

Teste de Hipóteses

André Martins 2006130025, Cláudia Campos 2018285941, Dário Félix 2018275530

1 Introdução

Com base na análise dos resultados da fase de Análise Exploratória de Dados e Regressão Linear, formulámos várias hipóteses, que apresentamos na Secção 1.1. Neste relatório, apresentaremos os resultados da nossa análise dessas hipóteses e se os dados as suportam.

1.1 Hipóteses

1. O algoritmo **Dinic** é melhor que o **EK** e, por sua vez, o algoritmo **EK** é melhor que o **MPM**, para valores de $|V|$ e $|E|$ baixos ($v = 100, 150, \dots, 250$ e $p = 0.1$)?
2. O algoritmo **Dinic** é melhor que o **MPM** e o **MPM** é melhor que o **EK**, para os restantes valores de $|V|$ e $|E|$ ($v = 300, 350, \dots, 1000$ e $p = 0.1$)?
3. O algoritmo **Dinic** é melhor que o **MPM** e o **MPM** é melhor que o **EK**, para os restantes valores de $|V|$ e $|E|$ ($v = 100, 150, \dots, 1000$ e $p = 0.2, 0.3, \dots, 0.7$)?
4. Os termos não presentes na expressão da complexidade de cada algoritmo têm impacto?

2 Métodos

Para minimizar o impacto da variação do *input*, realizámos experiências com *inputs* emparelhados, isto é, para a mesma medição, todos os algoritmos receberam o mesmo *input*, dado pela *seed*.

Em todas as hipóteses considerámos um nível de significância $\alpha = 0.05$.

2.1 Equipamento

Todas as experiências e medições foram realizadas na mesma máquina, cuja descrição técnica se encontra na Tabela 1. A máquina escolhida é vantajosa, pois tem muito poucos processos ativos, incluído as do sistema operativo, minimizando a tendência não determinística do tempo de execução caso fosse executada num computador pessoal normal [1].

Sistema Operativo	Raspberry Pi OS (baseado no Debian), Linux version 5.15.61-v8+
Memória	4GB LPDDR4-3200 SDRAM
Processador	Broadcom BCM2711, Quad core Cortex-A72 (ARM v8) 64-bit SoC @ 1.5GHz
Placa Gráfica	VideoCore VI
Disco	SD Card 32GB 100MB/s

Tabela 1: Especificações da máquina [2]

2.2 Compilação dos Algoritmos

Todos os algoritmos foram compilados conforme o *Makefile* disponibilizado.

2.3 Parametrização

2.3.1 Sementes Aleatórias

A *seed* é gerada aleatoriamente através de um gerador de números pseudoaleatórios. De modo a garantir a reprodutibilidade das experiências, a *seed* desse gerador foi definida com o valor zero.

2.3.2 Número de Medições (Repetições)

Segundo o Teorema do Limite Central, são necessárias pelo menos 30 amostras para poder efetuar uma análise segundo a distribuição normal (*Z-Test*). Por isso, para cada combinação de fatores (exceto a *seed*), realizámos **30 repetições**.

2.3.3 Intervalos dos Fatores

À semelhança da Análise Exploratória de dados, definimos o conjunto presente na Equação 1 para o número de vértices, v , e o conjunto presente na Equação 2 para a probabilidade, p .

$$V = \{50n + 50 \mid n \in \mathbb{N} \wedge n \geq 1 \wedge n \leq 19\} \quad (1)$$

$$P = \{0.1n \mid n \in \mathbb{N} \wedge n \geq 1 \wedge n \leq 7\} \quad (2)$$

2.3.4 Tempo Máximo de Execução

Decidimos escolher **10 segundos** como tempo máximo de execução por vários motivos: para o intervalo de fatores escolhido, especialmente em relação aos limites superiores, verificou-se em testes preliminares que seria o tempo suficiente para a execução dos algoritmos, e, além disso, para se poder obter os resultados em tempo útil, considerando o número de repetições e de medições definidos a partir da variação dos parâmetros.

3 Primeira Hipótese

Para analisar o impacto que o número de vértices e a probabilidade possuem na média do tempo de execução dos três algoritmos e se existem interações entre estes, identificamos três fatores, **vertices**, **probability** e **algorithm**, e definimos os níveis de cada um. Consideramos os níveis 100, 150, 200, 250 do fator **vértice**, o nível 0.1 do fator **probabilidade** e os níveis Dinic, EK e MPM do fator **algorithm**.

No entanto, como o fator **probability** apenas possui um nível, não o consideramos no teste estatístico que realizamos para esta hipótese. Por este motivo, e tendo em conta o *design* experimental da nossa experiência, optamos pelo teste estatístico *Two-Way Repeated Measures ANOVA*.

