João José Marques Lopes - joaolopes@student.dei.uc.pt - 2016244868
Tiago Daniel Fonseca Pessoa - tpessoa@student.dei.uc.pt - 2016242888
Roman Zhydyk - roman@student.dei.uc.pt - 2016231789

Regressão Linear - Meta 3

Introdução

Olhando para os resultados obtidos das duas metas anteriores, verificou-se que apenas a regressão linear, logarítmica e exponencial faziam lógica calcular, pois os gráficos descrevem funções com o mesmo aspecto. De modo a obter a equação da função que me melhor se adaptava e o coeficiente de determinação (R^2), utilizou-se a função curvefit em Python.

Análise da Maior subsequência

Foram então calculadas a regressão linear e logarítmica fixando a probabilidade de erro do algoritmo, de 0.01 até 0.91 e variando o N de 1000 até 10000. Também foram realizadas experiências fixando o N e variando a probabilidade, nas mesmas escalas, poŕem para este foi usada a regressão exponencial pois era a que melhor se adequava à distribuição dos pontos no gráfico.

- Regressão linear

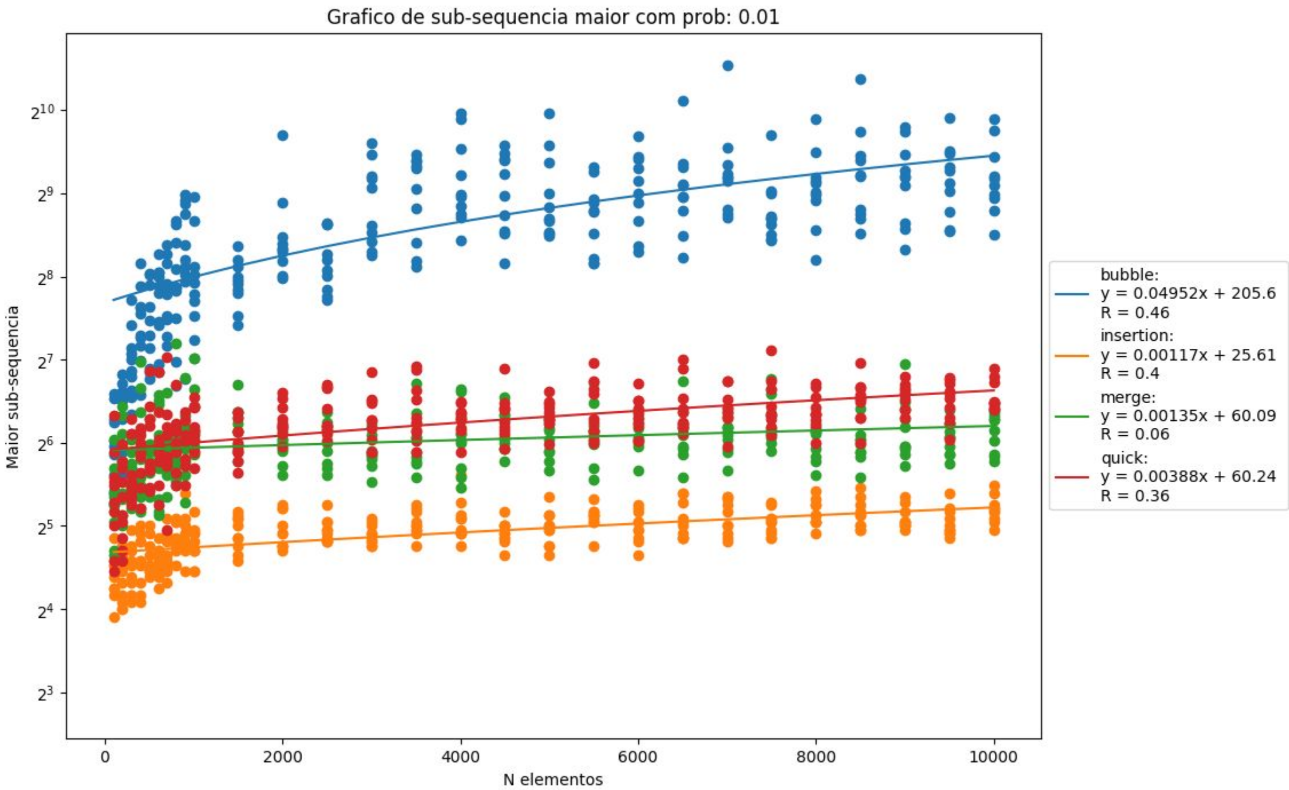

Figura 1.

