

Bruno Donato Banhos 587460
Carla Simões Gama 613843
Daniel Liberato de Jesus 552127
José Gabriel de Oliveira Santana 620459

c2ds1-2sp

De acordo com os resultados coletados pelos algoritmos feitos pelos integrantes e executado o algoritmo de índice Rand, concluiu-se que para o arquivo **c2ds1-2sp**, os cluster gerados pelo k-médias (com qualquer k definido) são diferentes, pois o índice é bem próximo de -1.

O single-link, por sua vez, nos mostra que os clusters produzidos ao serem comparados com os clusters reais, são muito parecidos, com o índice rand variando entre 1 (com k igual a dois) a 0.994 (com k igual a cinco).

Já o average-link, o resultado dos clusters reais e os produzidos pelo algoritmo, são idênticos para k igual a dois, ou seja, índice rand igual a um. Porém o índice rand vai diminuindo no decorrer do crescimento de k, como k igual a cinco, em que o índice rand é igual a 0.26. Ressaltando que quanto mais próximo de -1 o índice rand ajustado, mais diferentes são os clusters produzidos dos reais, com -1 significando totalmente diferentes e 1 completamente iguais.



Imagem 1: gráfico com os resultados do índice rand para cada valor de k da entrada de dados c2ds1-2sp

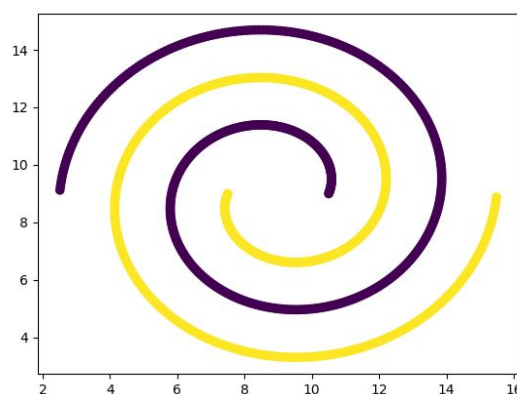


Imagem 2: plot do resultado do single-link para k = 2 (Cores Azul e Vermelho)

c2ds3-2g

Para a segunda entrada de dados, o algoritmo que o índice rand ajustado mostra que é mais parecido com o cluster real, é o k-médias com k igual a dois e índice rand igual a 0.92, com variação de 0.86 para k igual a 5. Já para o single-link, os resultados foram opostos ao k-médias, pois obteve índice rand ajustado bem próximo de -1. O average-link, por sua vez foi meio termo, com o índice rand ajustado variando próximo de 0.5.

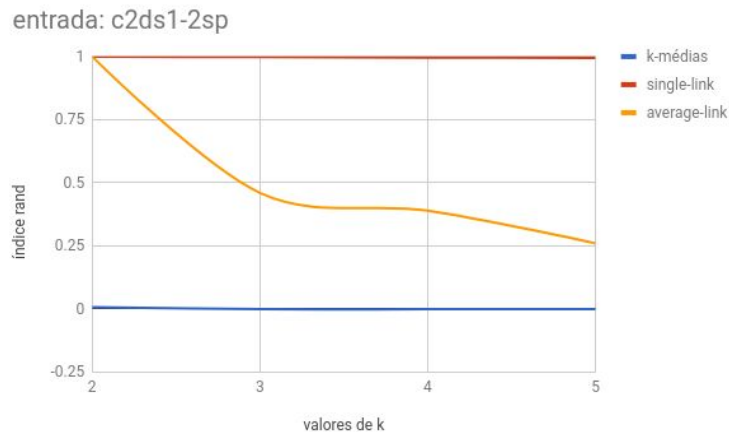


Imagem 3: gráfico com os resultados do índice rand para cada valor de k da entrada de dados c2ds3-2g