O *Repeated Measures ANOVA*, introduz os conceitos de *subjects*, *within-subjects* e *between-subjects*. No nosso caso, o *subject* da nossa experiência é a variável **seed**. O fator *within-subject* é o **algorithm** porque cada *input*, dado pela *seed*, é processado por cada nível do fator **algorithm**. E o fator *between-subject* é o fator **vertices**, dado que cada *input* apenas está sujeito a um nível deste fator.

3.1 Two-Way ANOVA Repeated Measures

Univariate Type III Repeated-Measures ANOVA Assuming Sphericity

	Sum Sq	num Df	Error SS	den Df	F value	Pr(>F)
(Intercept)	0.00090487	1	1.4360e-04	116	730.944	< 2.2e-16 ***
vertices	0.00049273	3	1.4360e-04	116	132.673	< 2.2e-16 ***
algorithm	0.00020356	2	9.2917e-05	232	254.134	< 2.2e-16 ***
vertices:algorithm	0.00012281	6	9.2917e-05	232	51.105	< 2.2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						

(a) Summary

Mauchly Tests for Sphericity

	Test statistic	p-value
algorithm	0.90158	0.0025871
vertices:algorithm	0.90158	0.0025871

(b) Teste de Mauchly

Figura 1: Resultados do *Two-way ANOVA Repeated Measures*

1. Não existe diferença na média do tempo de execução devido ao fator **vertices**
2. Não existe diferença na média do tempo de execução devido ao fator **algorithm**
3. Não existem interações entre os fatores **vertices** e **algorithm**

$$H_0^v : \mu_1. = \mu_2. = \dots = \mu_v. \quad (3)$$

$$H_1^v : \exists i, j : \mu_i. \neq \mu_j. \quad (4)$$

$$H_0^a : \mu_{.1} = \mu_{.2} = \dots = \mu_{.a} \quad (5)$$

$$H_1^a : \exists i, j : \mu_{.i} \neq \mu_{.j} \quad (6)$$

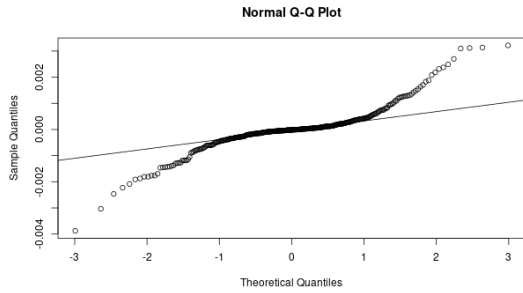
$$H_0^{v:a} : \forall i, j : \gamma_{ij} = 0 \quad (7)$$

$$H_1^{v:a} : \exists i, j : \gamma_{ij} \neq 0 \quad (8)$$

Observando os resultados da figura Figura 1a, obtivemos um *p-value* inferior a $2.2e^{-16}$ para os fatores **vertices**, **algorithm** e para a interação entre estes. Por isso, como o *p-value* é inferior ao nível de significância $\alpha = 0.05$, rejeitamos a hipótese nula para as três hipóteses do *Two-way ANOVA Repeated Measures* apresentadas anteriormente.

3.1.1 Pressupostos

No entanto, para podermos aceitar os resultados deste teste estatístico temos de analisar os pressupostos do *Two-Way ANOVA Repeated Measures*, a normalidade dos resíduos e a esfericidade da variância.



(a) Normal QQPlot

Anderson-Darling normality test

data: r\$.residuals
A = 16.868, p-value < 2.2e-16

(b) Anderson Darling Normality Test

Figura 2: Normalidade dos Resíduos

Na Figura 2a, observamos *outliers* nos extremos do Normal QQplot sugerindo que os resíduos não aparentam ser normalmente distribuídos. No entanto, para termos a certeza que os resíduos são ou não normalmente distribuídos, recorreremos ao teste de normalidade de *Anderson-Darling*. Como *p-value* deste teste, na Figura 2b, é inferior ao nível de significância $\alpha = 0.05$, podemos afirmar que o pressuposto da normalidade não se cumpre com um nível de confiança de 95%. Também avaliamos a esfericidade dos dados através do teste de *Mauchly*, na Figura 1b. Como os *p-values* para este teste são inferiores ao nível de significância $\alpha = 0.05$, concluímos que o pressuposto da esfericidade da variância também não se cumpre.

Dado que os pressupostos do *Two-Way ANOVA Repeated Measures* não se verificam, não iremos realizar análise *Post-Hoc*. Em vez disso, consideramos uma alternativa não paramétrica ao *Two-Way ANOVA Repeated Measures*, o *Aligned Ranks Transformed ANOVA*, que apresentaremos na secção seguinte.

