



Licenciatura em Ciência de Dados - 3º ano

Trabalho 1

Análise de Redes

11 de dezembro de 2024

Discentes: João Dias nº 110305 / Felipe Pereira nº 110861 /
David Franco nº 110733 / Diogo Aqueu nº 110705

2024/2025

ÍNDICE

Questão 1	3
Questão 2	5
Questão 3	6
Questão 4	7
Questão 5	9
Questão 6	11
Apêndice	12
Apêndice I - Importação das bibliotecas necessárias	12
Apêndice II - Importação da rede através do ficheiro disponibilizado no Moodle	12
Apêndice III - Plot visual da rede	12
Apêndice IV - Cálculos de Coeficientes de Clustering	13
Apêndice V - Análise da Densidade e Classificação da Rede	13
Apêndice VI - Cálculo do Grau Médio	13
Apêndice VII - Distribuição e Resumo Estatístico do Grau dos Nodos	13
Apêndice VIII - Plot da Distribuição de Graus	14
Apêndice IX - Cálculo do Parâmetro de Heterogeneidade e Identificação de Hubs	14
Apêndice X - Cálculo da Associação de Grau (Degree Assortativity)	15
Apêndice XI - Cálculo da Distância Média (Média dos Comprimentos dos Caminhos Mais Curtos)	15
Apêndice XII - Cálculo do Logaritmo de Base 10 do Número de Nodos	15
Apêndice XIII - Cálculo dos Coeficientes de Clustering dos Nodos e da Rede	15
Apêndice XIV - Cálculo do Índice de Coreness e Determinação do Número de Conchas	16
Apêndice XV - Determinação da Dimensão de Cada Concha	16
Apêndice XVI - Distribuição Percentual de Nodos por Concha (Gráfico Circular)	17
Apêndice XVII - Visualização da Rede com Decomposição K-Core	18

Q1. Indique a dimensão e o número de ligações da rede. Determine a densidade e classifique a rede. Obtenha o grau médio e a distribuição de grau. Caracterize a distribuição de grau (mediana, quartis, etc.) e comente. Calcule o parâmetro de heterogeneidade. Indique o que pode concluir quanto à existência de hubs

A rede analisada, que pode ser visualizada na Fig. 1, representa contactos presenciais entre pessoas, onde os nodos correspondem a pessoas, e as ligações indicam a existência de contacto directo entre elas. A dimensão da rede, ou seja, o número de pessoas analisadas, é 51, enquanto o número de ligações identificadas é 155. Isso implica que foram estabelecidos 155 contactos presenciais entre essas pessoas.

