

Redes não orientadas - Gestão de contactos sociais diretos entre habitantes de uma zona residencial

Trabalho de Grupo I realizado no âmbito da Unidade Curricular
de Análise de Redes do 3º ano da Licenciatura em Ciência de
Dados

Allan Kardec Rodrigues, 103380
aksrs@iscte-iul.pt

André Plancha, 105289
Andre_Plancha@iscte-iul.pt

Diogo Freitas, 104841
daafs@iscte-iul.pt

João Francisco Botas, 104782
Joao_Botas@iscte-iul.pt

Marco Esperança, 110451
mdeao@iscte-iul.pt

Índice

Introdução	2
Q1 : Análise da rede	3
Q2: Análise da componente gigante	7
Q3: Comparação entre a rede e a componente gigante	10
Conclusão	11
Apêndices	12
Rede	12
Apêndice A1 - Importação das bibliotecas necessárias	12
Apêndice B1 - Importação da rede através do ficheiro “trab_links.txt”	12
Apêndice C1 - Representação visual	12
Apêndice D1 - Dimensão e número de ligações	12
Apêndice E1 - Densidade	13
Apêndice F1 - Grau médio e distribuição de grau	13
Apêndice G1 - Verificação se a rede é conexa e, caso não o seja, indicação do número componentes e as dimensões mínima e máxima das componentes conexas	14
Apêndice H1 - Associação de grau	14
Apêndice J1 - Diâmetro da rede	14
Apêndice K1 - Média dos comprimentos dos caminhos mais curtos e averiguação se a distância média é pequena ...	15
Apêndice L1 - Estudo e caracterização de triângulos	15
Apêndice M1 - Cálculo do parâmetro de heterogeneidade	15
Apêndice N1 - Decomposição de <i>core</i> da rede e identificação do número de conchas (<i>shells</i>) existentes e a dimensão de cada uma	15
Componente Gigante	18
Apêndice A2 - Filtragem da componente gigante e representação visual	18
Apêndice B2 - Dimensão e número de ligações	18
Apêndice C2 - Densidade	18
Apêndice D2 - Grau médio e distribuição de grau	18
Apêndice E2 - Associação de grau	19
Apêndice F2 - Diâmetro da componente gigante	20
Apêndice G2 - Média dos comprimentos dos caminhos mais curtos e averiguação se a distância média é pequena ...	20
Apêndice H2 - Estudo e caracterização de triângulos	20
Apêndice I2 - Cálculo do parâmetro de heterogeneidade	20
Apêndice J2 - Decomposição de <i>core</i> da subrede e identificação do número de conchas (<i>shells</i>) existentes e a dimensão de cada uma	20

Introdução

O presente trabalho pretende analisar uma rede não orientada, que representa os contactos sociais diretos entre os habitantes de uma zona residencial. Os nodos (N) representam os habitantes da zona residencial e cada ligação (L) representa um contacto social direto entre dois habitantes. Na primeira questão pretende-se a análise de toda a rede e na segunda questão a componente gigante. Para isto, foi utilizado o *package* *igraph* da linguagem de programação R. Para importar esta rede para o R, utilizou-se a função `read_graph` da seguinte forma:

```
library(igraph) # importação da biblioteca
rede <- read_graph("trab_links.txt", format = c("edgelist"), directed=F)
plot(rede, vertex.size = 7, vertex.label.cex = .35) # com estes args para reduzir tamanho do grafo
```

A rede encontra-se representada na Figura 1.

