



Análise de Redes - Trabalho de Grupo n.º 1

Ciência de Dados - PL - 3º ano | Professora: Maria João Frazão Lopes

Catarina Castanheira, 92478

João Martins, 93259

Joel Paula, 93392

26/11/2021

Enunciado

Este trabalho consiste na análise de duas redes.

Numa destas redes representam-se as finais da Taça de Portugal desde a época de 2000-01 até à época de 2020-21, com excepção das épocas de 2003-04, 2007-08 e 2012-13. Cada nodo representa uma equipa e cada ligação uma final. Nas tabelas seguintes indicam-se as equipas e as finais.

Table 1: Tabela 1 - Finais da Taça de Portugal - Nodos da rede

| Nodos | Equipa |
|-------|----------------------|
| 1 | Académica |
| 2 | Belenenses |
| 3 | Benfica |
| 4 | Braga |
| 5 | Chaves |
| 6 | Desportivo das Aves |
| 7 | Leixões |
| 8 | Marítimo |
| 9 | Paços de Ferreira |
| 10 | Porto |
| 11 | Rio Ave |
| 12 | Sporting CP |
| 13 | União de Leiria |
| 14 | Vitória de Guimarães |
| 15 | Vitória de Setúbal |

Table 2: Tabela 2 - Finais da Taça de Portugal - Arestas da rede

| Equipa | Equipa |
|---------------------|----------------------|
| Académica | Sporting CP |
| Belenenses | Sporting CP |
| Benfica | Braga |
| Benfica | Porto |
| Benfica | Rio Ave |
| Benfica | Vitória de Guimarães |
| Benfica | Vitória de Setúbal |
| Braga | Porto |
| Braga | Sporting CP |
| Chaves | Porto |
| Desportivo das Aves | Sporting CP |
| Leixões | Sporting CP |
| Marítimo | Porto |
| Paços de Ferreira | Porto |
| Porto | Sporting CP |
| Porto | União de Leiria |
| Porto | Vitória de Guimarães |
| Porto | Vitória de Setúbal |

A segunda rede é uma rede aleatória com 15 nodos e 18 ligações. Para obter esta rede, deve utilizar-se a função `sample_gnm` com os parâmetros (n=15,m=18).

Q1.

Estude cada uma das redes quanto à densidade, ao grau dos nodos e grau médio.

Compare os resultados obtidos para as duas redes.

Interprete os valores obtidos para a primeira rede, no contexto da sua natureza.

Grafo 1 - finais da Taça de Portugal

```
### Graph 1
graph1 <- read_graph("Trab1.txt",
                     format = c("edgelist"),
                     n = 15,
                     directed = F)
plot(graph1, ) #, edge.curved=0.2, edge.arrow.size=0.5, edge.arrow.width=1)
```

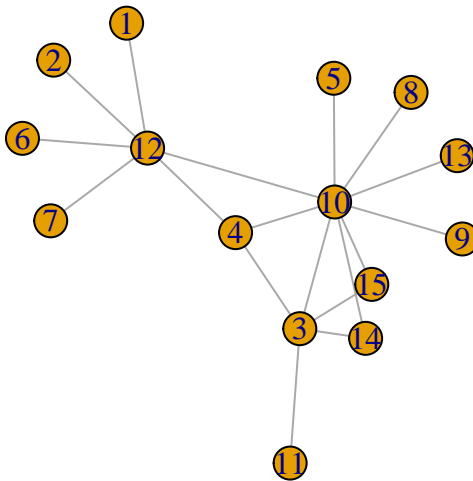


Figure 1: Figura 1 - rede de Finais da Taça de Portugal

```
edge_density(graph1, loops = F)
```

```
## [1] 0.1714286
```

```
# 0.08571429
sum(degree(graph1, mode = "total"))
```

```
## [1] 36
```

```
# grau total -> 36
mean(degree(graph1, mode = "total"))
```

```
## [1] 2.4
```

```
# grau médio -> 2.4
```

A densidade é baixa, uma vez que nem todas as equipas tiveram finais entre si e portanto não existem todas as arestas que poderiam existir para ser um grafo completo.

O grau é 36, uma vez que correspondem a 18 jogos entre um par de equipas, em 18 épocas.

O grau médio é elevado, mas não é um bom indicador do grau dos seus nodos, pois apesar de 9 das equipas apenas terem estado presentes uma vez e 3 delas apenas 2 ou 3 vezes, cada uma das 3 equipas mais presentes - Benfica, Sporting e Porto - estiveram presentes em todas as finais: Benfica = 5 (uma com o Porto), Porto = 9 (uma com Benfica e outra com Sporting), Sporting = 6 (1 com porto).

Grafo 2 - rede aleatória

```
### Graph 2
set.seed(42)
graph2 <- sample_gnm(15, 18)
plot(graph2)
```

```
edge_density(graph2, loops = F)
```

```
## [1] 0.1714286
```

```
# 0.1714286
sum(degree(graph2, mode = "total"))
```

```
## [1] 36
```

```
# grau total (grau incidente + grau divergente) -> 36
mean(degree(graph2, mode = "total"))
```

```
## [1] 2.4
```

```
# grau médio -> 2.4
```

A densidade é baixa, uma vez que nem todos os nodos estão ligados entre si diretamente - não é um grafo completo.

O grau é 36, o mesmo do anterior, já que o número de nodos e de arestas é o mesmo da rede anterior.