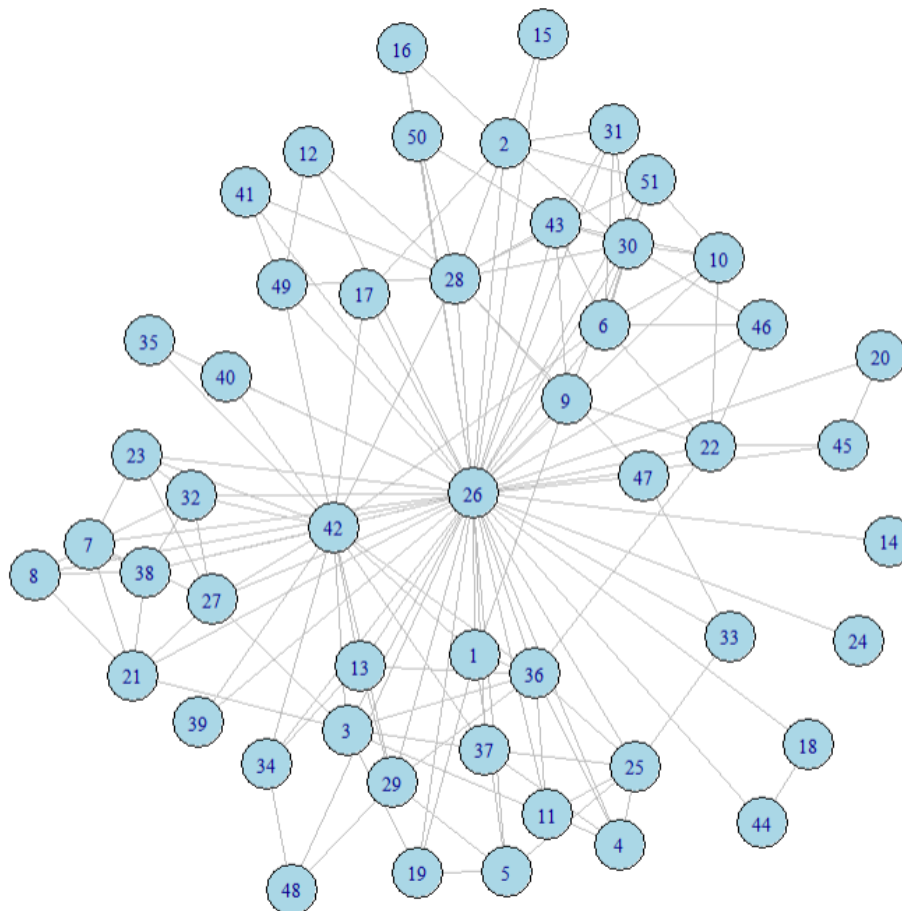


Fig. 1 - Grafo da rede

A densidade da rede é de 0,1216, o que caracteriza a rede como esparsa. Em termos práticos, nem todas as pessoas estão diretamente conectadas, o que é comum em redes sociais humanas, onde a interação é limitada a um subconjunto de indivíduos.

O grau médio, que reflete o número médio de contactos presenciais por pessoa, é 6,08. Isso significa que, em média, cada pessoa tem cerca de seis contactos diretos. Os valores do grau variam entre um mínimo de 1 contacto e um máximo de 50, com a mediana em 5 contactos, tal como se pode ver na Tabela 1. Os quartis mostram que 25% das pessoas têm 3 ou menos contactos estabelecidos, enquanto outros 25% têm mais de 6.5 contactos.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1,00	3,00	5,00	6,08	6,50	50,00

Tabela 1 - Distribuição de grau na rede

O gráfico da Fig. 2 mostra um histograma dos graus na rede. Observa-se uma distribuição assimétrica, com a maior parte dos nodos a apresentar um grau relativamente baixo, enquanto apenas alguns nodos apresentam graus muito elevados. A maior parte dos valores concentram-se nos valores antes do grau 10. No gráfico verifica-se um comportamento conhecido como "cauda longa" na distribuição de grau. A forma geral do gráfico sugere que a rede segue uma distribuição de grau heterogénea, com uma minoria de nodos muito conectados e uma maioria de nodos com "poucos" contatos (≤ 10).

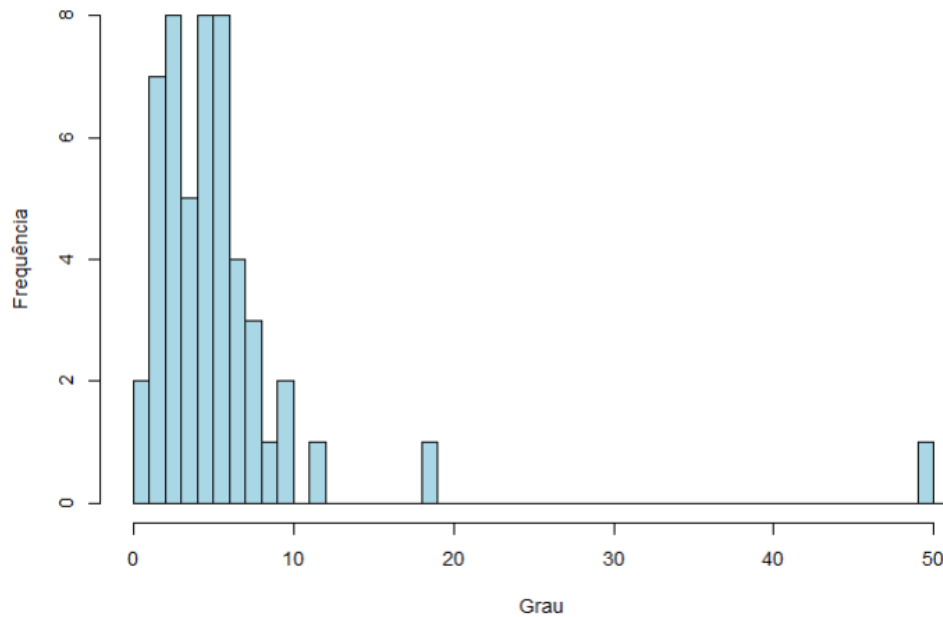


Fig. 2 - Distribuição de grau na rede

A existência de nodos com graus muito elevados (grau máximo de 50) reflete a presença de hubs, indivíduos altamente conectados que podem desempenhar um papel central na disseminação de informações ou na formação de subgrupos sociais. Note-se que se trata de uma rede ego. Essa observação é reforçada pelo cálculo do parâmetro de heterogeneidade, que resulta em 2,305. Este valor elevado confirma que a rede é heterogênea e que há uma distribuição desigual de conexões, com alguns indivíduos que acabam por concentrar grande parte dos contactos.