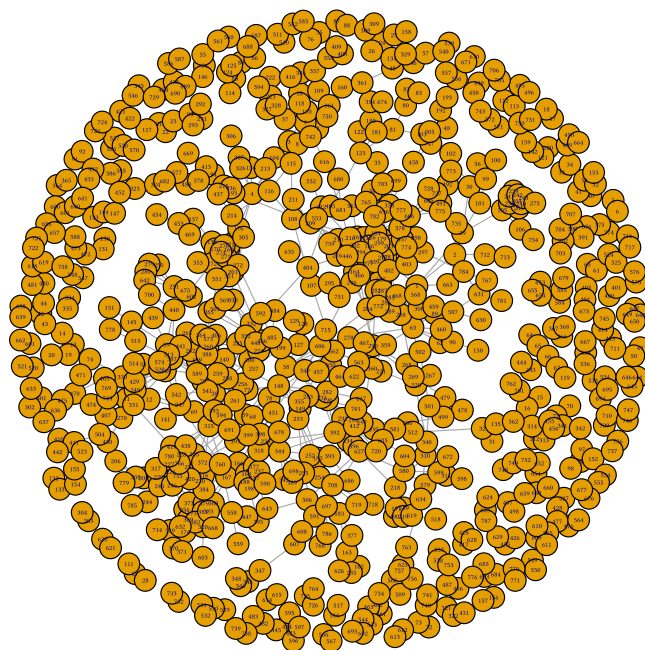


Figura 1: Representação visual da rede

A análise de uma rede pode ser útil para:

- **Identificação de Grupos e Comunidades:** Pelas componentes conectadas é possível descobrir que grupos de pessoas interagem mais frequentemente entre si, indicando possíveis comunidades na zona residencial.
- **Centralidade dos Indivíduos:** Através de algumas métricas de centralidade como o grau (número de conexões) ou a centralidade de informação (quão importante é um indivíduo no fluxo de informações) podem ajudar a identificar pessoas mais influentes ou centrais na rede social.
- **Propagação de Informação ou Ideia:** Estudar como uma informação ou ideia pode-se espalhar na comunidade através dos contactos sociais diretos.
- **Identificação de Pontos Críticos:** Encontrar os nós mais importantes na rede cuja remoção poderia interromper a propagação de informação ou de uma influência específica.
- **Monitorizar Mudanças:** Observar como a rede evolui ao longo do tempo, para oferecer *insights* sobre mudanças nas relações sociais e na estrutura da comunidade.

A segunda rede é a **componente gigante**, isto é, a maior componente conexa da rede original. Uma componente conexa numa rede é um conjunto de vértices onde cada par de vértices está conectado por um caminho. Assim, na componente gigante, cada vértice está acessível a partir de qualquer outro vértice dentro dessa componente através de um caminho.

A identificação da componente gigante é essencial para compreender a estrutura e conectividade de uma rede, destacando a sua parte mais extensa e predominante, revelando a sua coesão estrutural. Assim, no contexto deste trabalho, pode ser muito útil para representar uma comunidade social coesa, permitindo detetar os contactos sociais mais frequentes e fortes, permitindo evidenciar uma forte interação entre os habitantes.

A representação da componente gigante encontra-se na Figura 2.

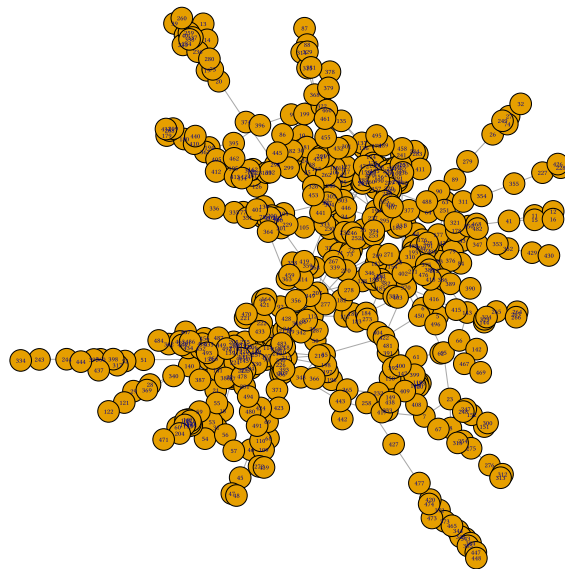


Figura 2: Representação visual da componente gigante

Q1 : Análise da rede

Na questão 1, primeiramente, pretende-se indicar a dimensão, e o número de ligações, assim como a densidade, o grau médio e a distribuição do grau.

O grau de um nodo i de uma rede não orientada é o número de ligações incidentes nesse nodo e representa-se por k_i . O grau médio é a média dos graus dos nodos de uma rede e representa-se por $\langle k \rangle$.

A densidade é uma medida relativa que relaciona o número de ligações existentes numa rede com o número máximo possível de ligações. Como a rede é não orientada, esta pode ser calculada por $d = \frac{2L}{N(N-1)}$.