O grau médio acaba por espelhar melhor o tipo de grafo do que no caso anterior.

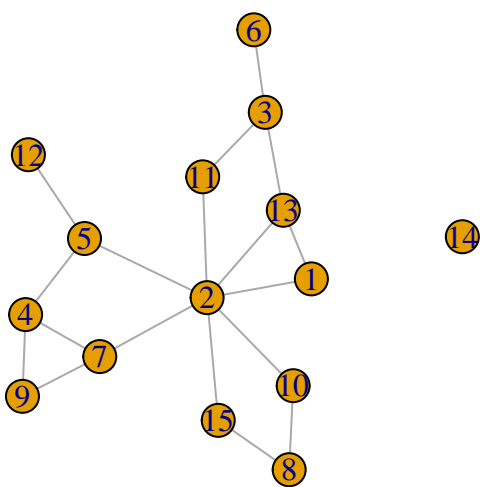


Figure 2: Figura 2 - rede aleatória

Q2.

Estude cada uma das redes quanto à associação de grau, aos comprimentos dos caminhos mais curtos, à média dos comprimentos dos caminhos mais curtos, à conectividade, aos coeficientes de *clustering* dos nodos e da rede.

Compare os resultados obtidos para as duas redes.

Interprete os valores obtidos para a primeira rede, no contexto da sua natureza.

Rede 1 - finais da Taça

Começamos pelo Grafo 1, calculando a correlação de grau.

```
### Graph 1
assortativity_degree(graph1, directed = F)
```

```
## [1] -0.6627219
```

Verificamos que a correlação é negativa, pelo que concluímos que a *rede é não associativa*.

Este resultado não surpreende, pois os nodos de menor grau encontram-se ligados aos de maior grau, não existindo paridade de grau entre os nodos ligados entre si.

Olhando para as distancias:

```
options(width = 90)
distances(graph1)
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## [1,]    0    2    3    2    3    2    2    3    3    2    4    1    3    3    3
## [2,]    2    0    3    2    3    2    2    3    3    2    4    1    3    3    3
## [3,]    3    3    0    1    2    3    3    2    2    1    1    2    2    1    1
## [4,]    2    2    1    0    2    2    2    2    2    1    2    1    2    2    2
## [5,]    3    3    2    2    0    3    3    2    2    1    3    2    2    2    2
## [6,]    2    2    3    2    3    0    2    3    3    2    4    1    3    3    3
## [7,]    2    2    3    2    3    2    0    3    3    2    4    1    3    3    3
## [8,]    3    3    2    2    2    3    3    0    2    1    3    2    2    2    2
## [9,]    3    3    2    2    2    3    3    2    0    1    3    2    2    2    2
## [10,]   2    2    1    1    1    2    2    1    1    0    2    1    1    1    1
## [11,]   4    4    1    2    3    4    4    3    3    2    0    3    3    2    2
## [12,]   1    1    2    1    2    1    1    2    2    1    3    0    2    2    2
## [13,]   3    3    2    2    2    3    3    2    2    1    3    2    0    2    2
## [14,]   3    3    1    2    2    3    3    2    2    1    2    2    2    0    2
## [15,]   3    3    1    2    2    3    3    2    2    1    2    2    2    2    0
```

```
# média dos comprimentos dos caminhos mais curtos
mean_distance(graph1, directed = F)
```

```
## [1] 2.219048
```

```
# -> 1.485714
diameter(graph1, directed = F) # 4, para não direcionado
```

```
## [1] 4
```

A rede diz-se *conexa* porque existe um caminho entre qualquer par de nodos.

O diâmetro de 4 acaba por confirmar o facto de existirem equipas que participaram em mais do que uma final e que esta rede representa um “*small world*”.

```
transitivity(graph1)
```

```
## [1] 0.1818182
```

```
# Global clustering coefficient -> 0.1818182
transitivity(graph1, type = "average")
```

```
## [1] 0.5240741
```

```
# Average clustering coefficient -> 0.5240741
```

O Coeficiente de *clustering* é bem menor que 1, o que indica que existe fraca probabilidade que duas equipas que tenham jogado uma final, tenham ambas jogado uma final com uma outra equipa. Isso indica que não é fácil para algumas equipas chegarem várias vezes à final da Taça.

Rede 2 - aleatória

```
### Graph 2
assortativity_degree(graph2, directed = F)
```

```
## [1] -0.25
```

```
# 0.02857143
options(width = 90)
distances(graph2)
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## [1,]    0    1    2    3    2    3    2    3    3    2    2    3    1  Inf    2
## [2,]    1    0    2    2    1    3    1    2    2    1    1    2    1  Inf    1
## [3,]    2    2    0    4    3    1    3    4    4    3    1    4    1  Inf    3
## [4,]    3    2    4    0    1    5    1    4    1    3    3    2    3  Inf    3
## [5,]    2    1    3    1    0    4    2    3    2    2    2    1    2  Inf    2
## [6,]    3    3    1    5    4    0    4    5    5    4    2    5    2  Inf    4
## [7,]    2    1    3    1    2    4    0    3    1    2    2    3    2  Inf    2
## [8,]    3    2    4    4    3    5    3    0    4    1    3    4    3  Inf    1
## [9,]    3    2    4    1    2    5    1    4    0    3    3    3    3  Inf    3
## [10,]   2    1    3    3    2    4    2    1    3    0    2    3    2  Inf    2
## [11,]   2    1    1    3    2    2    2    3    3    2    0    3    2  Inf    2
```

```
## [12,] 3 2 4 2 1 5 3 4 3 3 3 0 3 Inf 3
## [13,] 1 1 1 3 2 2 2 3 3 2 2 3 0 Inf 2
## [14,] Inf Inf Inf Inf Inf Inf Inf Inf Inf Inf Inf Inf 0 Inf
## [15,] 2 1 3 3 2 4 2 1 3 2 2 3 2 Inf 0
```

```
mean_distance(graph2, directed = F)
```

```
## [1] 2.483516
```

```
# média dos comprimentos dos caminhos mais curtos -> 2.483516
diameter(graph2, directed = F) # 5
```

```
## [1] 5
```

