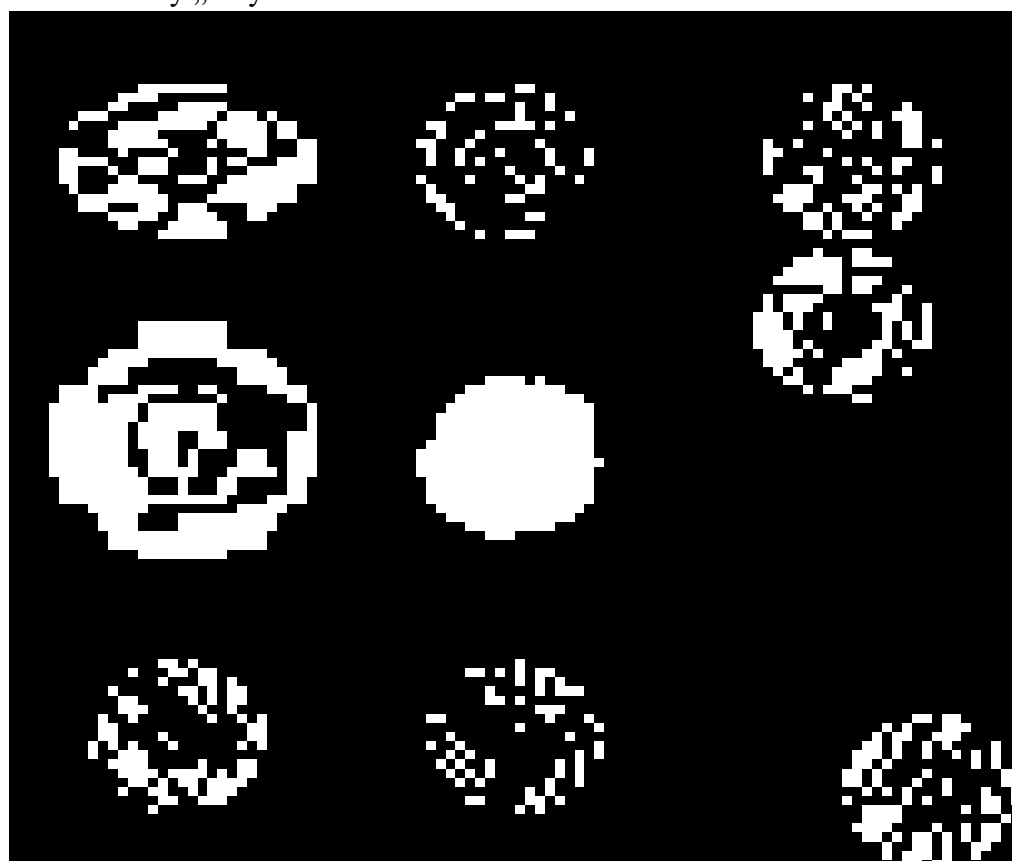


Marcin Sosniak 291547  
k-means

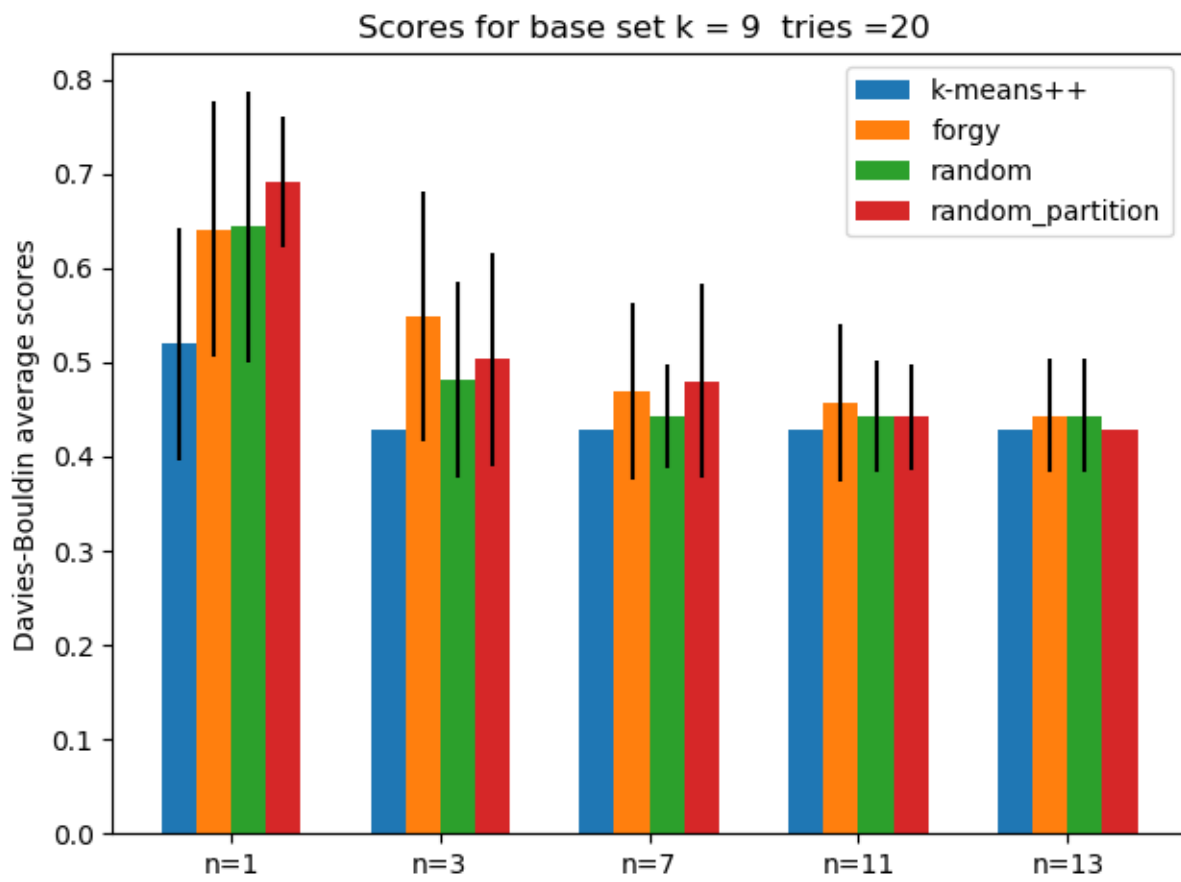