Pela inspeção visual da representação gráfica da rede apresentada na Introdução pode-se constatar que é uma rede de dimensões consideráveis e não se conseguindo identificar os caminhos a olho nu. Isto pode ser corroborado pelo número de vértices da rede que são 787 e o número de arestas que são 1197.

Relativamente ao grau médio $\langle k \rangle = 3.041931$, significa que, em média, cada nodo está ligado a aproximadamente 3 nodos. No contexto do problema, significa que cada habitante na zona residencial está diretamente conectado a cerca de 3 outros habitantes através de contactos sociais diretos, representados na rede.

Quanto à densidade, obteve-se o valor de 0.003870142, um valor muito inferior a 1, pelo que a rede é esparsa. Desta forma, existe apenas uma pequena fração do número total de conexões possível entre os habitantes que está

realmente representada na rede. Além disso, pode indicar também a presença de subgrupos ou comunidades dentro da sua zona residencial que têm poucos contactos fora dos seus grupos, resultando nessa densidade reduzida na rede global.

Quanto à distribuição de graus, os resultados podem ser observados na Figura 3.

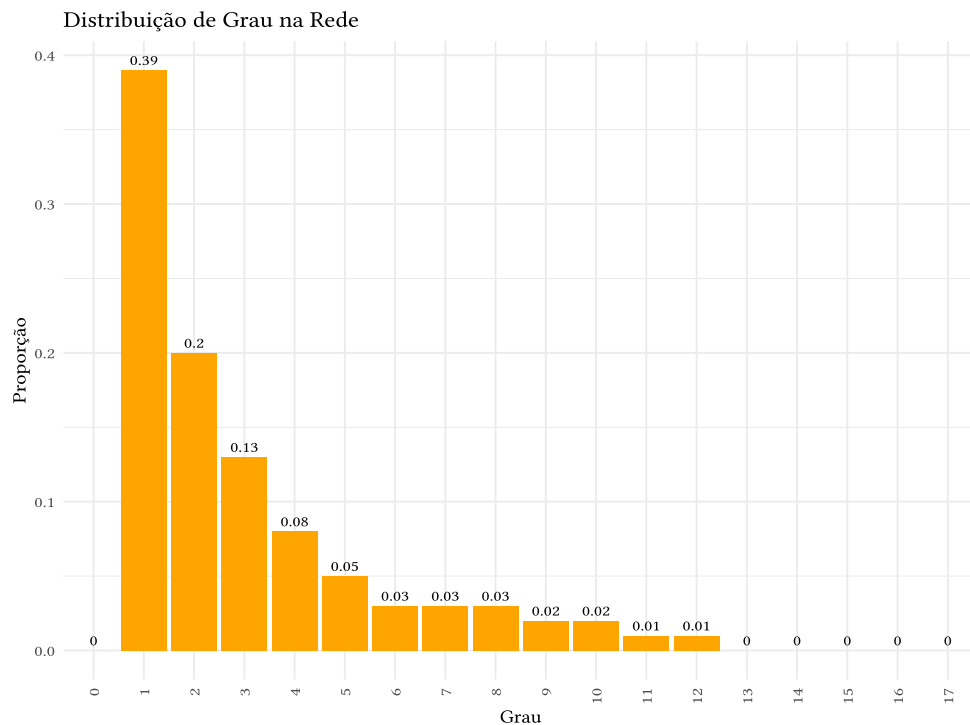


Figura 3: Gráficos de barras com a proporção da distribuição dos graus por cada nó da rede

Dela pode retirar-se que:

- **Grau 1:** Cerca de 39% dos habitantes têm apenas um contacto social direto na rede.
- **Grau 2:** Aproximadamente 20% dos habitantes têm dois contactos sociais diretos na rede.
- **Grau 3:** Cerca de 13% dos habitantes têm três contactos diretos sociais na rede.
- **Graus mais elevados:** À medida que o grau aumenta, a proporção de habitantes com esse grau diminui significativamente, indicando que é menos comum ter um número maior de contactos sociais diretos na rede.
- **Predominância de graus baixos:** A maior parte dos habitantes possui poucos contactos sociais diretos na rede, sugerindo uma rede onde a maioria dos habitantes possui um número limitado de conexões. Além disso, a concentração em graus mais baixos sugere a presença de grupos onde a interação é mais restrita.
- **Poucos Indivíduos Altamente Conectados:** A proporção diminui rapidamente à medida que o grau aumenta, indicando que há poucos indivíduos com muitos contatos diretos na rede.