Concluimos que, sendo a correlação negativa, a rede é *não associativa*. O que se pode observar por nodos de grau maior se ligarem a nodos de grau menor.

A rede é composta por duas subredes - uma desconexa composta do nodo 14, sem qualquer conexão, e outra componente gigante e conexa.

A componente gigante diz-se *conexa* uma vez que existe um caminho entre qualquer par de nodos.

```
transitivity(graph2)
```

```
## [1] 0.1428571
```

```
# Global clustering coefficient -> 0.1428571
transitivity(graph2, type = "average")
```

```
## [1] 0.2539683
```

```
# Average clustering coefficient -> 0.2539683
```

O Coeficiente de *clustering* também é bem menor que 1, o que indica que não existem relações próximas entre todos os nodos.

Q3.

Determine para cada uma das redes as medidas de centralidade e o parâmetro de heterogeneidade. Interprete os valores obtidos.

O que pode concluir quanto à existência de *hubs*? Quais serão os *hubs*, no caso de existirem? Justifique.

Efetue a decomposição de core de cada uma das redes.

Compare os resultados obtidos para as duas redes.

Interprete os valores obtidos para a primeira rede, no contexto da sua natureza.

Q3.1 - Determine para cada uma das redes as medidas de centralidade e o parâmetro de heterogeneidade. Interprete os valores obtidos.

Rede 1 - Finais da Taça de Portugal

```
# 1. Análise dos graus de cada nodo  
(degrees <- degree(graph1, mode="all"))
```

```
## [1] 1 1 5 3 1 1 1 1 1 9 1 6 1 2 2
```

```
# Representação gráfica dos nodos, atribuindo uma proporção no tamanho da sua representação  
# de acordo com o grau respectivo  
plot(graph1, vertex.size=15+degrees*5)
```

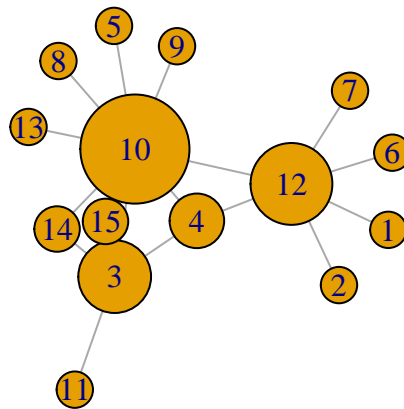


Figure 3: Figura 3 - rede de Finais da Taça de Portugal assinalando grau

```
df_taca <- data.frame(  
  Nodo=1:length(degree(graph1, mode="all")),  
  Equipa=c(  
    "Académica",  
    "Belenenses",  
    "Benfica",
```

```

    "Braga",
    "Chaves",
    "Desportivo das Aves",
    "Leixões",
    "Marítimo",
    "Paços de Ferreira",
    "Porto",
    "Rio Ave",
    "Sporting CP",
    "União de Leiria",
    "Vitória de Guimarães",
    "Vitória de Setúbal"
  ),
  Grau=degrees
)
kable(df_taca, caption="Tabela 3 - Nodos da rede das finais da Taça de Portugal indicando o grau")

```

Table 3: Tabela 3 - Nodos da rede das finais da Taça de Portugal indicando o grau

| Nodo | Equipa | Grau |
|------|----------------------|------|
| 1 | Académica | 1 |
| 2 | Belenenses | 1 |
| 3 | Benfica | 5 |
| 4 | Braga | 3 |
| 5 | Chaves | 1 |
| 6 | Desportivo das Aves | 1 |
| 7 | Leixões | 1 |
| 8 | Marítimo | 1 |
| 9 | Paços de Ferreira | 1 |
| 10 | Porto | 9 |
| 11 | Rio Ave | 1 |
| 12 | Sporting CP | 6 |
| 13 | União de Leiria | 1 |
| 14 | Vitória de Guimarães | 2 |
| 15 | Vitória de Setúbal | 2 |

```

# análise gráfica das frequências de cada nó:
par(mfrow=c(1,2))
barplot(table(df_taca$Grau), xlab="Grau", ylab="Frequência Absoluta",
        ylim=c(0,10), xlim=c(0,10), main="Distribuição de Centralidade I" )
barplot(prop.table(table(df_taca$Grau)), xlab="Grau", ylab="Frequência Relativa",
        ylim=c(0,1), xlim=c(0,10), main="Distribuição de Centralidade II")

```

2. Proximidade e Centralidade de Proximidade

```

#cálculo da Proximidade de cada um dos nodos aos restantes e guardamos na tabela df_taca:
df_taca$Proximidade <- round(closeness(graph1, mode="all", weights=NA), 3)

```

centralidade de proximidade normalizada

```

(df_taca$Centralidade_de_Proximidade <- round(centr_clo(graph1, mode="all", normalized=T)$res, 3))