Portanto, a rede é caracterizada por uma estrutura esparsa, mas com hubs que desempenham um papel crucial no seu funcionamento e organização.

Q2. Estude a associação de grau e indique o que poderá concluir-se.

A Associação de Grau (Degree Assortativity) é um conceito que descreve a correlação entre os graus dos nodos que estão ligados entre si. Esta associação pode ser positiva, negativa ou neutra:

- Rede Associativa: Se o coeficiente de correlação de grau for significativo e positivo, significa que os nodos de maior grau estão mais frequentemente ligados a nodos de maior grau, formando uma rede com grupos de nodos de alta conectividade.
- Rede Não Associativa (Disassortative): Se o coeficiente for negativo, indica que os nodos de maior grau estão ligados a nodos de menor grau, criando uma rede onde os hubs se ligam a pessoas com menos contactos presenciais estabelecidos.
- Rede Neutra: Quando a correlação de grau é próxima de zero, a rede é neutra, sem uma tendência clara entre o grau dos nodos conectados.

No nosso caso, a associação de grau foi calculada e o valor do coeficiente obtido foi -0.288, o que apesar do baixo valor indica que a rede é não associativa. Em redes não associativas, os nodos de maior grau tendem a estar conectados a nodos de menor grau. Ou seja, pessoas com muitos contactos presenciais estabelecidos tendem a ligar-se com pessoas com menos contactos presenciais estabelecidos.

A polarização que pode surgir neste tipo de rede, onde pessoas com muitos contactos se ligam principalmente a indivíduos com menos, pode ter consequências negativas, como a criação de bolhas de informação, o que pode levar a desinformação ou manipulação de opinião.

Q3. Determine a média dos comprimentos dos caminhos mais curtos. Indique o que pode concluir-se quanto à distância média.

A análise da distância média dos caminhos mais curtos em redes de contactos humanos está intimamente ligada ao conceito de Small Worlds. Estas redes apresentam uma característica fundamental: embora possam existir muitos nodos (pessoas, no nosso caso) e ligações (contactos presenciais), a distância média entre quaisquer dois nodos é pequena, facilitando a propagação de informações ou recursos.

A nossa rede apresenta uma distância média de 1,878431, o que significa que, em média, são necessários aproximadamente 1,88 “passos”/contactos para que uma pessoa na rede possa estabelecer contacto com outra pessoa. Para determinar se esta distância média dos caminhos mais curtos é pequena ou grande, comparamos o valor obtido com o logaritmo de base 10 do número de nodos da nossa rede. Como a função de logaritmo cresce muito lentamente em relação a N , significa que, em redes de Small World, a distância média entre os nodos tende a ser próxima ao valor do logaritmo de N . O logaritmo de base 10 do número de nodos na nossa rede foi 1,70757, e a distância média obtida foi 1,878431, que está ligeiramente acima deste valor, mas ainda assim bastante próxima. Este comportamento sugere que a rede possui uma estrutura eficiente, com distâncias médias reduzidas.

Adicionalmente, uma rede com uma distância média baixa é característica de uma rede não associativa. Por outras palavras, nesta rede, os hubs (pessoas altamente conectadas, com muitos contactos presenciais estabelecidos) tendem a estar ligados a pessoas menos conectadas, o que contribui para a redução da distância média entre os nodos.

Q4. Determine os coeficientes de clustering dos nodos e da rede. Calcule a média dos coeficientes de clustering dos nodos. Compare esta média com o coeficiente de clustering da rede. Explique por que diferem estes dois valores. Diga o que pode concluir-se quanto à existência de triângulos.

Os coeficientes de clustering dos nodos e o coeficiente de clustering da rede fornecem informações importantes sobre a coesão e a estrutura de uma rede de contactos. O coeficiente de clustering local de um nodo avalia a densidade das conexões entre os seus vizinhos, ou seja, a probabilidade de que dois nodos que estão ligados a um terceiro nodo também estejam ligados entre si. O coeficiente de clustering global, por sua vez, refere-se à proporção de triângulos presentes na rede em comparação com o número total de triplets possíveis.