De seguida, averiguar-se-á se a rede é conexa e, caso não seja, será indicado o número de componentes conexas e as dimensões mínima e máxima das componentes conexas.

Recorrendo à função `is.connected` podemos verificar que a rede não é conexa. Constatamos que a rede possui três componentes conectadas distintas, sendo que a menor componente possui 2 nodos, enquanto a maior possui 496 nodos. A variação considerável nos tamanhos das componentes, sugere que alguns grupos podem ser muito pequenos, enquanto outros podem representar uma parte significativa da população da zona residencial.

A associação de grau é 0.4765607, e, sendo positiva, há uma tendência para nodos com características semelhantes se conectarem, sugerindo que pessoas com características semelhantes têm uma chance maior de estabelecer contatos sociais diretos entre si. A média dos comprimentos dos caminhos mais curtos é 7.914034, o que significa que, em média, dois habitantes quaisquer na zona residencial estão conectados por um caminho de 7.914 unidades na rede de contatos sociais diretos. Uma distância média é pequena se cresce muito lentamente com o número de nodos da rede.

De forma a permitir avaliar se a distância média é pequena, recorremos a $\log_{10}(N)$, que é uma função que cresce muito lentamente e caso este último cálculo seja próximo da distância média então considera-se pequena.

Ora, temos que:

$$\begin{aligned} \text{distância média} &\approx 7.9 \\ \log_{10}(N) &\approx 2.9 \end{aligned} \tag{1}$$

Desta forma o valor da distância média é significativamente maior que $\log_{10}(N)$, pelo que a distância média é grande.

O diâmetro da rede é 21, o que indica que o caminho mais longo entre dois habitantes na rede social direta é composto por 21 conexões sociais diretas.

Numa rede social, podem observar-se diversos triângulos, isto é, conjuntos de três nodos em que existe uma ligação entre cada par de nodos.

Para estudar e caracterizar a existência de triângulos na rede, calculamos o coeficiente de *clustering* e o número de triângulos existentes na rede. O coeficiente de *clustering* é a fração de pares de nodos adjacentes desse nodo que estão ligados entre si. Obtivemos o valor de 0.4199126, o que indica a existência de subgrupos coesos na rede e, pelo seu valor considerável, a existência de muitos triângulos. Para calcular o número de triângulos existentes na rede utilizamos o seguinte código:

```
sum(count_triangles(rede))
```

A função `count_triangles` mostra o número de triângulos para cada nó da rede e, em seguida, utilizamos a função `sum` para somar todos esses valores, dando o número total de triângulos na rede, que são 2307.

Em diversas redes, alguns nodos têm bastante mais ligações do que os restantes, pelo que convém calcular a heterogeneidade, que é 1.837585. Como o valor é superior a 1, significa que alguns nós têm um número de conexões (grau) significativamente maior do que outros. Neste sentido, pode significar a presença de alguns indivíduos (*hubs*) que têm um número excecionalmente alto de conexões, em comparação com a maioria dos outros indivíduos na rede.

A decomposição de *k-cores* de uma rede identifica as conchas (*shells*) ou camadas hierárquicas, onde cada camada de concha é formada por nós que têm pelo menos *k* conexões naquela camada específica. Desta forma, permite estudar a rede em diferentes níveis de conectividade, tal como acontece no LinkedIn.¹

¹<https://www.linkedin.com/help/linkedin/answer/a545636/your-network-and-degrees-of-connection>