```

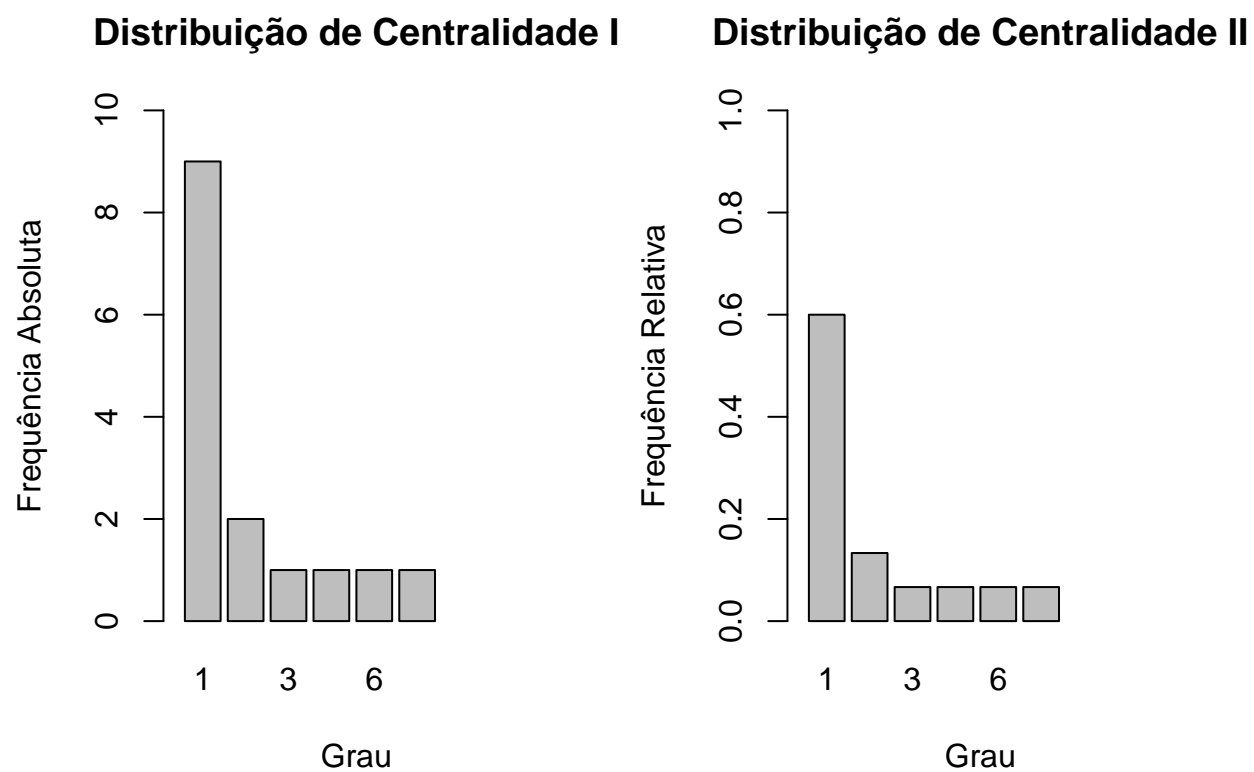


Figure 4: Figura 4 - Distribuição de centralidade da rede de Finais da Taça de Portugal

```
## [1] 0.389 0.389 0.519 0.560 0.438 0.389 0.389 0.438 0.438 0.737 0.350 0.609 0.438 0.467
## [15] 0.467
```

3. Intermediação e Centralidade de Intermediação

#intermediação de ligações:

```
intermediacao_ligacoes <- round(edge_betweenness(graph1, directed=F, weights=NA),3)
cat("Intermediação de Ligações: ", intermediacao_ligacoes)
```

```
## Intermediação de Ligações: 14 14 8 15 14 3 3 6 10 14 14 14 14 40 14 11 11
```

centralidade de intermediação de nodos

```
df_taca$Centralidade_de_Intermediacao <- round(betweenness(graph1, directed=F, normalized=T), 3)
cat("Centralidade de Intermediação Normalizada: ", df_taca$Centralidade_de_Intermediacao)
```

```
## Centralidade de Intermediação Normalizada: 0 0 0.159 0.055 0 0 0 0 0 0.687 0 0.505 0 0 0
```

4. Heterogeneidade

```
(heterogeneidade <- mean(degrees^2)/(mean(degrees)^2))
```

```
## [1] 1.944444
```

```
kable(df_taca, caption = "Tabela 4 - Finais da Taça de Portugal - Nodos da rede com graus e medidas de centralidade")
```

Table 4: Tabela 4 - Finais da Taça de Portugal - Nodos da rede com graus e medidas de centralidade

| Nodo | Equipa | Grau | Proximidade | Centralidade_de_Proximidade | Centralidade_de_Intermediacao |
|------|----------------------|------|-------------|-----------------------------|-------------------------------|
| 1 | Académica | 1 | 0.028 | 0.389 | 0.000 |
| 2 | Belenenses | 1 | 0.028 | 0.389 | 0.000 |
| 3 | Benfica | 5 | 0.037 | 0.519 | 0.159 |
| 4 | Braga | 3 | 0.040 | 0.560 | 0.055 |
| 5 | Chaves | 1 | 0.031 | 0.438 | 0.000 |
| 6 | Desportivo das Aves | 1 | 0.028 | 0.389 | 0.000 |
| 7 | Leixões | 1 | 0.028 | 0.389 | 0.000 |
| 8 | Marítimo | 1 | 0.031 | 0.438 | 0.000 |
| 9 | Paços de Ferreira | 1 | 0.031 | 0.438 | 0.000 |
| 10 | Porto | 9 | 0.053 | 0.737 | 0.687 |
| 11 | Rio Ave | 1 | 0.025 | 0.350 | 0.000 |
| 12 | Sporting CP | 6 | 0.043 | 0.609 | 0.505 |
| 13 | União de Leiria | 1 | 0.031 | 0.438 | 0.000 |
| 14 | Vitória de Guimarães | 2 | 0.033 | 0.467 | 0.000 |
| 15 | Vitória de Setúbal | 2 | 0.033 | 0.467 | 0.000 |