3.2 Aligned Ranks Transformed ANOVA

Analysis of Variance of Aligned Rank Transformed Data

Table Type: Repeated Measures Analysis of Variance Table (Type I)
Model: Repeated Measures (aov)
Response: art(time)

	Error Df	res	F value	Pr(>F)
1 vertices	seed 3	116	199.559	< 2.22e-16 ***
2 algorithm	sd:lg 2	232	442.008	< 2.22e-16 ***
3 vertices:algorithm	sd:lg 6	232	83.267	< 2.22e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(a) Summary do ART ANOVA

Aligned Rank Transform of Factorial Model

Call:
art(formula = time ~ vertices * algorithm + Error(seed/algorithm),
data = data)

Column sums of aligned responses (should all be ~0):
vertices algorithm vertices:algorithm
0 0 0

F values of ANOVAs on aligned responses not of interest (should all be ~0):
Min. 1st Qu. Median Mean 3rd Qu. Max.
0 0 0 0 0 0

(b) Summary do Modelo ART

Figura 3: Resultados do *Aligned Ranks Transformed ANOVA*

Para verificar se o *Aligned Ranks Transformed* foi corretamente aplicado e se é apropriado para os nossos dados, analisamos o *summary* do modelo ART, Figura 3b. Como todos os valores são aproximadamente zero, podemos aplicar o *Aligned Ranks Transformed ANOVA*. Na Figura 3a, como esperado, rejeitamos as hipóteses para os fatores *vertices*, *algorithm* e para a suas interações, Equações 3, 5 e 7, e passamos à análise *Post-Hoc*.

3.2.1 Análise Post-Hoc

contrast	estimate	SE	df	t.ratio	p.value
Dinic - EK	-156.87	6.26	232	-25.044	<.0001
Dinic - MPM	-165.36	6.26	232	-26.400	<.0001
EK - MPM	-8.49	6.26	232	-1.356	0.3660

Results are averaged over the levels of: vertices
P value adjustment: tukey method for comparing a family of 3 estimates

Figura 4: Contrast Test para o fator *algorithm*

Para a análise *Post-Hoc* do *Aligned Ranks Transformed ANOVA* utilizamos os *Contrast Tests* (ART-C) apenas para o fator *algorithm* dado que a nossa hipótese apenas contempla diferenças na média do tempo de execução para os algoritmos.

Segundo a Figura 4, rejeitamos as hipóteses dos efeitos *Dinic - EK* e *Dinic - MPM*. Isto significa que existem diferenças significativas entre o algoritmo **Dinic** e os restantes, **EK** e **MPM**. Por outro lado, não rejeitamos a hipótese do efeito *MPM - EK* evidenciando que não existem diferenças significativas entre os algoritmos **EK** e **MPM**, ou seja, não podemos afirmar que o algoritmo **EK** é mais rápido que o **MPM** para valores baixos de vértices e probabilidades.

4 Segunda Hipótese

Esta hipótese é semelhante à hipótese descrita na Secção 3, variando apenas os níveis do fator *vertices*, ou seja, considerámos os níveis 300, 350, ..., 1000.

4.1 Two-Way ANOVA Repeated Measures

À semelhança dos resultados da Secção 3.1, obtivemos um *p-value* inferior ao nível de significância para os fatores *vertices*, *algorithm* e *vertices:algorithm*, rejeitando, assim, a hipótese nula para as três hipóteses do *Two-way ANOVA*, Equações 3, 5 e 7.