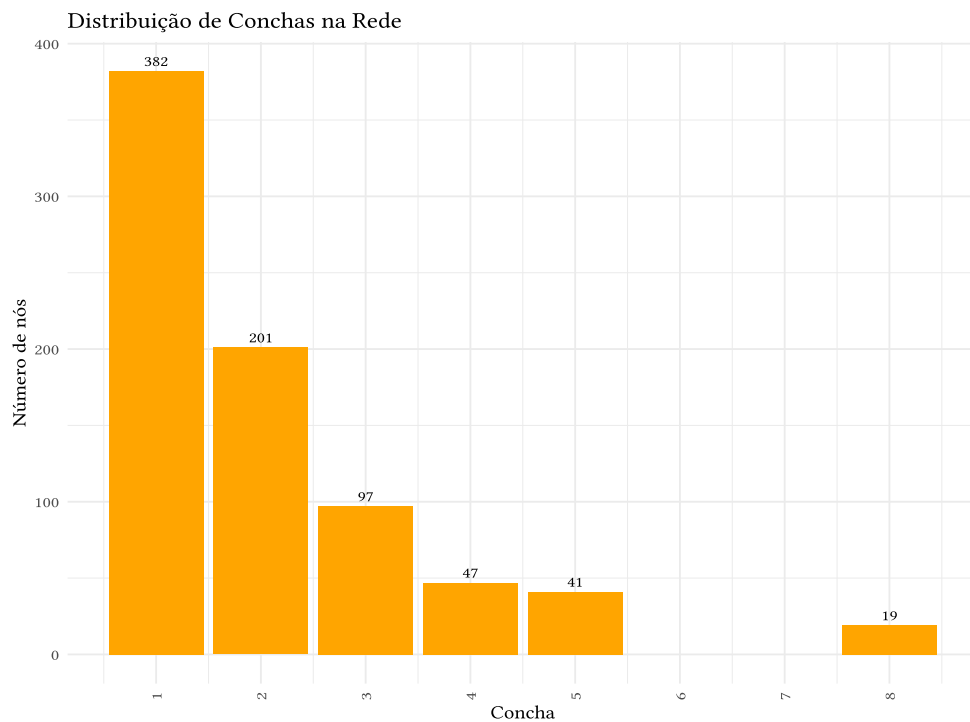


Figura 4: Gráficos de barras com a distribuição do número de nós por concha

Percentagem de nós em cada concha

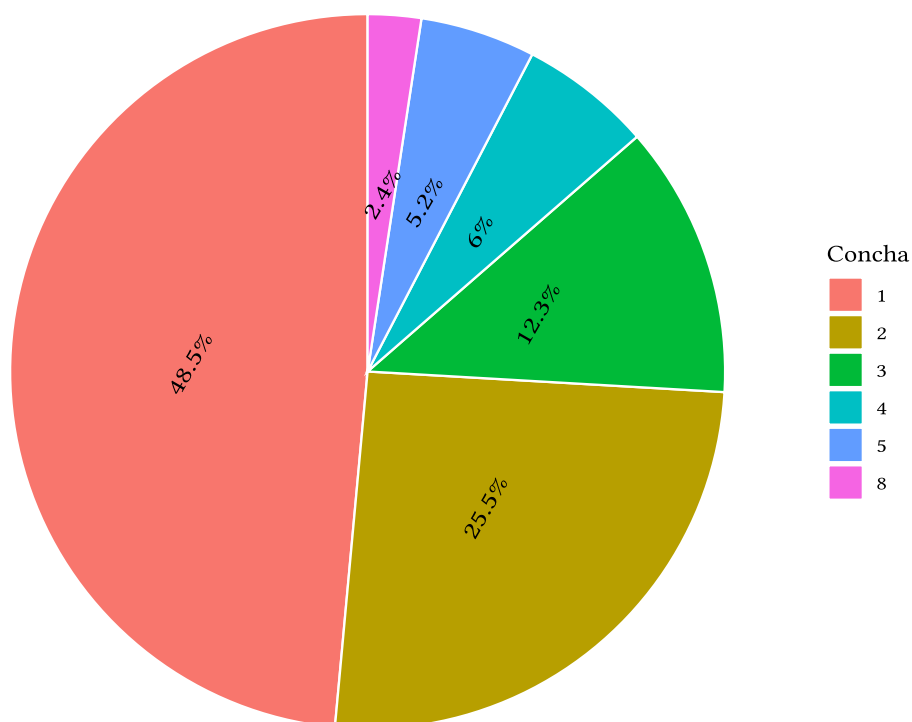


Figura 5: Gráfico circular com a percentagem do número de nós por concha

Ao observar o número de nós em cada concha percebemos que a maioria dos nós está nas conchas de menor número (1 e 2), que têm 382 e 281 nodos, respectivamente, representando quase 75% (concha 1 - 48.5% e concha 2 - 25.5%). As restantes conchas possuem percentagens menores, indicando uma diminuição gradual do número de nós conforme avançamos para conchas de números mais altos. Assim, a maior parte da rede está concentrada nas conchas iniciais indicando uma estrutura fortemente conectada e, possivelmente, alguns subgrupos menores mais afastados nas conchas mais altas.