No que diz respeito à distribuição da centralidade dos nodos desta rede, é possível perceber que a maioria destes tem como grau o valor um, ou seja, a maioria das equipas (60% delas) só participou uma vez numa final da Taça no período em análise - têm somente uma ligação. Depois, temos duas equipas que já tem

duas participações e finalmente quatro que têm mais que três participações no período: Braga (3), Benfica (5), Sporting CP (6) e Porto (9).

Se olharmos para a medida de centralidade de proximidade, o que verificamos é que quanto maior o grau de dado nodo, maior o valor para a centralidade de proximidade. Estando perante valores normalizados, quanto mais próximo de 1 mais próximo determinado nodo está de ser “o” nodo central (a centralidade de proximidade será máxima no caso de um grafo do tipo estrela (<https://igraph.org/r/doc/centralize.html>)) - o equivalente a termos uma mesma equipa em participar em todas as finais. Significa isto que aquelas equipas já identificadas anteriormente como tendo os graus mais elevados, são também elas cujos nodos se encontram mais próximos de todos os restantes (Porto com 0.737, Sporting CP com 0.609, Braga com 0.56, e Benfica com 0.518).

As medidas de intermediação, também referentes a centralidade, dizem-nos que somente há quatro equipas ou nodos em que passam caminhos mais curtos (*shortest path*): Braga com 5.5% dos caminhos mais curtos da rede a passarem pelo seu nodo; Benfica com 15.9%; Sporting CP com 50.5%, e Porto com 68.7%.

Finalmente, a análise da heterogeneidade indica-nos um valor de 1.94. Sendo um valor superior a 1, sugere a existência de alguma heterogeneidade nos dados, ou seja, a existência de nodos com graus diferentes dos restantes.

Rede 2 - Rede Aleatória

```
# 1. Análise dos graus de cada nodo
(degrees2 <- degree(graph2, mode="all"))
```

```
## [1] 2 7 3 3 3 1 3 2 2 2 2 1 3 0 2
```

```
# Representação gráfica dos nodos, atribuindo uma proporção no tamanho da sua representação
# de acordo com o grau respectivo
plot(graph2, vertex.size=16+degrees2*5)
```

```
df_rede_aleatoria <- data.frame(
  Nodo=1:length(degree(graph2, mode="all")),
  Grau=degrees2
)
kable(df_rede_aleatoria, caption = "Tabela 5 - rede aleatória - Nodos da rede com graus")
```

Table 5: Tabela 5 - rede aleatória - Nodos da rede com graus

| Nodo | Grau |
|------|------|
| 1 | 2 |
| 2 | 7 |
| 3 | 3 |
| 4 | 3 |
| 5 | 3 |
| 6 | 1 |
| 7 | 3 |
| 8 | 2 |
| 9 | 2 |
| 10 | 2 |

| Nodo | Grau |
|------|------|
| 11 | 2 |
| 12 | 1 |
| 13 | 3 |
| 14 | 0 |
| 15 | 2 |

```
# análise gráfica das frequências de cada nó:
```

```
par(mfrow=c(1,2))
```

```
barplot(table(df_rede_aleatoria$Grau), xlab="Grau", ylab="Frequência Absoluta", ylim=c(0,10), xlim=c(0,
```

```
barplot(prop.table(table(df_rede_aleatoria$Grau)), xlab="Grau", ylab="Frequência Relativa", ylim=c(0,1)
```

```
# 2. Proximidade e Centralidade de Proximidade
```

```
# cálculo da Proximidade de cada um dos nodos aos restantes e
```

```
# guardamos na tabela df_rede_aleatoria:
```

```
df_rede_aleatoria$Proximidade <- closeness(graph2, mode="all", weights=NA)
```

```
## Warning in closeness(graph2, mode = "all", weights = NA): At centrality.c:2874 :closeness
```

```
## centrality is not well-defined for disconnected graphs
```

```
# centralidade de proximidade normalizada
```

```
(df_rede_aleatoria$Centralidade_de_Proximidade <- centr_clo(graph2,
```

```
mode="all", normalized=T)$res)
```

```
## Warning in centr_clo(graph2, mode = "all", normalized = T): At
```

```
## centrality.c:2874 :closeness centrality is not well-defined for disconnected graphs
```

```
## [1] 0.31818182 0.40000000 0.28000000 0.28000000 0.33333333 0.22580645 0.32558140
```

```
## [8] 0.25454545 0.26923077 0.31111111 0.32558140 0.25925926 0.33333333 0.06666667
```

```
## [15] 0.31111111
```

```
# 3. Intermediação e Centralidade de Intermediação
```

```
#intermediação de ligações:
```

```
intermediacao_ligacoes2 <- edge_betweenness(graph2, directed=F, weights=NA)
```

```
cat("Intermediação de Ligações: ", intermediacao_ligacoes2)
```

```
## Intermediação de Ligações: 10 23.5 9.5 13 23.5 6.5 3 10 17 7 19.5 11.5 13 3 18.5 13.5 17 7
```

```
# centralidade de intermediação de nodos
```

```
(df_rede_aleatoria$Centralidade_de_Intermediacao <- round(betweenness(graph2,
```

```
directed=F, normalized=T), 3))
```

```
## [1] 0.000 0.637 0.137 0.033 0.181 0.000 0.148 0.005 0.000 0.060 0.099 0.000 0.121 0.000
```

```
## [15] 0.060
```

```
# 4. Heterogeneidade
```

```
(heterogeneidade2 <- mean(degrees2^2)/(mean(degrees2)^2))
```

