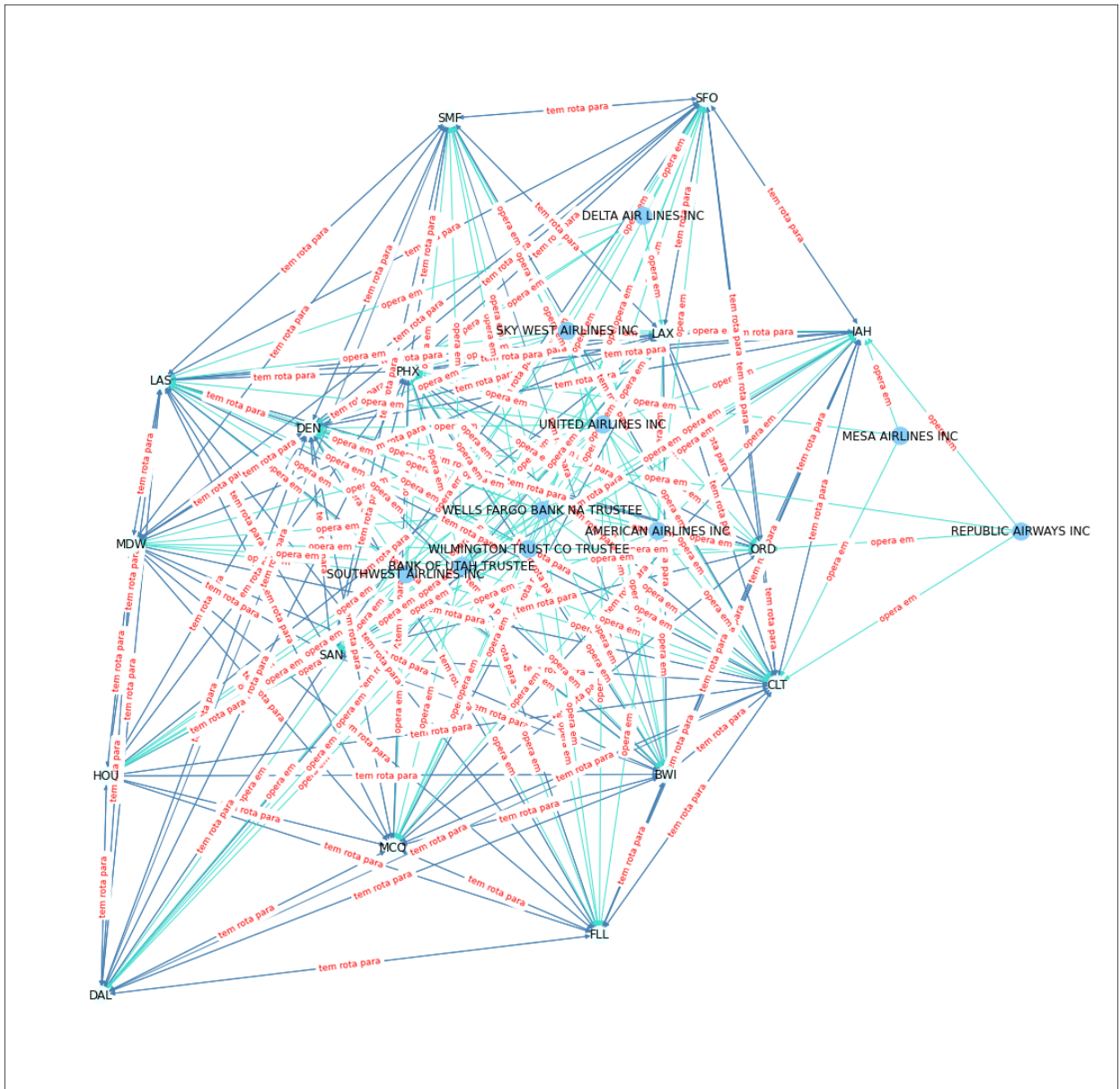


Figura 5: 10 maiores companhias e os 16 maiores aeroportos, sem fabricantes.