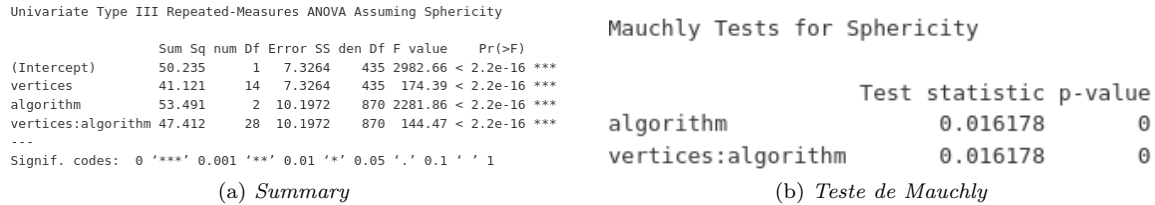


Figura 5: Resultados do *Two-way ANOVA Repeated Measures*

4.1.1 Pressupostos

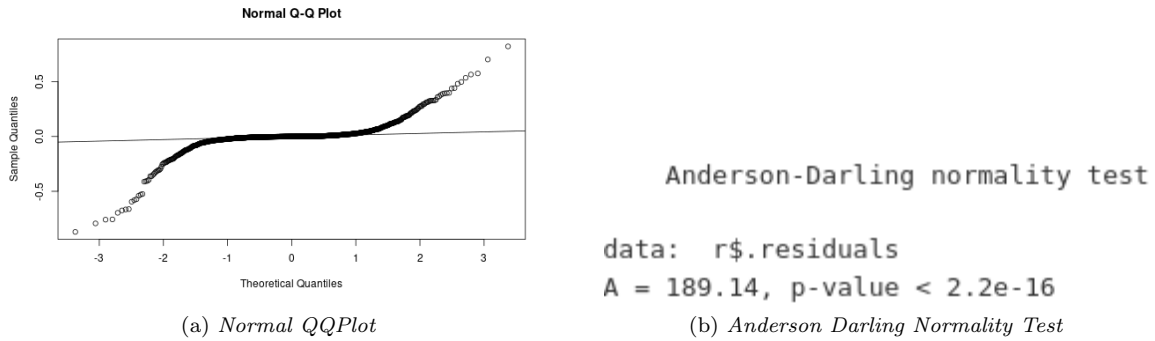


Figura 6: Normalidade dos Resíduos

À semelhança da análise dos pressupostos da Secção 3.1.1, verificamos, através dos testes estatísticos *Anderson-Darling* e de *Mauchly*, que os pressupostos da normalidade e da esfericidade não se cumprem. Por isso, aplicámos a alternativa não paramétrica ao *Two-Way ANOVA Repeated Measures*, o *Aligned Ranks Transformed ANOVA*.

4.2 Aligned Ranks Transformed ANOVA

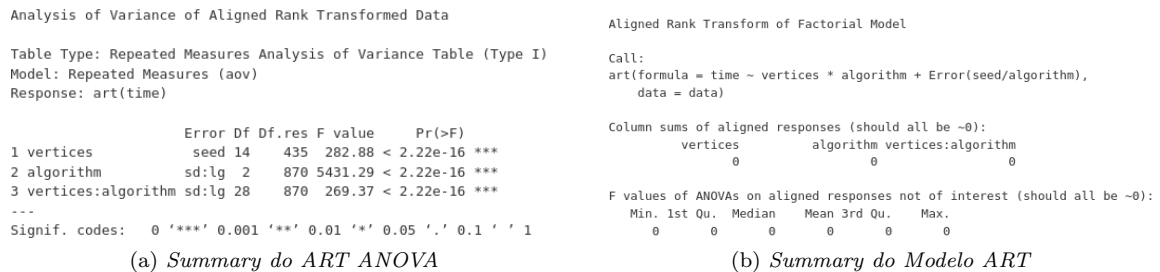


Figura 7: Resultados do *Aligned Ranks Transformed ANOVA*

Na Figura 7a, rejeitamos as hipóteses para os fatores *vertices*, *algorithm* e para a suas interações, Equações 3, 5 e 7. Por isso, analisamos à análise *Post-Hoc* do fator *algorithm*.

4.2.1 Análise Post-Hoc

contrast	estimate	SE	df	t.ratio	p.value
Dinic - EK	-832	7.98	870	-104.213	<.0001
Dinic - MPM	-406	7.98	870	-50.820	<.0001
EK - MPM	426	7.98	870	53.393	<.0001

Results are averaged over the levels of: vertices
P value adjustment: tukey method for comparing a family of 3 estimates

Figura 8: Contrast Test para o fator *algorithm*

Segundo a Figura 8, rejeitamos as hipóteses dos efeitos *Dinic - EK*, *Dinic - MPM*, *MPM - EK*. Isto significa que existem diferenças significativas entre todos os algoritmos, comprovando os resultados da Análise Exploratória de Dados, ou seja, o **Dinic** é mais rápido que o **MPM**, que, por sua vez, é mais rápido que o **EK**.

5 Terceira Hipótese

Para esta hipótese considerámos os níveis 100, 150, ..., 1000 do fator *vertices*, os níveis 0.2, 0.3, ..., 0.7 do fator *probability* e os níveis Dinic, EK e MPM do fator *algorithm*.

Como possuímos três fatores, optamos pelo teste estatístico *Three-Way ANOVA Repeated Measures*, onde o *subject* é a variável *seed*, o fator *within-subject* é o *algorithm* e os fatores *between-subject* são os fatores *vertices* e *probability*.