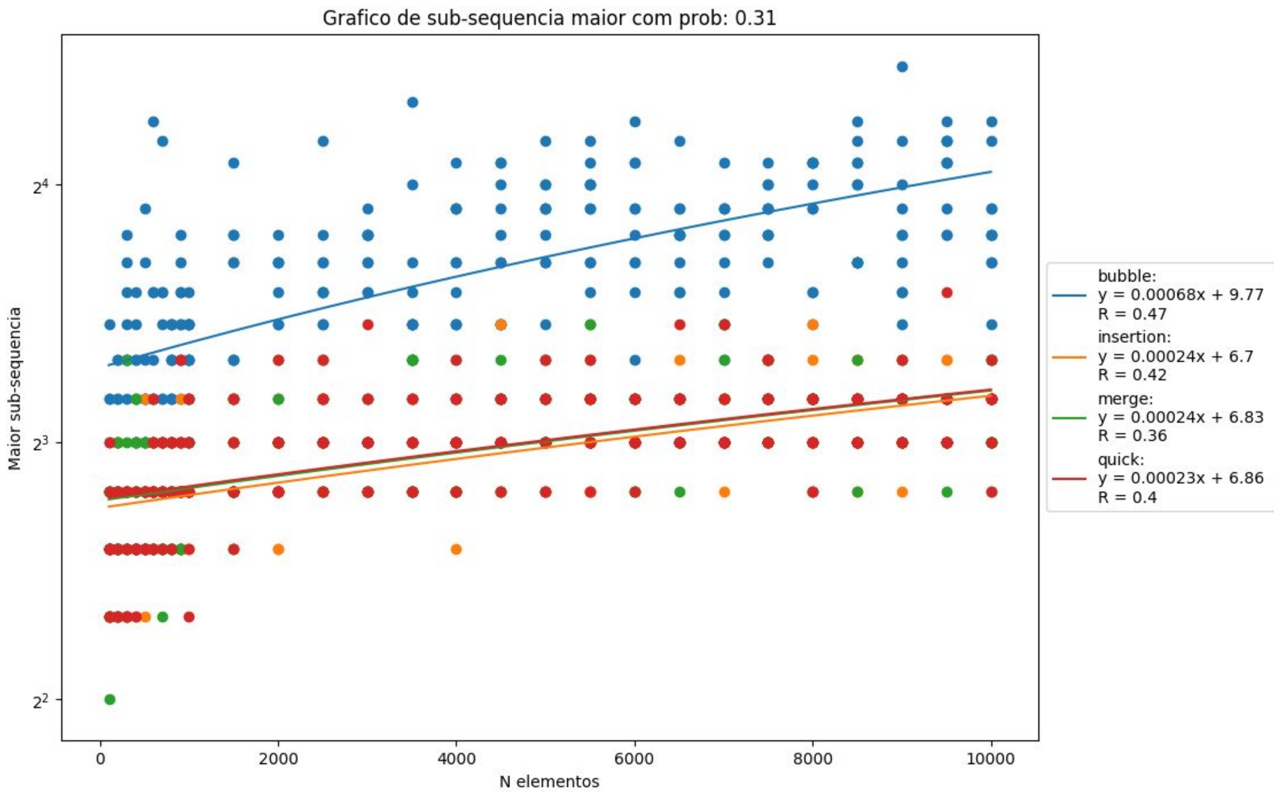

Figura 2.