Zbior testowy „ładny”:



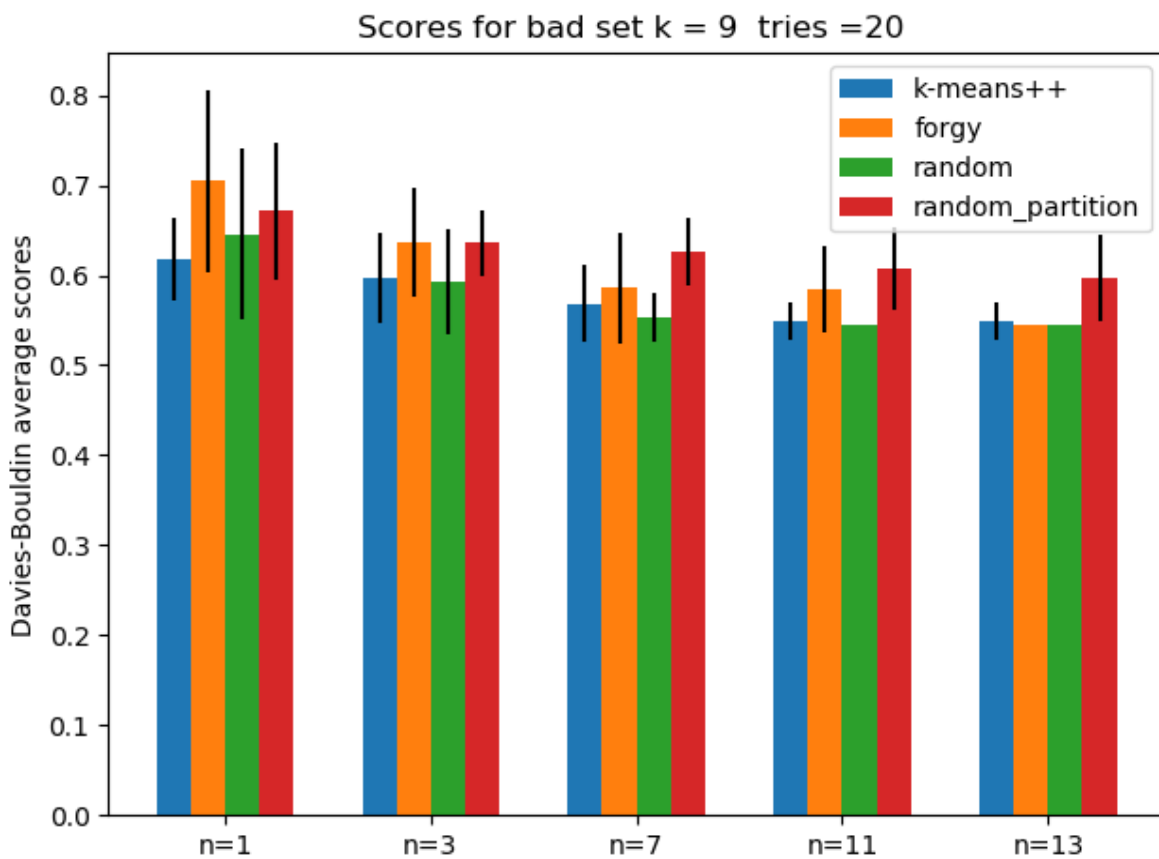
Zbior testowy „brzydki”:



Zmienne n dla „ładnego” zbioru :

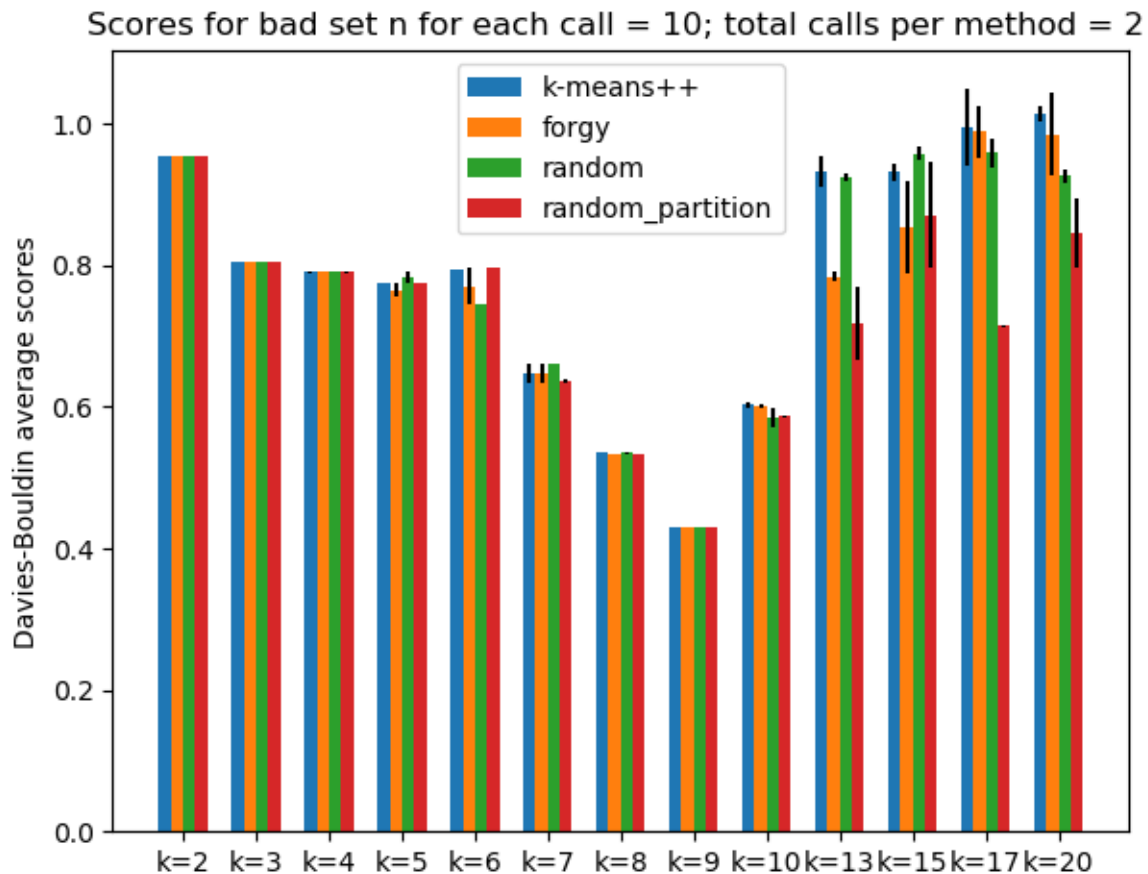


Oraz Zmienne n dla „brzydkiego” zbioru:

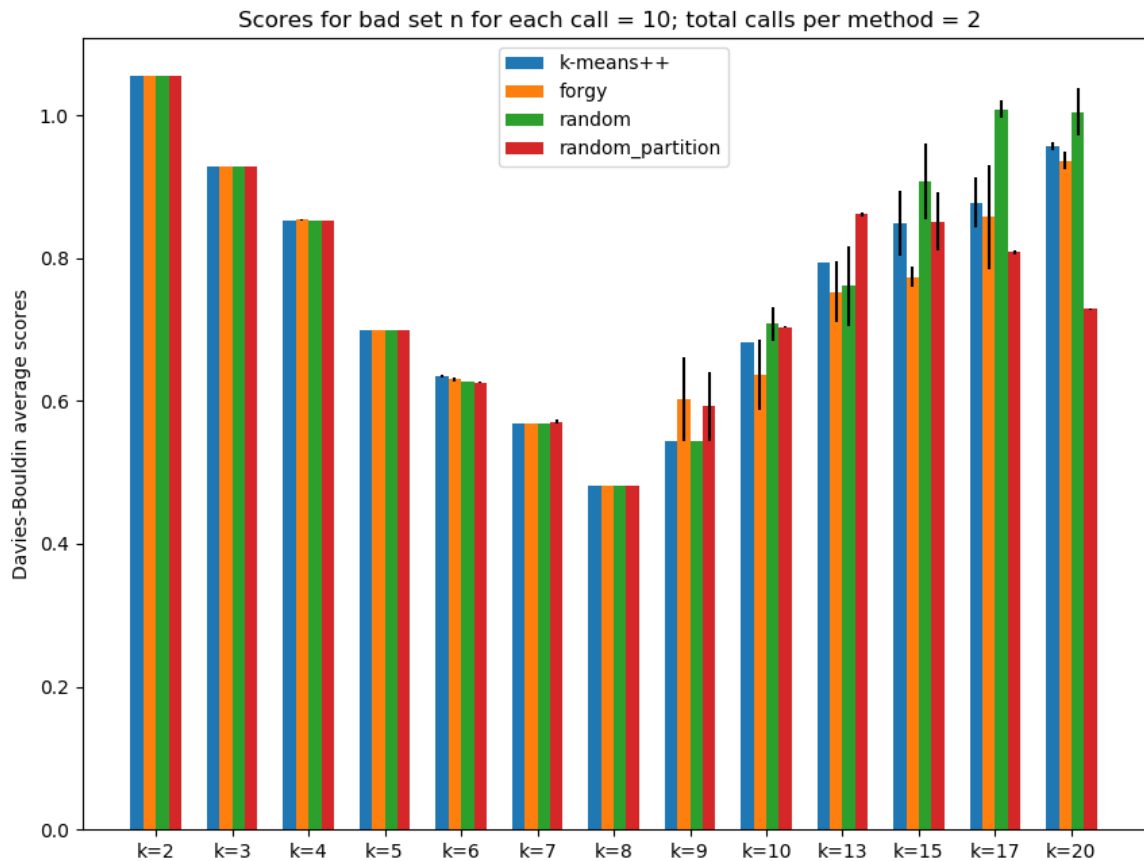


Oczywiście im niższy indeks tym lepszy wynik. Z porównania wykresów widać, że na „ładniejszym zbiorze” wszystkie metody inicjalizacji poradziły sobie lepiej niż na „brzydkim”. Możemy też zauważyć że k-means++ radzi sobie widocznie lepiej na prostym zbiorze niż na trudniejszym. Skuteczność uniform random, zwłaszcza na trudniejszym zbiorze najprawdopodobniej wynika z równomiernego ułożenia klastrów w zbiorze testowym, więc ma sporo szansę dobrze 'pasować' do zbioru. Ale nawet z tym założeniem k-means jest najczęściej najlepsza z inicjalizacji

Sprawdźmy teraz różne k:  
„ładny” zbior :



oraz „brzydki” zbior:



zauważmy że liczba prób była niska. Dla ładnego zbioru zdecydowanie widac że k-means'y 'uwazaja' że jest tam 9 klastrow. Co zreszta prawda jest tam 9 idealnych klastrow. W wersji skolei zmodyfikowanej, wychodzi iz dla 8 setow jest najlepszy wynik, co jest najprawdopodobniej zwiazane z tym że bardzo bliskie siebie 2 klastry zostały uznane za jeden klaster (co zreszta jest nie do stwierdzenia bez posiadania innych informacji o danych niż same punkty). Możliwym wynikiem było by też 10, ale najwyraźniej wrzeczono było zbyt okrągłe