5.1 Three-Way ANOVA Repeated Measures

1. Não existe diferença na média do tempo de execução devido ao fator *vertices*

$$H_0^v : \mu_{1..} = \mu_{2..} = \dots = \mu_{v..} \quad (9)$$

$$H_1^v : \exists i, j : \mu_{i..} \neq \mu_{j..} \quad (10)$$

2. Não existe diferença na média do tempo de execução devido ao fator *probability*

$$H_0^p : \mu_{..1} = \mu_{..2} = \dots = \mu_{..a} \quad (11)$$

$$H_1^p : \exists i, j : \mu_{..i} \neq \mu_{..j} \quad (12)$$

3. Não existe diferença na média do tempo de execução devido ao fator *algorithm*

$$H_0^a : \mu_{..1} = \mu_{..2} = \dots = \mu_{..a} \quad (13)$$

$$H_1^a : \exists i, j : \mu_{..i} \neq \mu_{..j} \quad (14)$$

4. Não existem interações entre os fatores *vertices* e *probability*

$$H_0^{v:p} : \forall i, j : \gamma_{ij.} = 0 \quad (15)$$

$$H_1^{v:p} : \exists i, j : \gamma_{ij.} \neq 0 \quad (16)$$

5. Não existem interações entre os fatores *vertices* e *algorithm*

$$H_0^{v:a} : \forall i, j : \gamma_{i.j} = 0 \quad (17)$$

$$H_1^{v:a} : \exists i, j : \gamma_{i.j} \neq 0 \quad (18)$$

6. Não existem interações entre os fatores *probability* e *algorithm*

$$H_0^{p:a} : \forall i, j : \gamma_{i.j} = 0 \quad (19)$$

$$H_1^{p:a} : \exists i, j : \gamma_{i.j} \neq 0 \quad (20)$$

7. Não existem interações entre os fatores *vertices*, *probability* e *algorithm*

$$H_0^{v:p:a} : \forall i, j, k : \gamma_{ijk} = 0 \quad (21)$$

$$H_1^{v:p:a} : \exists i, j, k : \gamma_{ijk} \neq 0 \quad (22)$$

Univariate Type III Repeated-Measures ANOVA Assuming Sphericity						
	Sum Sq	num Df	Error SS	den Df	F value	
(Intercept)	11979.0	1	451.51	3306	87711.72	
vertices	10306.9	18	451.51	3306	4192.69	
probability	2795.8	5	451.51	3306	4094.31	
vertices:probability	2091.5	90	451.51	3306	170.16	
algorithm	18725.2	2	692.05	6612	89452.33	
vertices:algorithm	15998.5	36	692.05	6612	4245.92	
probability:algorithm	4904.0	10	692.05	6612	4685.37	
vertices:probability:algorithm	3726.1	180	692.05	6612	197.78	
	Pr(>F)					
(Intercept)	< 2.2e-16 ***					
vertices	< 2.2e-16 ***					
probability	< 2.2e-16 ***					
vertices:probability	< 2.2e-16 ***					
algorithm	< 2.2e-16 ***					
vertices:algorithm	< 2.2e-16 ***					
probability:algorithm	< 2.2e-16 ***					
vertices:probability:algorithm	< 2.2e-16 ***					

(a) Summary

Mauchly Tests for Sphericity

	Test statistic	p-value
algorithm	0.013084	0
vertices:algorithm	0.013084	0
probability:algorithm	0.013084	0
vertices:probability:algorithm	0.013084	0

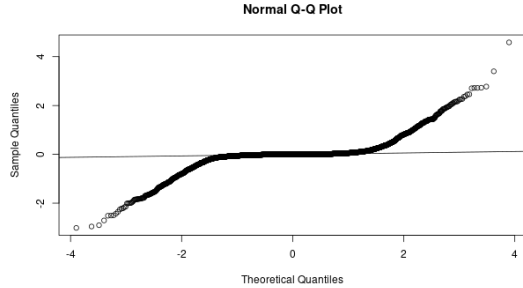
(b) Teste de Mauchly

Figura 9: Resultados do *Three-Way ANOVA Repeated Measures*

Observando os resultados da figura Figura 9a, obtivemos um p -value inferior a $2.2e^{-16}$ para os todos os fatores e para todas as interações. Como o p -value é inferior ao nível de significância $\alpha = 0.05$, rejeitamos a hipótese nula para das Equações 9, 11, 13, 15, 17, 19 e 21.

5.1.1 Pressupostos

Para podermos aceitar os resultados deste teste estatístico temos de analisar os pressupostos do *Three-Way ANOVA Repeated Measures*, a normalidade dos resíduos e a esfericidade da variância.