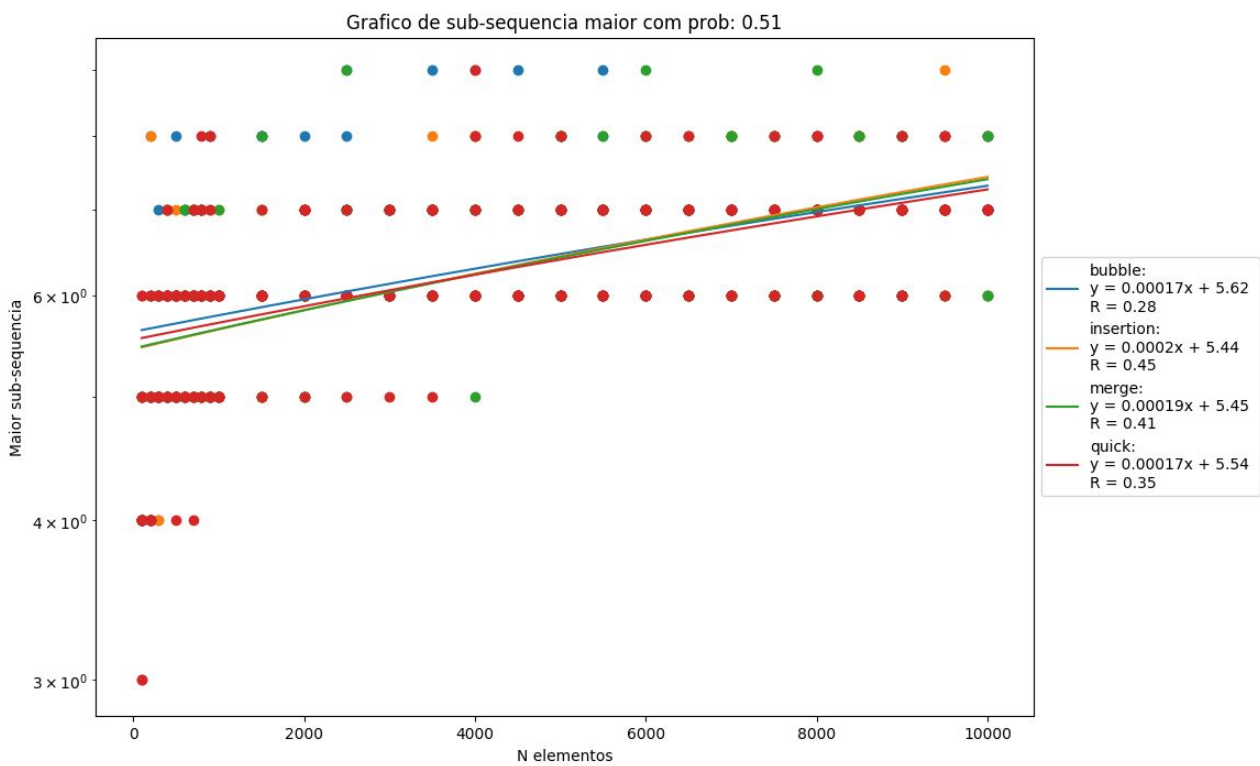

Figura 3.