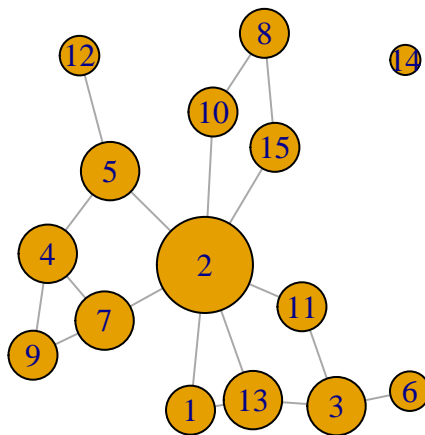


Figure 5: Figura 5 - rede aleatória assinalando grau

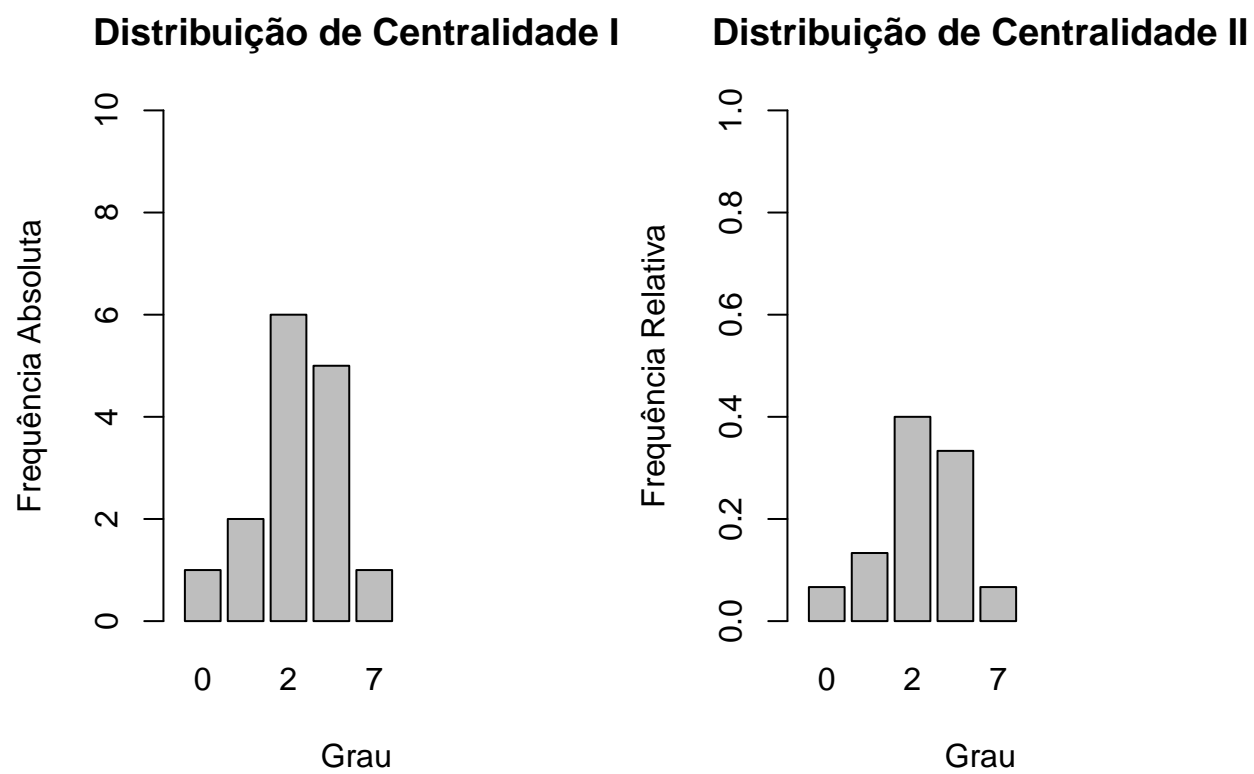


Figure 6: Figura 6 - rede aleatória medidas de centralidade

[1] 1.388889

Analisando os indicadores de centralidade da rede aleatória, a primeira nota que fazemos é que 11 dos seus 15 nodos têm grau 2 ou grau 3 (60% e 33.33% dos nodos, respectivamente). A representação gráfica da rede destaca claramente o nodo de grau 7 numa posição central do grafo. Este nodo é aquele que apresenta o valor de centralidade mais elevado (0.40) e também o valor mais elevado para a centralidade de intermediação (0.637), significando isto que este nodo é o mais próximo de todos os restantes e também que quase 64% dos caminhos mais curtos da rede atravessam este nodo. Dos restantes nodos, aqueles que apresentam valores mais elevados para a centralidade de proximidade e de intermediação são alguns de grau 3 (nodos 3, 5 e 7); contudo, os seus valores parecem mais próximos daqueles que os nodos de menor grau observam do que aqueles que temos no nodo de grau 7.