Na nossa análise, os coeficientes de clustering dos nodos foram calculados e obtivemos valores que variam de 0 até 1. A média dos coeficientes de clustering dos nodos foi de 0,683, o que indica que, em média, os vizinhos dos nós estão relativamente bem conectados, formando subgrupos locais com boa coesão. Já o coeficiente de clustering global da rede foi de 0,241, um valor significativamente mais baixo, o que sugere que, apesar da presença de subgrupos bem conectados localmente, a rede, como um todo, é mais dispersa, conforme já destacado anteriormente.

A diferença entre a média dos coeficientes de clustering dos nodos e o coeficiente de clustering global pode ser explicada pelo fato de que, enquanto o primeiro mede a coesão local, o segundo reflete a conectividade global da rede. A média dos coeficientes de clustering dos nodos é influenciada pela presença de comunidades ou subgrupos com alta densidade de conexões, mas o coeficiente global é atenuado pela quantidade de nodos isolados ou com poucas conexões, que não contribuem tanto para a formação de triângulos em toda a rede.

A existência de 162 triângulos na rede reforça a ideia de que há várias comunidades locais bem conectadas, onde os membros formam laços mútuos. Estes triângulos representam interações diretas entre três pessoas e são fundamentais para a coesão social, facilitando a circulação de informação e a construção de confiança entre os envolvidos.

A diferença entre a média dos coeficientes de clustering dos nodos e o coeficiente global da rede reflete a estrutura esparsa da rede em termos globais, mas com subgrupos bem conectados localmente. A presença de triângulos na rede indica que existem grupos de pessoas com laços fortes, o que contribui para a estabilidade e a resiliência da rede, promovendo uma maior troca de informações entre os seus membros.

Q5. Efetue a decomposição de core da rede. Indique o número de cores existentes e a dimensão de cada um. Indique o que poderá concluir-se.

A decomposição de core da rede foi realizada com o objetivo de identificar os diferentes níveis de conectividade entre os nodos, removendo recursivamente os nodos de menor grau, de forma a extrair/obter a parte mais densa da rede. Esse processo é útil para entender a estrutura interna da rede, identificando os nós centrais (hubs) e as áreas periféricas, mais dispersas. Após aplicar este algoritmo à rede, obtivemos os seguintes resultados:

A rede tem 5 cores (k-cores). A distribuição dos nós pode ver-se na Fig. 3 e é a seguinte:

- 2 nós no core 1;
- 8 nós no core 2;
- 10 nós no core 3;
- 7 nós no core 4;
- 24 nós no core 5.