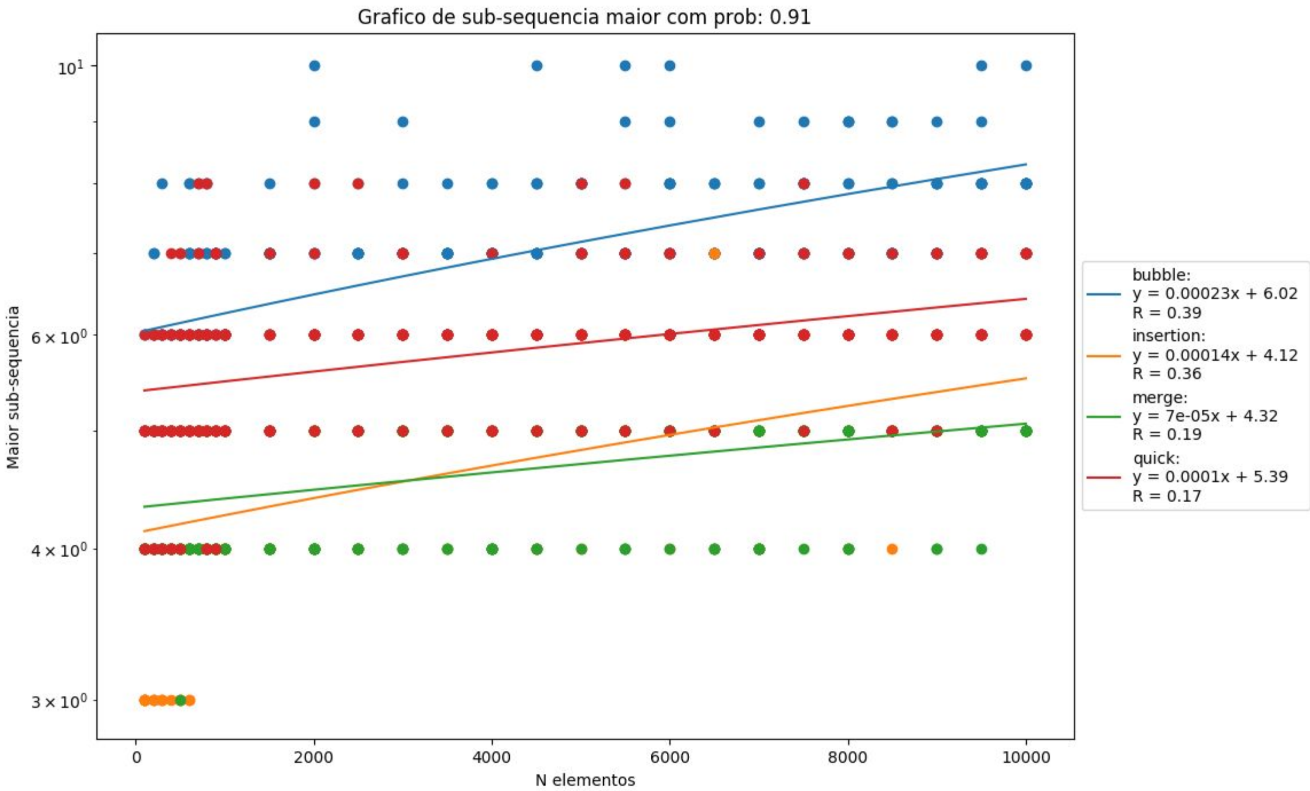

Figura 4.

- Regressão logarítmica

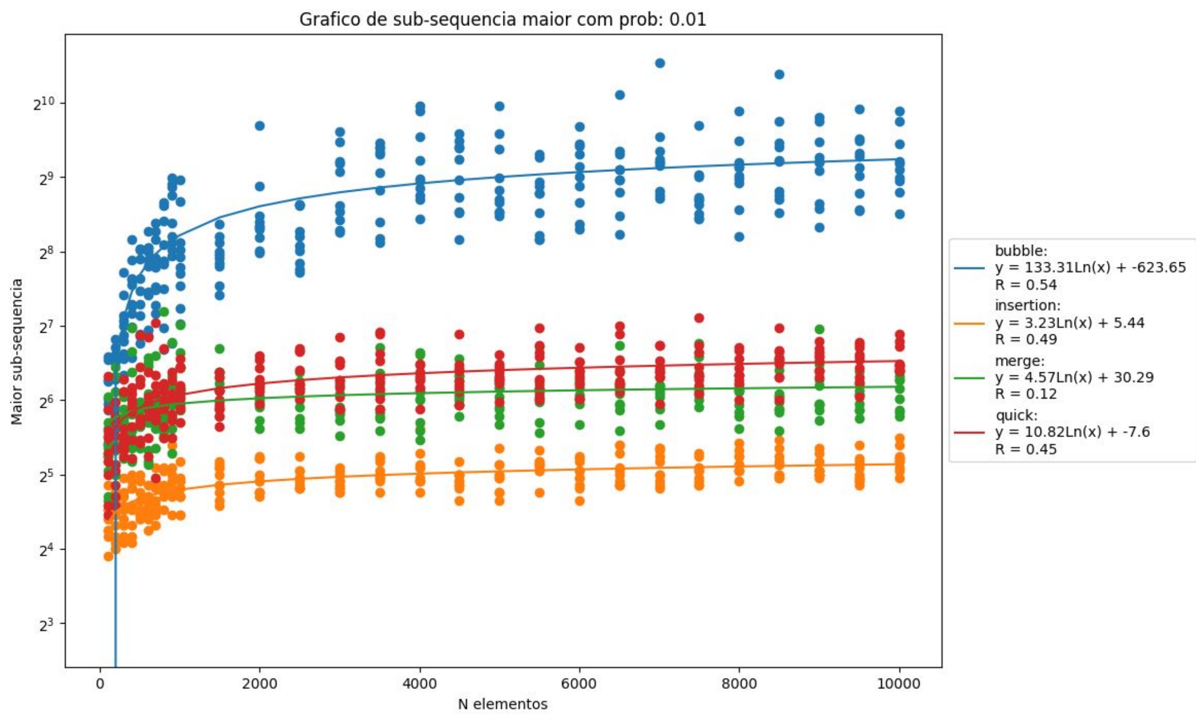

Figura 5.