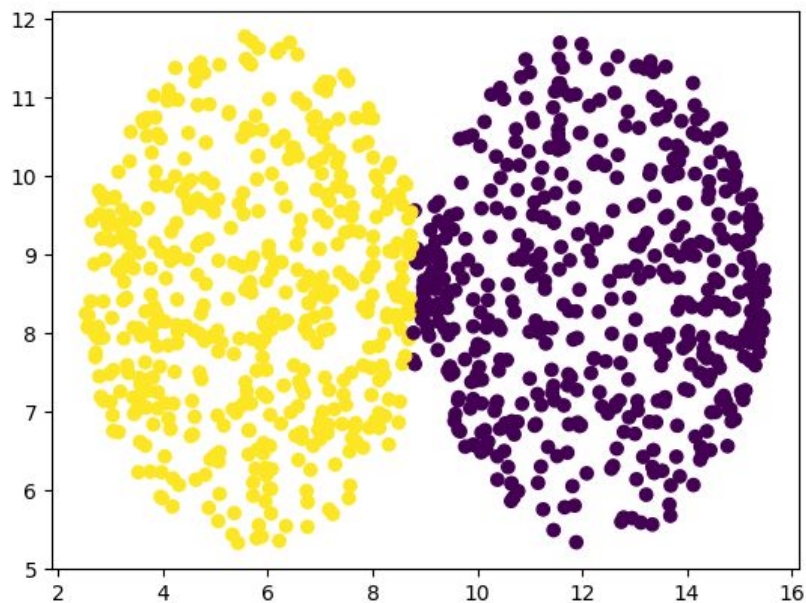


Imagem 4: Plot do resultado do algoritmo k-médias, onde $K = 2$

Monkey

Neste conjunto de dados, os resultados referentes ao algoritmo k-médias, mostraram-se um pouco mais próximos do conjunto de dados de teste, mas ainda muito diferentes do esperado.

O algoritmo que mais se aproximou do conjunto original foi o single-link, com ARI alcançando 0.87, enquanto o average-link foi o que obteve o menor ARI.

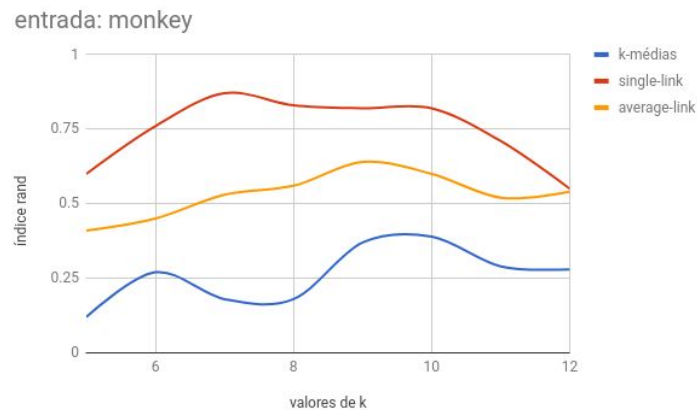


Imagem5: gráfico com os resultados do índice rand para cada valor de k da entrada de dados monkey

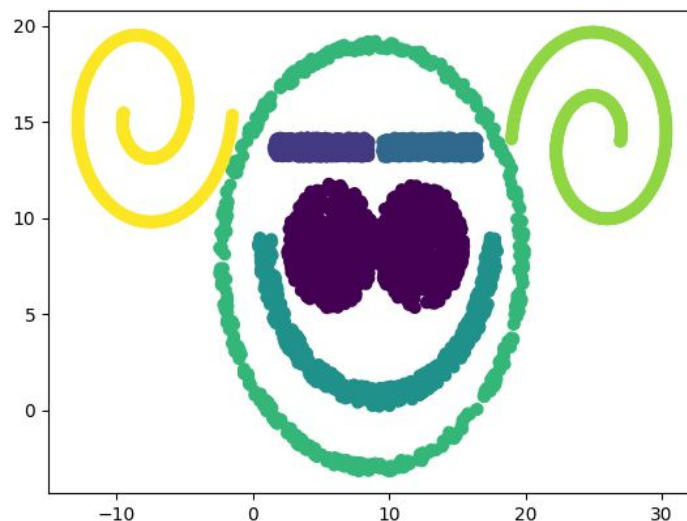


Imagem 6: Plot do resultado do algoritmo single-link, onde K = 7

Conclusão

A partir dos resultados obtidos podemos concluir que para cada entrada, um algoritmo diferente é mais adequado para agrupar. No caso da entrada do c2ds1-2sp, o single-link obteve os melhores resultados. Já para a entrada c2ds3-2g, o algoritmo que obteve os melhores resultados foi o k-médias. Por fim, para a entrada monkey, o melhor algoritmo foi o single-link mesmo. O grupo também concluiu que o algoritmo do average-link pode ser considerado caso o algoritmo definido como principal não tenha um resultado satisfatório, já que para todas as entradas, o índice rand ajustado apontou que o average-link ficava entre o melhor e o pior algoritmo.