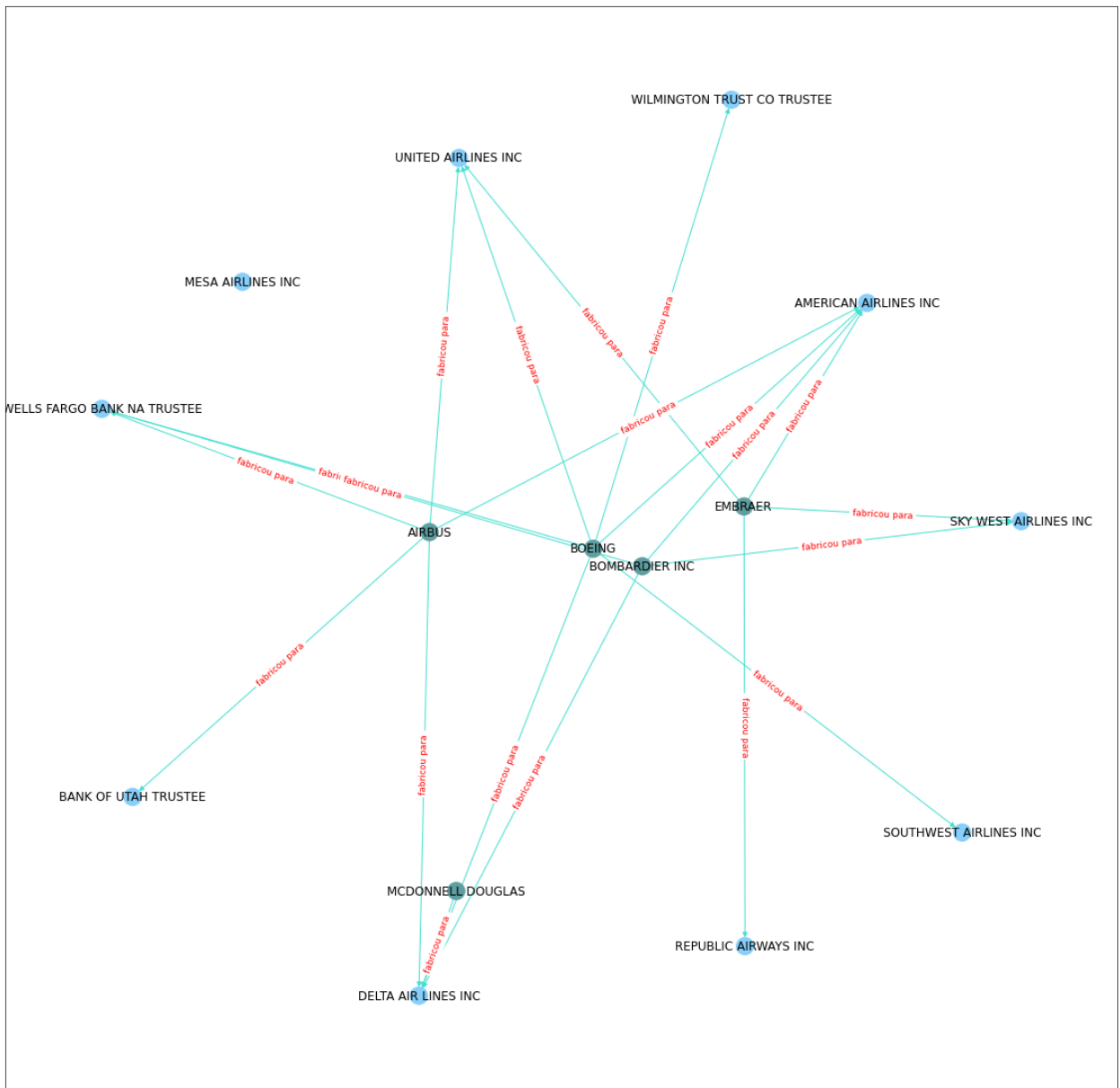


Figura 6: 10 maiores companhias e os 5 maiores fabricantes, sem aeroportos.