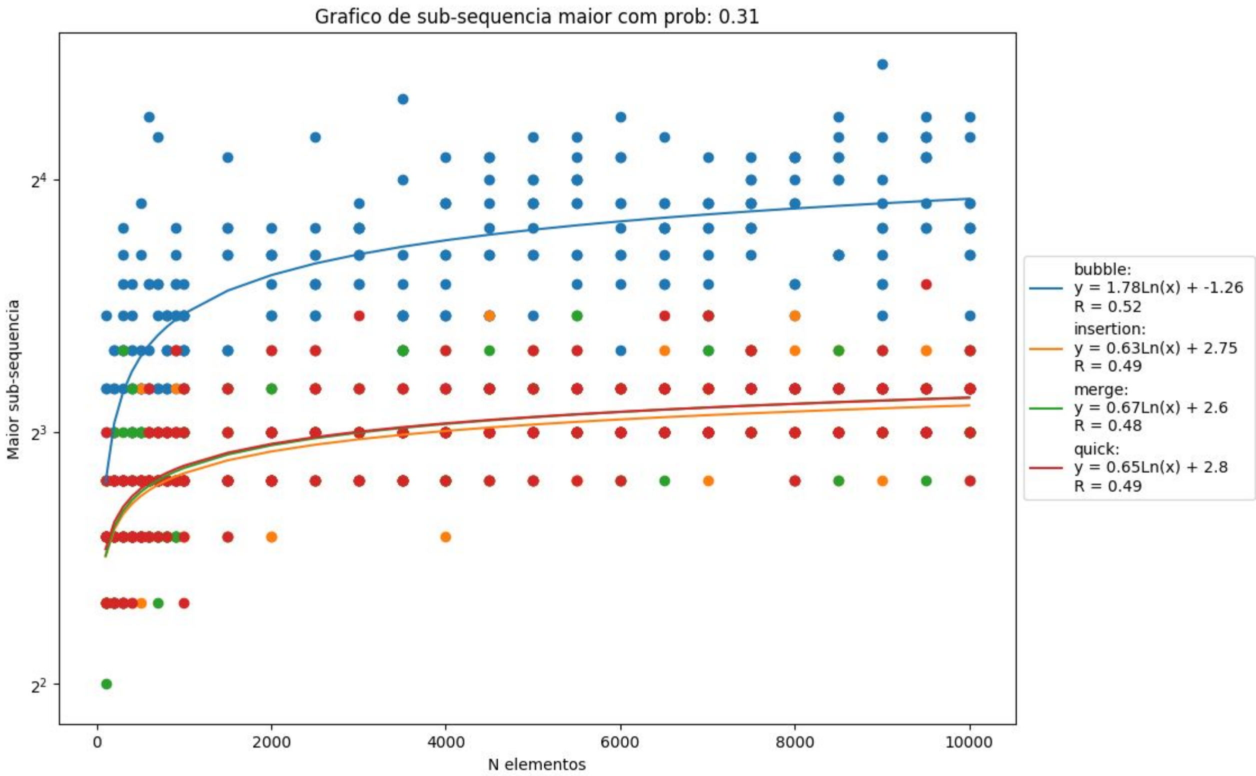

Figura 6.