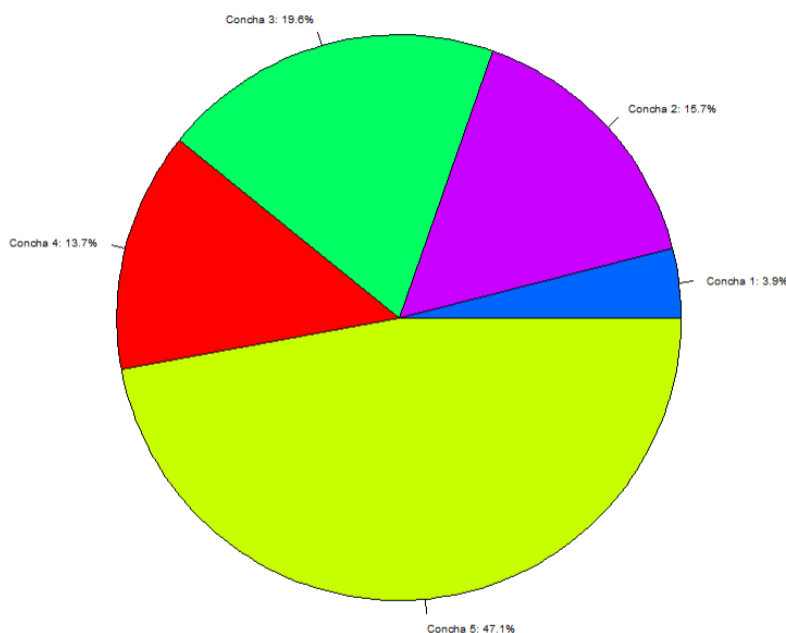


Fig. 3 - Distribuição percentual de nodos por concha

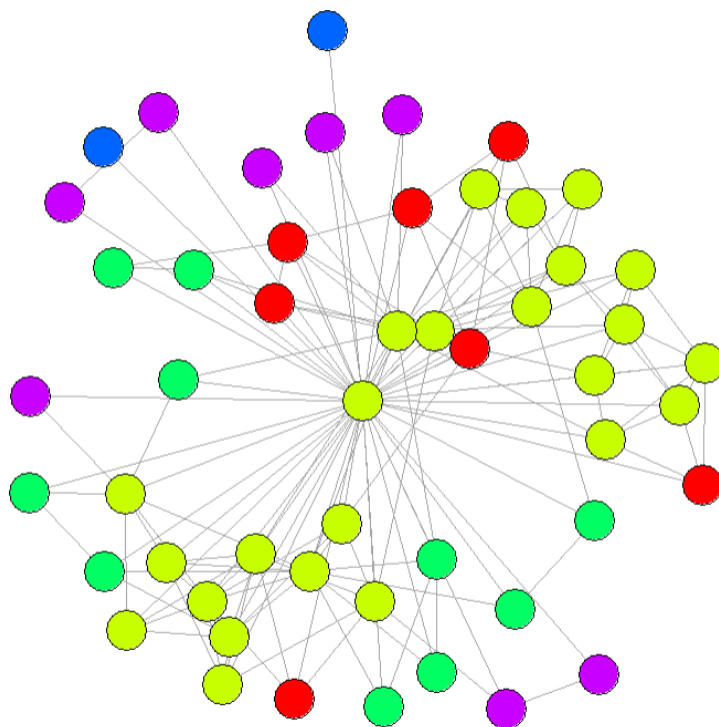


Fig. 4 - Visualização da rede com K-Core Decomposition

A maioria dos nodos encontra-se na concha 5 (24 nodos) representando quase 50% da rede. De seguida, as conchas 2, 3 e 4 representam percentagens semelhantes, o que indica que o número de nodos de cada concha são semelhantes entre si, restando apenas a concha 1 com o menor número de nodos.

Assim sendo, a maior parte da rede está concentrada em torno dos núcleos de maior grau (cores mais altos), onde a maioria das pessoas teve contacto presencial com um número relativamente grande de outros indivíduos. Por outro lado, as conchas externas (cores mais baixas) são compostas por nodos com menor grau, ou seja, pessoas com menos conexões diretas, que estão mais na periferia da rede.

Q6. Elabore um pequeno texto com comentários adicionais que considere pertinentes.

A análise desta rede trouxe algumas conclusões interessantes sobre como as pessoas se conectam. A presença de hubs, ou seja, indivíduos com muitos contactos, mostra que eles são importantes para manter a rede integrada e para facilitar a troca de informações entre os outros. Por outro lado, o facto de a rede ter um indicador negativo (ainda que baixo) na associação de grau (-0,288) indica que pessoas com muitos contactos tendem a ligar-se a outras com menos contactos, o que ajuda a evitar que grupos muito fechados se formem, mantendo a rede mais conectada no geral.

A distância média entre os nodos (1,88) também reflete uma rede bastante eficiente, já que é possível alcançar qualquer pessoa com poucos "passos". Isto está ligado ao conceito de Small World, onde as informações conseguem espalhar-se rapidamente, além disso, os valores de clustering mostram que existem subgrupos bem conectados na rede (com uma média de 0,683), mesmo que, globalmente, a rede seja mais esparsa.

A decomposição em cores também revelou que quase metade da rede está concentrada no núcleo mais denso (core 5), enquanto os nodos com menos conexões estão mais na periferia. Isto significa que a rede tem um centro forte e bem conectado, mas também dá espaço para quem tem menos contactos, o que a torna flexível e funcional.

No geral, a rede mostra um equilíbrio entre coesão e dispersão, o que reflete bem as dinâmicas sociais reais, onde há sempre pessoas mais conectadas a funcionar como "pontes" entre diferentes grupos.