(a) Normal QQPlot

Anderson-Darling normality test

```
data: r$.residuals
A = 1685.9, p-value < 2.2e-16
```

(b) Anderson Darling Normality Test

Figura 10: Normalidade dos Resíduos

Recorrendo ao teste de normalidade de *Anderson-Darling* e ao teste de esfericidade de *Mauchly*, Figuras 10b e 9b, concluímos que o pressuposto da normalidade dos resíduos e o pressuposto esfericidade da variância não se verificam. Uma vez que os pressupostos do *Three-Way ANOVA Repeated Measures* não se verificam, não iremos realizar análise *Post-Hoc*. Em vez disso, considerámos a alternativa não paramétrica *Aligned Ranks Transformed ANOVA*.

5.2 Aligned Ranks Transformed ANOVA

```
Analysis of Variance of Aligned Rank Transformed Data

Table Type: Repeated Measures Analysis of Variance Table (Type I)
Model: Repeated Measures (aov)
Response: art(time)

      Error Df Df.res F value    Pr(>F)
1 vertices      seed 18 3306 6812.87 < 2.22e-16 ***
2 probability      seed 5 3306 6084.77 < 2.22e-16 ***
3 algorithm      sd:lg 2 6612 60952.26 < 2.22e-16 ***
4 vertices:probability      seed 90 3306 343.64 < 2.22e-16 ***
5 vertices:algorithm      sd:lg 36 6612 9233.19 < 2.22e-16 ***
6 probability:algorithm      sd:lg 10 6612 9203.01 < 2.22e-16 ***
7 vertices:probability:algorithm      sd:lg 180 6612 367.72 < 2.22e-16 ***

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(a) Summary do ART ANOVA

```
Aligned Rank Transform of Factorial Model

Call:
art(formula = time ~ vertices * probability * algorithm + Error(seed/algorithm),
    data = data)

Column sums of aligned responses (should all be ~0):
      vertices      probability
      0          0
      algorithm      vertices:probability
      0          0
      vertices:algorithm      probability:algorithm
      0          0
      vertices:probability:algorithm
      0

F values of ANOVAs on aligned responses not of interest (should all be ~0):
      Min. 1st Qu. Median      Mean 3rd Qu.      Max.
      0         0         0         0         0         0
```

(b) Summary do Modelo ART

Figura 11: Resultados do *Aligned Ranks Transformed ANOVA*

Na Figura 11a, rejeitamos as hipóteses das Equações 9, 11, 13, 15, 17, 19, 21.

5.2.1 Análise Post-Hoc

```
contrast estimate SE df t.ratio p.value
Dinic - EK      -6650 19.1 6612 -348.767 <.0001
Dinic - MPM     -3056 19.1 6612 -160.249 <.0001
EK - MPM         3595 19.1 6612 188.517 <.0001

Results are averaged over the levels of: vertices, probability
P value adjustment: tukey method for comparing a family of 3 estimates
```

Figura 12: Contrast Test para o fator *algorithm*

Segundo a Figura 12, rejeitamos as hipóteses dos efeitos *Dinic - EK*, *Dinic - MPM*, *MPM - EK*. Existem diferenças significativas entre todos os algoritmos, ou seja, o **Dinic** é mais rápido que o **MPM**, que, por sua vez, é mais rápido que o **EK**.

6 Quarta Hipótese

Como foi referido no relatório da primeira meta, considerar apenas as complexidades temporais teóricas omitem termos de menor grau da equação. Assim, testamos se os termos não presentes nessas expressões têm impacto, conforme as Equações 23, 24 e 25.

$$y_{dinic} = b_0 + b_1v + b_3e + b_4ve + b_6e^2 + b_2v^2 + b_5v^2e \quad (23)$$

$$y_{ek} = b_0 + b_1v + b_2e + b_4ve + b_3e^2 + b_6v^2 + b_5ve^2 \quad (24)$$

$$y_{mpm} = b_0 + b_1v + b_2e + b_3ve + b_4e^2 + b_5v^2 + b_6ve^2 + b_7v^2e + b_8v^3 \quad (25)$$

6.1 Hipóteses Formais (*F-Test*)

A influência de cada variável independente na variável dependente é testada pelo valor não nulo do coeficiente da variável independente:

$$H_0^0 : b_0 = 0 \quad H_0^1 : b_1 = 0 \quad \dots \quad H_0^k : b_k = 0 \quad (26)$$

$$H_1^0 : b_0 \neq 0 \quad H_1^1 : b_1 \neq 0 \quad \dots \quad H_1^k : b_k \neq 0 \quad (27)$$

6.2 Resultados

```
Call:
lm(formula = time ~ poly(vertices, 2, raw = TRUE) * poly(arcs, 2, raw = TRUE), data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-0.15507 -0.01151 -0.00087  0.00500  0.42439