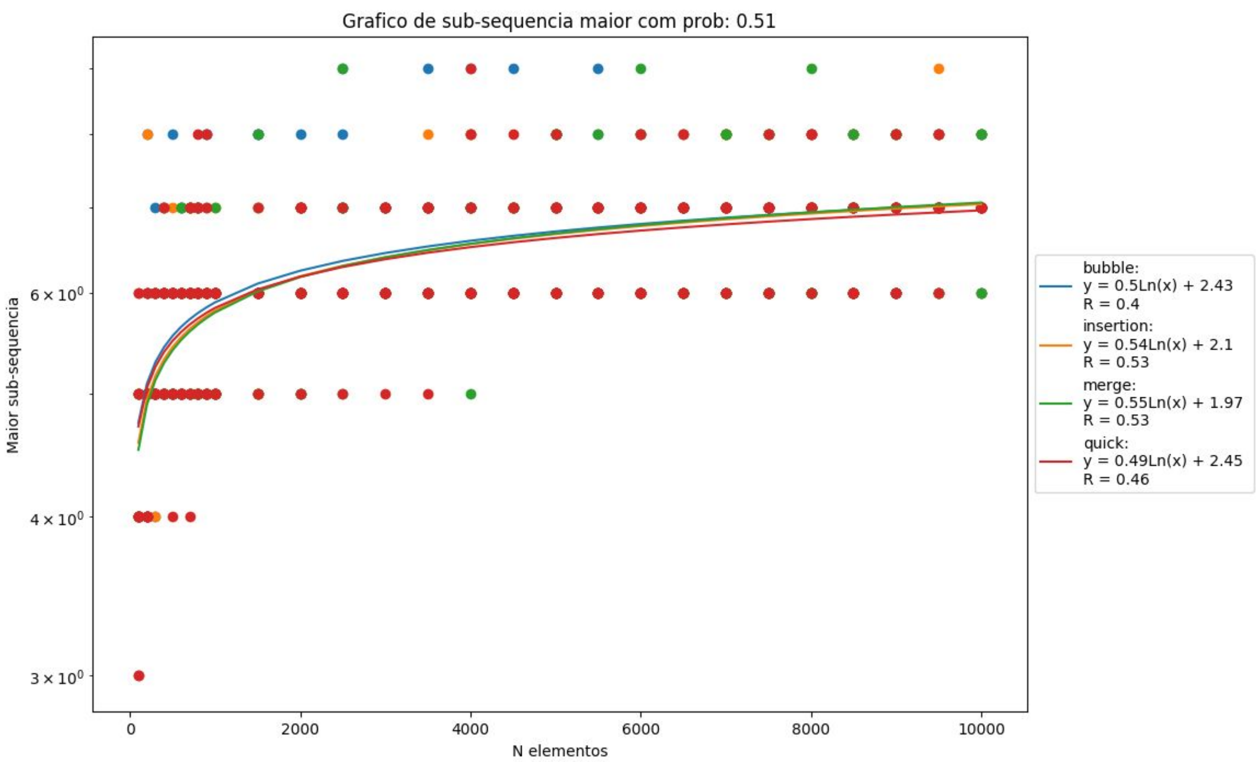

Figura 7.

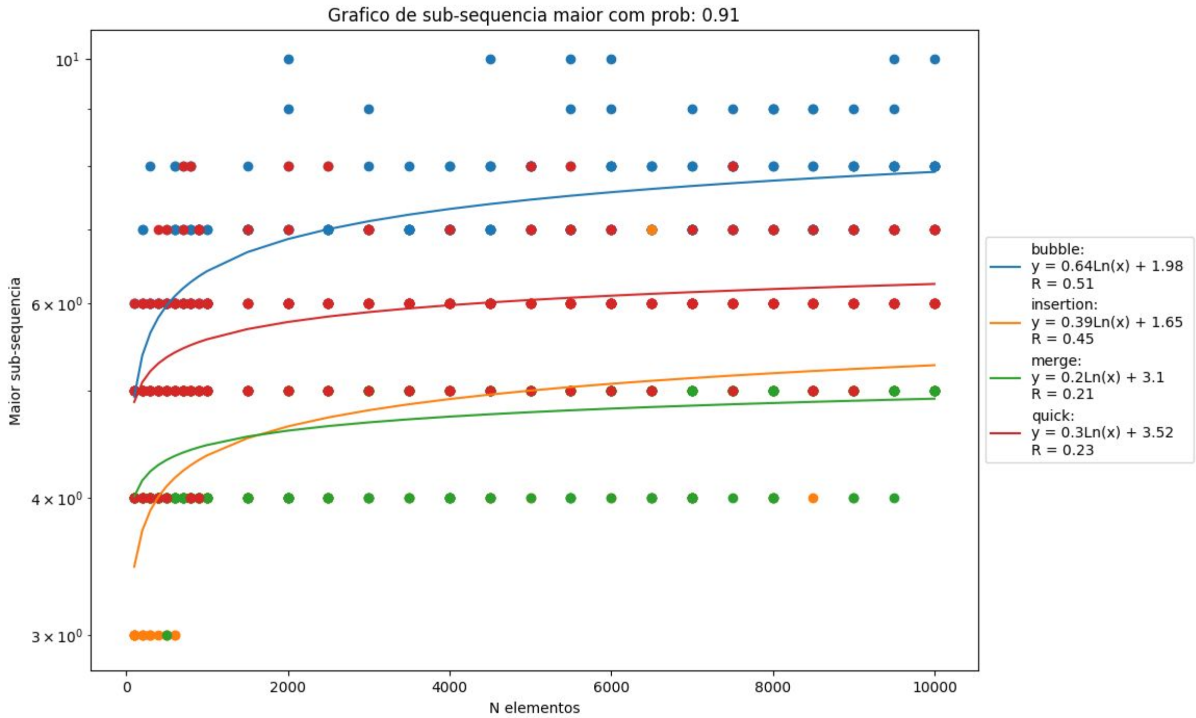

Figura 8.