Apêndice

Apêndice I - Importação das bibliotecas necessárias

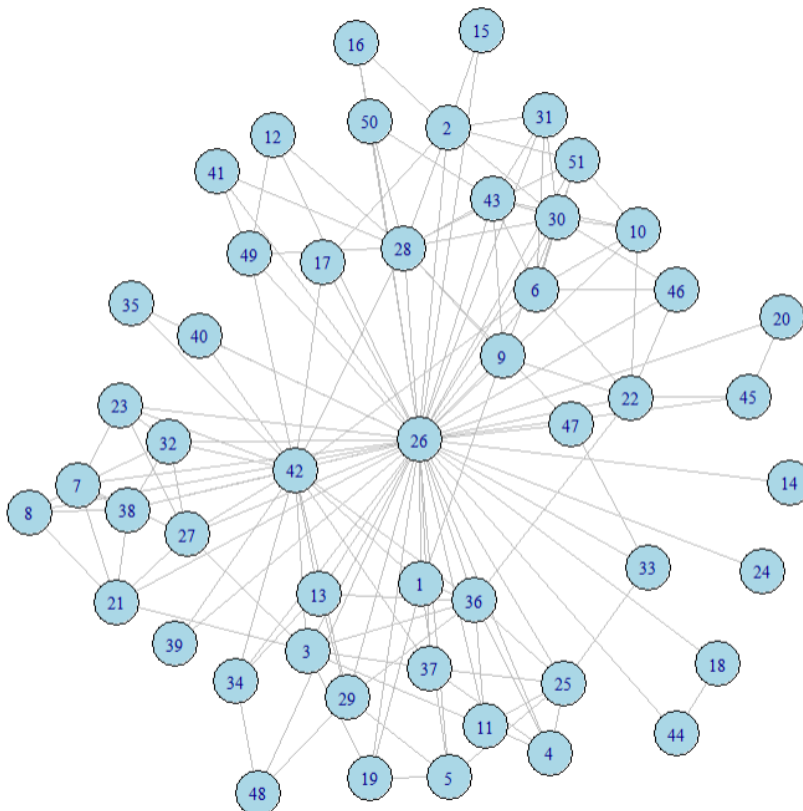
```
library(igraph)
```

Apêndice II - Importação da rede através do ficheiro disponibilizado no Moodle

```
rede <- read_graph("ligacoes_t1.txt", format = "edgelist", directed = F)
```

Apêndice III - Plot visual da rede

```
plot(  
  rede,  
  vertex.size = 12,          # Tamanho dos nodos  
  edge.color = "gray",       # Cor das arestas  
  vertex.color = "lightblue", # Cor dos nodos  
)
```



Apêndice IV - Cálculos de Coeficientes de Clustering

```
> ##### Q1
> # 1. Dimensão (número de nodos) e número de ligações
> dimensao <- vcount(rede)
> num_ligacoes <- ecount(rede)
> cat("Dimensão da rede (número de nodos):", dimensao, "\n")

Dimensão da rede (número de nodos): 51

> cat("Número de ligações:", num_ligacoes, "\n")

Número de ligações: 155
```

Apêndice V - Análise da Densidade e Classificação da Rede

```
> # 2. Densidade da rede
> densidade <- edge_density(rede)
> cat("Densidade da rede:", densidade, "\n")

Densidade da rede: 0.1215686

> # Classificação da rede
> if (densidade > 0.5) {
+   cat("Classificação: Rede Densa\n")
+ } else {
+   cat("Classificação: Rede Esparsa\n")
+ }
Classificação: Rede Esparsa
```

Apêndice VI - Cálculo do Grau Médio

```
> # 3. Grau médio
> graus <- degree(rede, mode = "total")
> grau_medio <- mean(graus)
> cat("Grau médio:", grau_medio, "\n")

Grau médio: 6.078431
```

Apêndice VII - Distribuição e Resumo Estatístico do Grau dos Nodos

```
> # 4. Distribuição do grau
> cat("Resumo estatístico do grau:\n")

Resumo estatístico do grau:
```

```

> summary(graus)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000  3.000  5.000  6.078  6.500  50.000
> quartis <- quantile(graus)
> cat("Quartis do grau:\n")
Quartis do grau:
> print(quartis)
  0%  25%  50%  75% 100%
  1.0  3.0  5.0  6.5 50.0

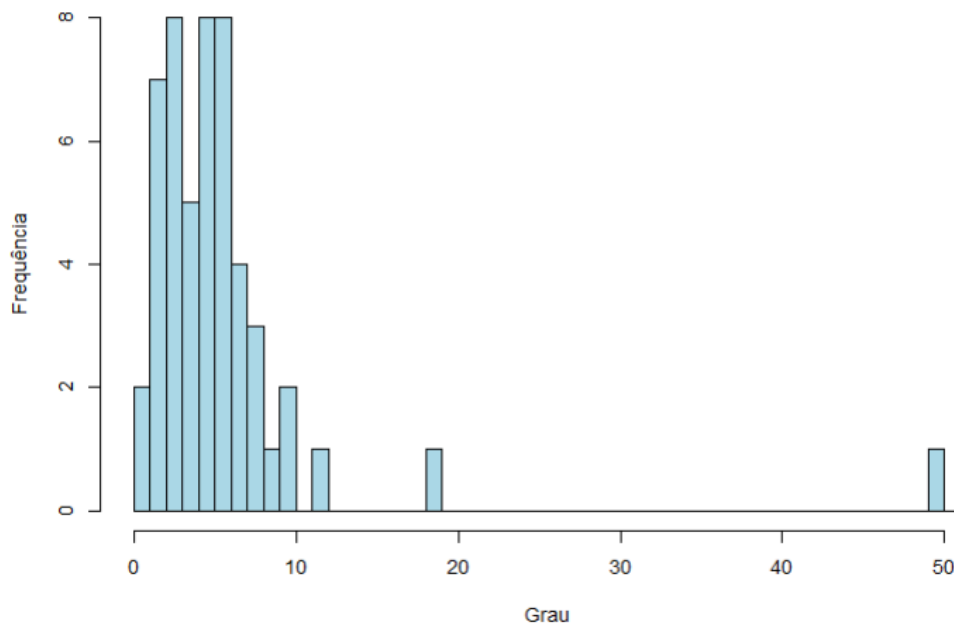
```