As conchas 1 e 2 podem ser compostas por pessoas com muitas conexões ou muita influência na rede, representando, possivelmente, grupos de habitantes altamente conectados ou indivíduos chaves na comunicação e interação social da zona residencial considerada.

Q2: Análise da componente gigante

Nesta questão pretende-se fazer uma análise análoga à anterior, mas desta vez para a componente gigante.

Desta subrede resultou uma rede com 496 nodos e com 327 ligações.

O grau médio obtido, $\langle k \rangle$, é de 3.967742, o que significa que, em média, aproximadamente, cada nodo está ligado a outros quatro nodos. Neste problema, representa que cada habitante na zona residencial está diretamente conectado a cerca de 4 outros habitantes através de contactos sociais diretos, representados na componente gigante.

Quanto à densidade obtivemos o valor de 0.008, o que era espectável, tendo em conta que a componente gigante representa a componente com mais conexões e como tal será mais densa. Contudo, continua a ser inferior a 1 e muito próxima de 0, pelo que esta subrede é muito esparsa.

Quanto à distribuição de graus da componente gigante obtiveram-se os seguintes resultados:

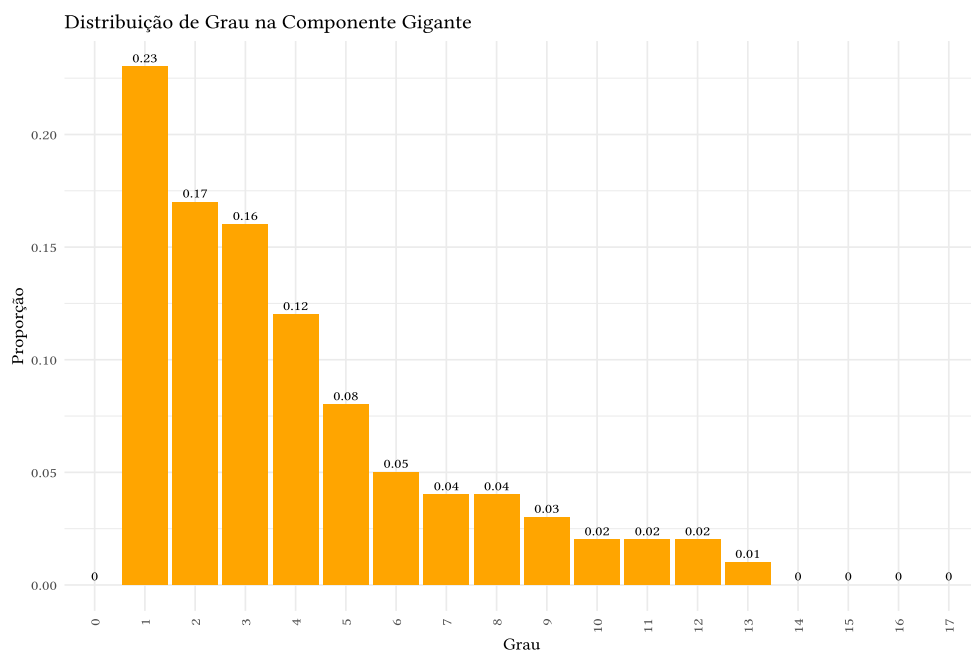


Figura 6: Gráficos de barras com a proporção da distribuição dos graus por cada nó da subrede

Do gráfico pode analisar-se que:

- **Grau 1:** Cerca de 23% dos habitantes têm apenas um contacto social direto na subrede.
- **Grau 2:** Aproximadamente 17% dos habitantes têm dois contactos sociais diretos na subrede.
- **Grau 3:** Cerca de 16% dos habitantes têm três contactos diretos sociais na subrede.
- **Grau 4:** Cerca de 12% dos habitantes têm três contactos diretos sociais na subrede.
- **Graus mais elevados:** À medida que o grau aumenta, a proporção de habitantes com esse grau diminui significativamente, indicando que é menos comum ter um número maior de contactos sociais diretos na componente gigante.
- **Predominância de número de contactos sociais até 4:** A maior parte das pessoas tem no máximo 4 contactos sociais dentro da sua área residencial, na componente gigante.
- **Poucos Indivíduos Altamente Conectados:** A proporção diminui rapidamente à medida que o grau aumenta, indicando que há poucos indivíduos com muitos contactos diretos na subrede.