- Regressão exponencial

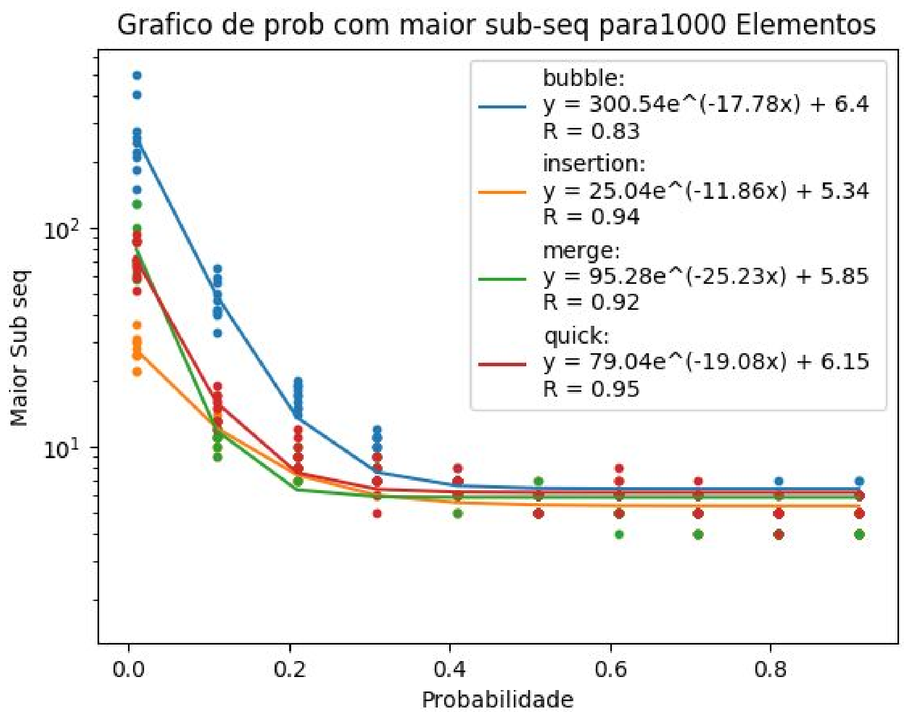

Figura 9.