Apêndice VIII - Plot da Distribuição de Graus

```

> # Plot da distribuição de graus
> hist(graus, breaks = 0:vcount(rede), col = "lightblue", border = "black",
+      main = "Distribuição de Grau", xlab = "Grau", ylab = "Frequência")

```



Apêndice IX - Cálculo do Parâmetro de Heterogeneidade e Identificação de Hubs

```

> # 5. Parâmetro de heterogeneidade
> heterogeneidade <- mean(graus^2) / (grau_medio^2)
> cat("Parâmetro de heterogeneidade:", heterogeneidade, "\n")

Parâmetro de heterogeneidade: 2.305349

> # Identificar a existência de hubs

```

```
> if (heterogeneidade > 1.5) {
+   cat("A rede possui hubs (nodos altamente conectados).\n")
+ } else {
+   cat("A rede não apresenta hubs significativos.\n")
+ }
```

A rede possui hubs (nodos altamente conectados).

Apêndice X - Cálculo da Associação de Grau (Degree Assortativity)

```
> ##### Q2
> # Calcular o degree assortativity da rede
> assortatividade <- assortativity_degree(rede)
> cat("Degree Assortativity:", assortatividade, "\n")
```

Degree Assortativity: -0.2881953

Apêndice XI - Cálculo da Distância Média (Média dos Comprimentos dos Caminhos Mais Curtos)

```
> ##### Q3
> # Calcular a média dos comprimentos dos caminhos mais curtos (distância média)
> distancia_media <- mean_distance(rede)
> cat("Média dos comprimentos dos caminhos mais curtos (distância média):",
distancia_media, "\n")
```

Média dos comprimentos dos caminhos mais curtos (distância média): 1.878431

Apêndice XII - Cálculo do Logaritmo de Base 10 do Número de Nodos

```
> # Calcular o logaritmo de base 10 do número de nodos
> log_num_nodos <- log10(dimensao)
> cat("Logaritmo de base 10 do número de nodos:", log_num_nodos, "\n")
Logaritmo de base 10 do número de nodos: 1.70757
```

Apêndice XIII - Cálculo dos Coeficientes de Clustering dos Nodos e da Rede

```
> ##### Q4
> # Determinar os coeficientes de clustering dos nodos e da rede
> # Coeficiente de clustering dos nodos
> coef_clustering_nodos <- transitivity(rede, type = "local")
> cat("Coeficientes de clustering dos nodos:\n", coef_clustering_nodos)
```

```

Coeficientes de clustering dos nodos:
 0.4666667 0.3928571 0.4285714 0.8 0.5 0.5 0.6666667 1 0.5333333 0.7333333 0.8 1
0.8 NaN 1 0.6666667 0.6666667 1 0.6666667 1 0.6666667 0.3809524 0.9 NaN 0.4761905
0.08571429 0.6666667 0.3181818 0.5333333 0.5333333 0.8 0.8 0.6666667 0.6666667 1
0.4 0.6 0.6666667 1 1 1 0.1812865 0.5714286 1 0.6666667 0.8333333 0.6666667
0.6666667 0.7 0.6666667 0.7333333

> # Média dos coeficientes de clustering dos nodos
> media_clustering_nodos <- mean(coef_clustering_nodos, na.rm = TRUE)
> cat("\nMédia dos coeficientes de clustering dos nodos:", media_clustering_nodos,
"\n")

Média dos coeficientes de clustering dos nodos: 0.6830309

> # Coeficiente de clustering da rede (global)
> coef_clustering_rede <- transitivity(rede, type = "global")
> cat("Coeficiente de clustering da rede (global):", coef_clustering_rede, "\n")
Coeficiente de clustering da rede (global): 0.2409519

> # Número total de triângulos na rede
> num_triangulos <- count_triangles(rede)
> cat("\nNúmero total de triângulos na rede:", sum(num_triangulos) / 3, "\n") #
Dividimos por 3 porque cada triângulo é contado três vezes.

Número total de triângulos na rede: 162