Sobre a heterogeneidade da rede, obtemos um valor de 1.38. Apesar de este valor sugerir alguma diferenciação nos dados, por ser superior a 1, esta diferenciação pode não ser muito relevante (1.38 não é muito maior que 1).

Q3.2 - O que pode concluir quanto à existência de *hubs*? Quais serão os *hubs*, no caso de existirem? Justifique.

Rede 1 - Finais da Taça de Portugal

Com base na informação obtida, podemos concluir que existem *hubs* no grafo que representa as finais da Taça de Portugal: verificamos que três equipas se destacam no número de participações das restantes (Porto, Sporting CP, Benfica; com 9, 6, e 5 participações, respectivamente, contra uma participação na maioria das restantes); quatro das equipas (Porto, Sporting CP, Benfica, Braga) têm valores razoáveis de centralidade de proximidade, o que significa que os seus nodos se encontram próximos dos restantes, e por consequência centrais aos restantes nodos; finalmente, os valores da centralidade de intermediação mostram que os caminhos mais curtos da rede passam em somente quatro nodos dos quinze - Porto, Sporting CP, Benfica e Braga -, embora com representações distintas. Verificamos que o parâmetro de heterogeneidade, sendo superior a 1, também nos dá sugestão de alguma diferenciação da representação de informação pelos nodos, ou seja, nodos mais centrais que outros.

Os nodos que se destacam nesta análise acabam por ser os referentes à equipa do Porto e do Sporting CP, cada um deles a obter elevados valores para as medidas de centralidade e a serem os dois nodos com maior grau. São estas as equipas que mais jogaram finais da Taça de Portugal no período referido. Com certeza, estes são os principais *hubs* do grafo. Consoante o “nível de inclusão” pretendido, os outros dois candidatos a integrarem o conjunto de *hubs* serão também os nodos do Benfica e do Braga.

Rede 2 - Rede Aleatória

Na rede aleatória, parece-nos existir um nodo que se destaca claramente dos outros e que se apresenta como único candidato a ser caracterizado como *hub*. Isto porque este nodo (nodo 2) é um nodo que se destaca dos restantes pelo seu grau (grau 7, contra um máximo de grau 3 dos restantes nodos), pela sua centralidade de proximidade normalizada (é o nodo mais próximo em termos de distância de qualquer um dos restantes), e por fim também pela sua centralidade de intermediação: é claramente o nodo pelo qual passam mais caminhos mais curtos de toda a rede - cerca de 64%.

Q3.3 - Efetue a decomposição de core de cada uma das redes. Compare os resultados obtidos para as duas redes.

Rede 1 - Finais da Taça de Portugal

```
# 6. Decomposição de Core
```

```
(df_taca$K_Core_Decomposition <- round(coreness(graph1, mode="all"),3))
```

```
## [1] 1 1 2 2 1 1 1 1 2 1 2 1 2 2
```

```
plot(graph1, vertex.size = 15+df_taca$K_Core_Decomposition*10,  
      vertex.label=df_taca$K_Core_Decomposition)
```

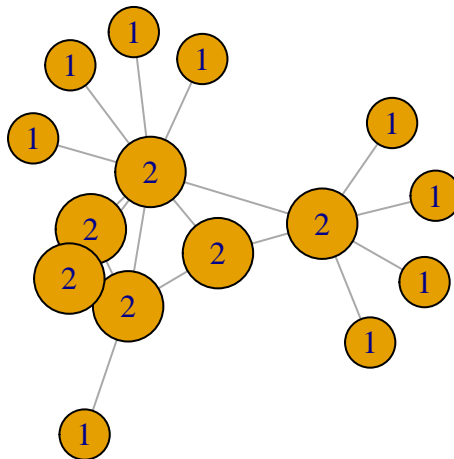


Figure 7: Figura 7 - rede de Finais da Taça de Portugal assinalando decomposição de core

```
kable(df_taca, caption = "Tabela 6 - Finais da Taça de Portugal - Nodos da rede com decomposição de core")
```

Table 6: Tabela 6 - Finais da Taça de Portugal - Nodos da rede com decomposição de core

| Nodo | Equipa | Grau | Proximidade | Centralidade_de_Proximidade | Centralidade_de_Intermediação | K_Core | Core_Decomposition |
|------|----------------------|------|-------------|-----------------------------|-------------------------------|--------|--------------------|
| 1 | Académica | 1 | 0.028 | 0.389 | 0.000 | 1 | 1 |
| 2 | Belenenses | 1 | 0.028 | 0.389 | 0.000 | 1 | 1 |
| 3 | Benfica | 5 | 0.037 | 0.519 | 0.159 | 2 | 2 |
| 4 | Braga | 3 | 0.040 | 0.560 | 0.055 | 2 | 2 |
| 5 | Chaves | 1 | 0.031 | 0.438 | 0.000 | 1 | 1 |
| 6 | Desportivo das Aves | 1 | 0.028 | 0.389 | 0.000 | 1 | 1 |
| 7 | Leixões | 1 | 0.028 | 0.389 | 0.000 | 1 | 1 |
| 8 | Marítimo | 1 | 0.031 | 0.438 | 0.000 | 1 | 1 |
| 9 | Paços de Ferreira | 1 | 0.031 | 0.438 | 0.000 | 1 | 1 |
| 10 | Porto | 9 | 0.053 | 0.737 | 0.687 | 2 | 2 |
| 11 | Rio Ave | 1 | 0.025 | 0.350 | 0.000 | 1 | 1 |
| 12 | Sporting CP | 6 | 0.043 | 0.609 | 0.505 | 2 | 2 |
| 13 | União de Leiria | 1 | 0.031 | 0.438 | 0.000 | 1 | 1 |
| 14 | Vitória de Guimarães | 2 | 0.033 | 0.467 | 0.000 | 2 | 2 |
| 15 | Vitória de Setúbal | 2 | 0.033 | 0.467 | 0.000 | 2 | 2 |