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    4.461e-03  3.742e-03   1.192  0.23336
poly(vertices, 2, raw = TRUE)1 -5.840e-05  2.309e-05  -2.530  0.01145 *
poly(vertices, 2, raw = TRUE)2  7.570e-08  2.430e-08   3.115  0.00185 **
poly(arcs, 2, raw = TRUE)1    5.823e-07  3.107e-07   1.874  0.06098 .
poly(arcs, 2, raw = TRUE)2 -2.623e-11  3.909e-12  -6.710  2.22e-11 ***
poly(vertices, 2, raw = TRUE)1:poly(arcs, 2, raw = TRUE)1  2.198e-09  7.850e-10  2.800  0.00514 **
poly(vertices, 2, raw = TRUE)2:poly(arcs, 2, raw = TRUE)1 -9.562e-13  5.560e-13  -1.720  0.08556 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04174 on 3981 degrees of freedom
Multiple R-squared:  0.6841, Adjusted R-squared:  0.6835
F-statistic: 1078 on 8 and 3981 DF, p-value: < 2.2e-16
```

(a) Dinic

```
Call:
lm(formula = time ~ poly(vertices, 2, raw = TRUE) * poly(arcs, 2, raw = TRUE), data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-3.6253 -0.1670  0.0190  0.2302  3.6675

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    -2.913e-01  5.925e-02  -4.916  9.19e-07 ***
poly(vertices, 2, raw = TRUE)1  3.735e-03  3.655e-04  10.218  < 2e-16 ***
poly(vertices, 2, raw = TRUE)2 -5.846e-06  3.847e-07 -15.198  < 2e-16 ***
poly(arcs, 2, raw = TRUE)1    -3.765e-05  4.919e-06  -7.654  2.43e-14 ***
poly(arcs, 2, raw = TRUE)2    1.581e-09  6.180e-11  25.554  < 2e-16 ***
poly(vertices, 2, raw = TRUE)1:poly(arcs, 2, raw = TRUE)1  2.112e-08  1.243e-08   1.700  0.0893 .
poly(vertices, 2, raw = TRUE)2:poly(arcs, 2, raw = TRUE)2 -2.718e-12  1.266e-13 -21.463  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6608 on 3981 degrees of freedom
Multiple R-squared:  0.9607, Adjusted R-squared:  0.9606
F-statistic: 1.217e+04 on 8 and 3981 DF, p-value: < 2.2e-16
```

(b) EK

```
Call:
lm(formula = time ~ poly(vertices, 3, raw = TRUE) * poly(arcs,
2, raw = TRUE), data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-0.30316 -0.01436 -0.00022  0.00994  0.47507

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    -9.527e-05  9.001e-03  -0.011  0.992
poly(vertices, 3, raw = TRUE)1  1.796e-06  8.864e-05  0.020  0.984
poly(vertices, 3, raw = TRUE)2 -3.673e-08  2.138e-07  -0.172  0.864
poly(vertices, 3, raw = TRUE)3  1.529e-10  1.435e-10  1.066  0.287
poly(arcs, 2, raw = TRUE)1    1.925e-07  1.078e-06  0.179  0.858
poly(arcs, 2, raw = TRUE)2   -1.167e-11  2.263e-11  -0.515  0.606
poly(vertices, 3, raw = TRUE)1:poly(arcs, 2, raw = TRUE)1  4.602e-09  5.022e-09  0.916  0.360
poly(vertices, 3, raw = TRUE)2:poly(arcs, 2, raw = TRUE)1 -4.738e-12  7.817e-12  -0.606  0.544
poly(vertices, 3, raw = TRUE)3:poly(arcs, 2, raw = TRUE)1  1.325e-15  3.898e-15  0.340  0.734
poly(vertices, 3, raw = TRUE)1:poly(arcs, 2, raw = TRUE)2  3.945e-14  7.461e-14  0.529  0.597
poly(vertices, 3, raw = TRUE)2:poly(arcs, 2, raw = TRUE)2 -4.027e-17  8.272e-17  -0.487  0.626
poly(vertices, 3, raw = TRUE)3:poly(arcs, 2, raw = TRUE)2  1.531e-20  3.087e-20  0.496  0.620