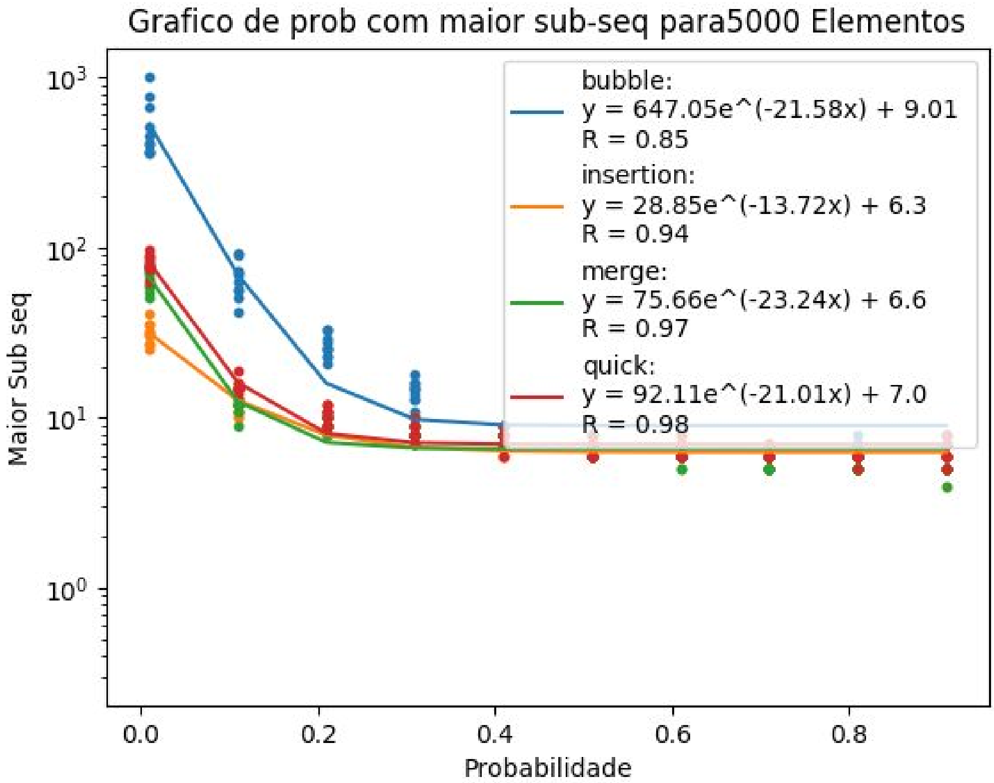

Figura 10.

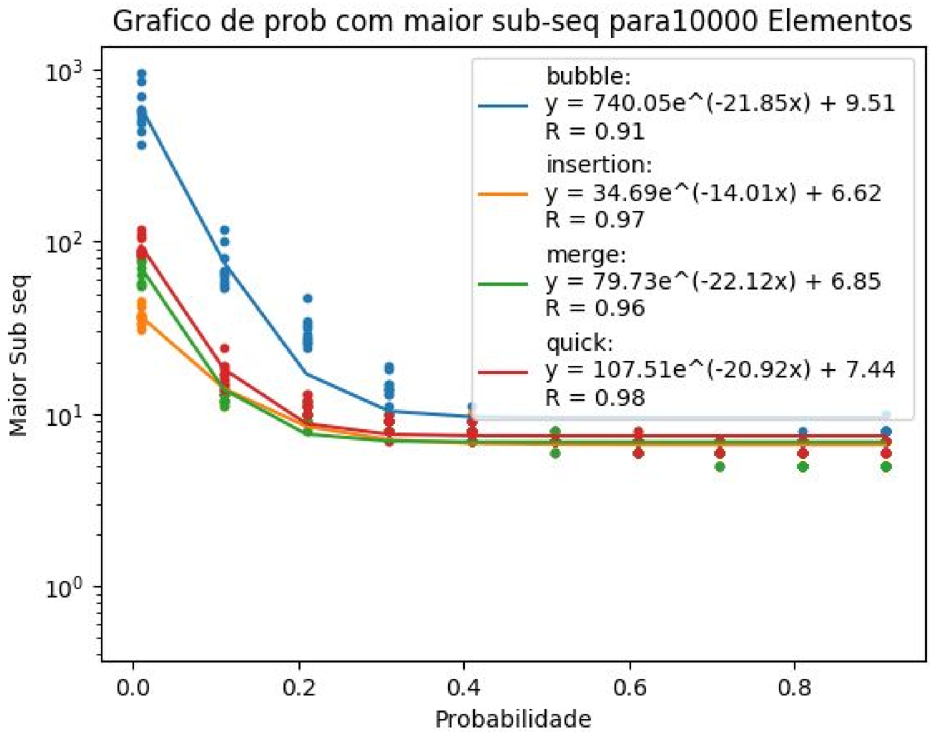

Figura 11.

Análise dos resultados e conclusão

Começou-se por fazer a regressão linear (Figura 1. até Figura 4.) para os casos em que se variou o tamanho e manteve-se fixo o valor da probabilidade. Como se pode observar, os valores de R^2 deram muito abaixo dos valores ideais. Na regressão logarítmica (Figura 5. à Figura 8.), ao analisarmos os gráficos, notou-se que os valores de R eram igualmente baixos. Assim, vimos que não fazia sentido fazer mais nenhum tipo de regressão porque não havia nenhuma função que se fosse adaptar àquela distribuição de pontos. Com isto podemos concluir que a variação de tamanho não tem grande influência no resultado final da análise da maior subsequência.

Para a segunda experiência efectuada, fixar o tamanho e variar a probabilidade (Figura 9. Figura 11.), ao observar os gráficos, pode-se notar que a função que melhor se adaptaria aos dados seria uma função exponencial. Foram então realizados os testes para tamanho do array de 1000 a 10000 e pode-se observar pela análise da função que o R neste caso dava um valor muito próximo de 1, o que é um valor ideal.

Com isto podemos confirmar os dados analisados na meta 2, em que se viu que a variável independente que teria mais influência no resultado era a probabilidade e não o tamanho do array.