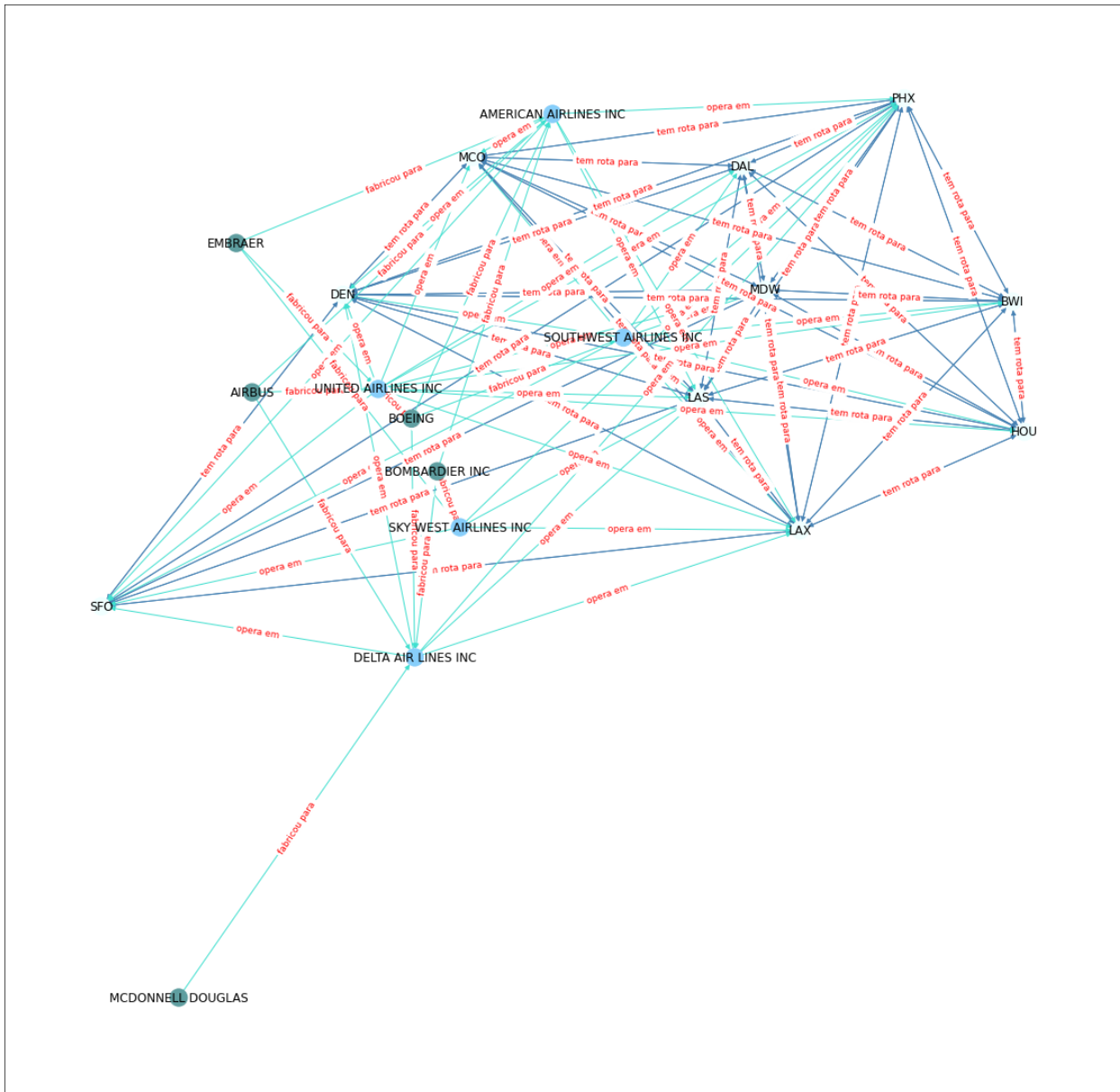


Figura 8: 5 maiores companhias, os 10 maiores aeroportos e os 5 maiores fabricantes.

Referências

- [1] Federal Aviation Administration. *FAA Registry*. URL: https://registry.faa.gov/aircraftinquiry/nnun_inquiry.aspx. (acessado: DD.MM.AAAA).
- [2] Igor Patrício Michels e Igor Cortes Junqueira. *Domestics Flights USA - Graph*. URL: https://github.com/IgorMichels/USA_Flights_Graphs.
- [3] Igor Patrício Michels et al. *Domestics Flights USA*. URL: https://github.com/IgorMichels/USA_Flights.
- [4] Renato Rocha Souza. *Databases Course*. URL: https://github.com/rsouza/Databases_Course/.
- [5] Bureau of Transportation Statistics. *Reporting Carrier On-Time Performance (1987-present)*. URL: https://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=236. (acessado: 13.09.2020).