```

Apêndice XIV - Cálculo do Índice de Coreness e Determinação do Número de Conchas

```

> ##### Q5
> # Calcular os índices de coreness para cada nodo
> coreness_values <- coreness(rede)
> # Determinar o número total de conchas (cores)
> num_conchas <- max(coreness_values)
> cat("Número total de conchas (cores):", num_conchas, "\n")
Número total de conchas (cores): 5

```

Apêndice XV - Determinação da Dimensão de Cada Concha

```

> # Determinar a dimensão de cada concha
> for (i in 1:num_conchas) {
+   dimensao_concha <- sum(coreness_values == i)
+   cat(paste("Número de nodos na concha", i, ":", dimensao_concha, "\n"))
}

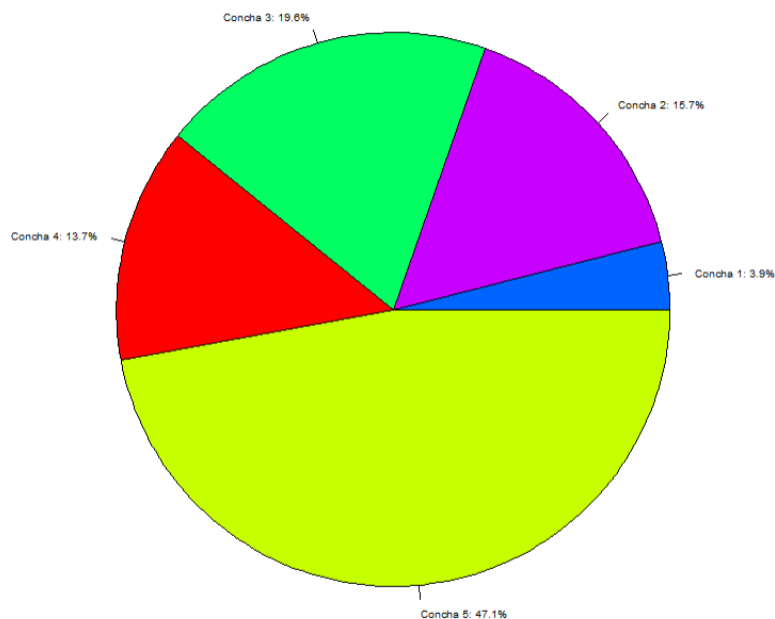
```



```
+ }
Número de nodos na concha 1 : 2
Número de nodos na concha 2 : 8
Número de nodos na concha 3 : 10
Número de nodos na concha 4 : 7
Número de nodos na concha 5 : 24
```

Apêndice XVI - Distribuição Percentual de Nodos por Concha (Gráfico Circular)

```
> # Definir cores únicas para cada concha (k-core)
> core_colors <- rainbow(length(unique(coreness_values)))
>
> # Mapear cada valor de coreness para a cor correspondente
> color_mapping <- setNames(core_colors, unique(coreness_values))
>
> # Gráfico circular da percentagem dos nodos por concha
> percentages <- table(coreness_values) / length(coreness_values) * 100
> pie(percentages,
+   main = "Distribuição Percentual de Nodos por Concha",
+   col = color_mapping[as.character(names(percentages))],
+   labels = paste0("Concha ", names(percentages), ": ", round(percentages, 1),
+ "%"),
+   cex = 0.7)
```



Apêndice XVII - Visualização da Rede com Decomposição K-Core

```
> # Visualização do grafo com k-core onde cada concha tem uma cor diferente
> V(rede)$color <- color_mapping[as.character(coreness_values)]
> plot(rede,
+       vertex.size = 12,
+       vertex.color = V(rede)$color,
+       main = "Visualização da Rede com K-Core Decomposition",
+       vertex.label = NA)
```