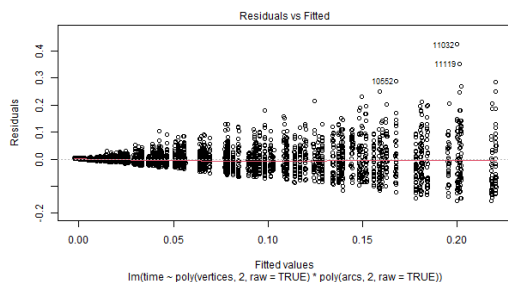
Residual standard error: 0.05624 on 3978 degrees of freedom
Multiple R-squared:  0.9235, Adjusted R-squared:  0.9233
F-statistic: 4365 on 11 and 3978 DF, p-value: < 2.2e-16
```

(c) MPM

Figura 13: Resultados da regressão polinomial

Analisando a Figura 13, e assumindo $\alpha = 0.05$, pelos valores do *p-value*, no **Dinic**, Figura 13a, rejeitam-se as hipóteses nulas nos termos v , ve , v^2 e e^2 . No **EK**, Figura 13b, rejeitam-se as hipóteses nulas nos termos da ordenada na origem, v , v^2 , e , e^2 , v e ve^2 . No **MPM**, Figura 13c, não se rejeita nenhuma hipótese nula.

6.3 Pressupostos



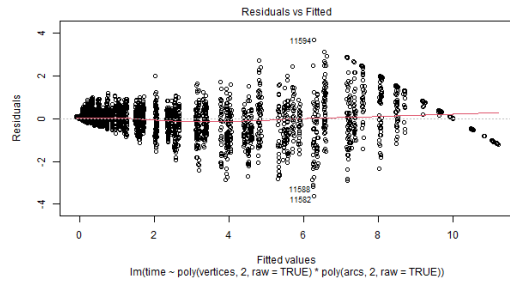
(a) Residual vs Fitted Plot

```
shapiro-wilk normality test

data: residuals(aov.out)
W = 0.81892, p-value < 2.2e-16
```

(b) Shapiro Test

Figura 14: Regressão polinomial do Dinic



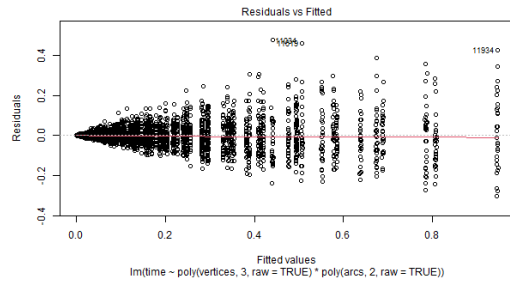
(a) *Residual vs Fitted Plot*

shapiro-wilk normality test

```
data: residuals(aov.out)
w = 0.90625, p-value < 2.2e-16
```

(b) *Shapiro Test*

Figura 15: Regressão polinomial do EK



(a) *Residual vs Fitted Plot*

shapiro-wilk normality test

```
data: residuals(aov.out)
w = 0.81835, p-value < 2.2e-16
```

(b) *Shapiro Test*

Figura 16: Regressão polinomial do MPM

Para todos os 3 algoritmos, **Dinic**, **EK** e **MPM**, os pressupostos da homocedasticidade, observado nos *Residual vs Fitted Plots*, Figuras 14a, 15a e 16a, e da normalidade dos resíduos, corroborado no *Shapiro-Wilk Normality Test*, Figuras 14b, 15b e 16b, não se verificam. Assim, e não podendo concluir nada acerca destas regressões polinomiais, não é possível ter certeza acerca das rejeições das hipóteses nulas referidas na Subseção 6.2.

7 Conclusão

Neste trabalho verificamos a validade de várias hipóteses relativas ao desempenho de três algoritmos. Para tal, recorremos a vários testes estatísticos.

Além de efetuar os testes estatísticos, avaliámos também o quão bem os testes utilizados se adequavam aos nossos dados, através da análise dos seus pressupostos.

Usando teste paramétricos, não conseguimos validar os pressupostos, mas aplicando transformações não paramétricas, conseguimos rejeitar as hipóteses nulas das 2.^a e 3.^a hipóteses, sendo as restantes inconclusivas.

8 Referências

- [1] L. Paquete, “Measurements,” 2022.
- [2] R. Pi, “Raspberry pi documentation - processors,” <https://www.raspberrypi.com/documentation/computers/processors.html#bcm2711>, 2022.
- [3] D. J. Lilja, *Measuring computer performance: a practitioners guide*. Cambridge University Press, 2005.
- [4] C. C. McGeoch, *A Guide to experimental algorithmics*. Cambridge University Press, 2012.
- [5] L. Paquete, “Exploratory data analysis,” 2022.
- [6] —, “Linear regression,” 2022.
- [7] M. Kay, L. A. Elkin, J. J. Higgins, and J. O. Wobbrock, “Artool: R package for the aligned rank transform for nonparametric factorial anovas,” 2021. [Online]. Available: <https://github.com/mjskay/ARTool>
- [8] M. Kay, L. A. Elkin, and J. O. Wobbrock, Oct 2021. [Online]. Available: <https://cran.r-project.org/web/packages/ARTool/vignettes/art-contrasts.html>