

MAE 399 - Análise de Dados e Simulação
Prof. Fábio Machado
Lista 5 - entrega - 06/06/2022

1. Recentemente introduzimos o tema *Cadeias de Markov* dentro do tópico *Processos Estocásticos*. O Objetivo desta lista é entender a dinâmica envolvendo o deslocamento das bicicletas pelas várias estações do sistema Citi Bike, como uma *Cadeia de Markov*.

O problema inicial é o fato de termos mais de 1000 estações. Se usarmos todas individualmente teríamos uma matriz 1000×1000 com muitos zeros pois há uma série de pares de estações para as quais nenhuma bicicleta fez o percurso de uma estação para a outra. A partir desta constatação a proposta é definir clusters de estações, seguindo algum critério, para depois entender os deslocamentos de bicicletas entre as estações de cada cluster e entre os clusters.

Para cada mês pares do ano de 2018:

- a) Encontre a estação de partida mais popular da base de dados, bem como os 9 destinos mais populares a partir desta estação. Vamos pensar este conjunto de estações como um primeiro cluster.
- b) Apresente a matriz de probabilidades de transição entre estas 10 estações, descartando todas as transições com origem ou destino que não seja uma destas 10 estações.
- c) Verifique se há estações sem nenhuma ligação direta (origem-destino) com alguma das 10 estações do cluster acima. Para este conjunto de estações, repita o que foi feito acima. Vamos pensar este conjunto de estações como um segundo cluster.
- d) Apresente um mapa com a localização geográfica destas estações, com uma cor para cada um dos clusters.