Neste caso, observamos que a decomposição da rede gera conchas com *coreness* ou de de 1 ou 2: todos aqueles nodos que integram o core com k mais elevado são aqueles cujo grau é igual ou superior a 2 - no caso concreto, estão aqui as equipas que jogaram pelo menos 2 finais da Taça de Portugal. Além disto, verifica-se que 6 dos 15 nodos integram o core com o k mais elevado e 9 em 15 fazem parte de uma concha com o k mínimo. Ou seja, somente 40% dos nodos formam a concha de *coreness* superior.

Rede 2 - Rede Aleatória

```
# 6. Decomposição de Core
```

```
(df_rede_aleatoria$K_Core_Decomposition <- round(coreness(graph2, mode="all"), 3))
```

```
## [1] 2 2 2 2 2 1 2 2 2 2 2 1 2 0 2
```

```
plot(graph2, vertex.size = 15+df_rede_aleatoria$K_Core_Decomposition*10,
      vertex.label = df_rede_aleatoria$K_Core_Decomposition)
```

```
kable(df_rede_aleatoria, caption = "Tabela 7 - rede aleatória - Nodos da rede com decomposição de core")
```

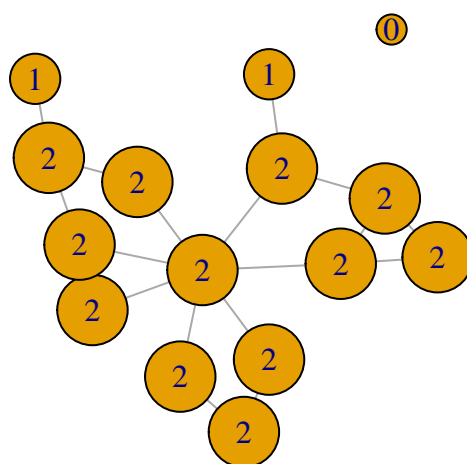


Figure 8: Figura 8 - rede aleatória assinalando decomposição de core

Table 7: Tabela 7 - rede aleatória - Nodos da rede com decomposição de core

| Nodo | Grau | Proximidade | Centralidade_de_Proximidade | Centralidade_de_Intermediação | K_Core_Decomposition |
|------|------|-------------|-----------------------------|-------------------------------|----------------------|
| 1 | 2 | 0.0227273 | 0.3181818 | 0.000 | 2 |
| 2 | 7 | 0.0285714 | 0.4000000 | 0.637 | 2 |
| 3 | 3 | 0.0200000 | 0.2800000 | 0.137 | 2 |
| 4 | 3 | 0.0200000 | 0.2800000 | 0.033 | 2 |
| 5 | 3 | 0.0238095 | 0.3333333 | 0.181 | 2 |
| 6 | 1 | 0.0161290 | 0.2258065 | 0.000 | 1 |
| 7 | 3 | 0.0232558 | 0.3255814 | 0.148 | 2 |
| 8 | 2 | 0.0181818 | 0.2545455 | 0.005 | 2 |
| 9 | 2 | 0.0192308 | 0.2692308 | 0.000 | 2 |
| 10 | 2 | 0.0222222 | 0.3111111 | 0.060 | 2 |
| 11 | 2 | 0.0232558 | 0.3255814 | 0.099 | 2 |
| 12 | 1 | 0.0185185 | 0.2592593 | 0.000 | 1 |
| 13 | 3 | 0.0238095 | 0.3333333 | 0.121 | 2 |
| 14 | 0 | 0.0047619 | 0.0666667 | 0.000 | 0 |
| 15 | 2 | 0.0222222 | 0.3111111 | 0.060 | 2 |

Nesta rede aleatória, verificamos que existem três níveis de *coreness*: 0, 1, 2. Isto porque sendo uma rede desconexa, o nodo isolado vai integrar somente a concha de $k = 0$. Aqui, como existem muitos nodos com grau igual ou superior a 2, significa que temos mais nodos a integrar a concha $k = 2$ que a proporção observada na rede da Taça de Portugal: na rede aleatória somente 3/15 nodos não integram a concha de k mais elevado, ou seja, temos 80% dos nodos a formarem o core de k superior. Esta análise permite-nos concluir que a centralidade pode ser considerada mais “difusa” nesta rede aleatória que na da Taça de Portugal, uma vez que a maior parte dos nodos fazem parte do core com grau mais elevado, e a medida de heterogeneidade é relativamente mais próxima de 1. Enquanto na primeira rede, temos nodos que se destacam na formação da concha de k superior, na segunda o menos comum é termos nodos a integrarem somente nodos de k inferiores.