Relativamente à associação de grau, esta é de 0.345145, e, sendo positiva e um valor um pouco baixo, existe alguma tendência para nodos com características semelhantes se conectarem, sugerindo que pessoas com características seme-

lhantes têm uma chance maior de estabelecer contatos sociais diretos entre si. A média de comprimento dos caminhos mais curtos é 7.933447, representando que, em média, dois habitantes quaisquer na zona residencial estão conectados por um caminho de 7.93 unidades na componente gigante.

Quanto à avaliação da distância média ser pequena, temos o seguinte:

$$\begin{aligned} \text{distância média} &= 7.933447 \\ \log_{10}(N) &\approx 2.7 \end{aligned} \quad (2)$$

Desta forma o valor da distância média é significativamente maior que $\log_{10}(N)$, pelo que a distância média é significativa.

O diâmetro da componente gigante é 21, o que indica que o caminho mais longo entre dois habitantes na subrede é composto por 21 contactos sociais diretos.

Quanto ao estudo e caracterização de triângulos nesta subrede, o coeficiente de *clustering* obtido foi de 0.419774, que indica a existência de subgrupos coesos nesta componente. Existem 2209 número de triângulos fechados existentes na componente gigante.

A heterogeneidade na componente gigante é de 1.612086. O valor é superior a 1, e, tal como explicado anteriormente, significa que alguns nós dentro da subrede tem um número de conexões (grau) significativamente maior do que outros. Assim, podem existir *hubs*, com algumas pessoas com muitos contactos sociais diretos, comparativamente a outras.

Relativamente à decomposição de *cores* na subrede obtivemos os seguintes resultados:

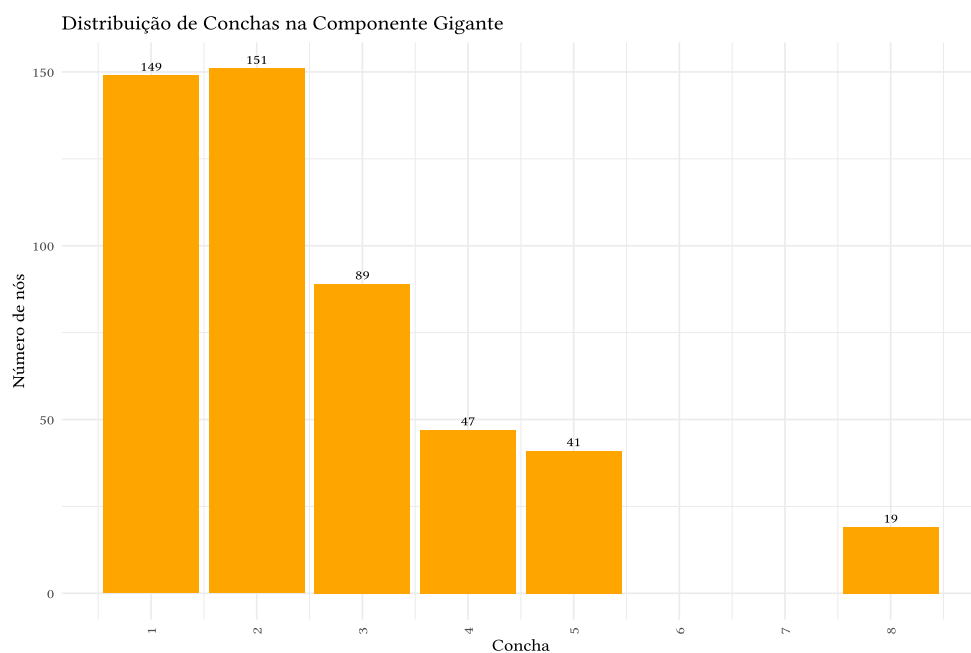


Figura 7: Gráficos de barras com a distribuição do número de nós por concha na componente gigante

Percentagem de nós em cada concha na Componente Gigante

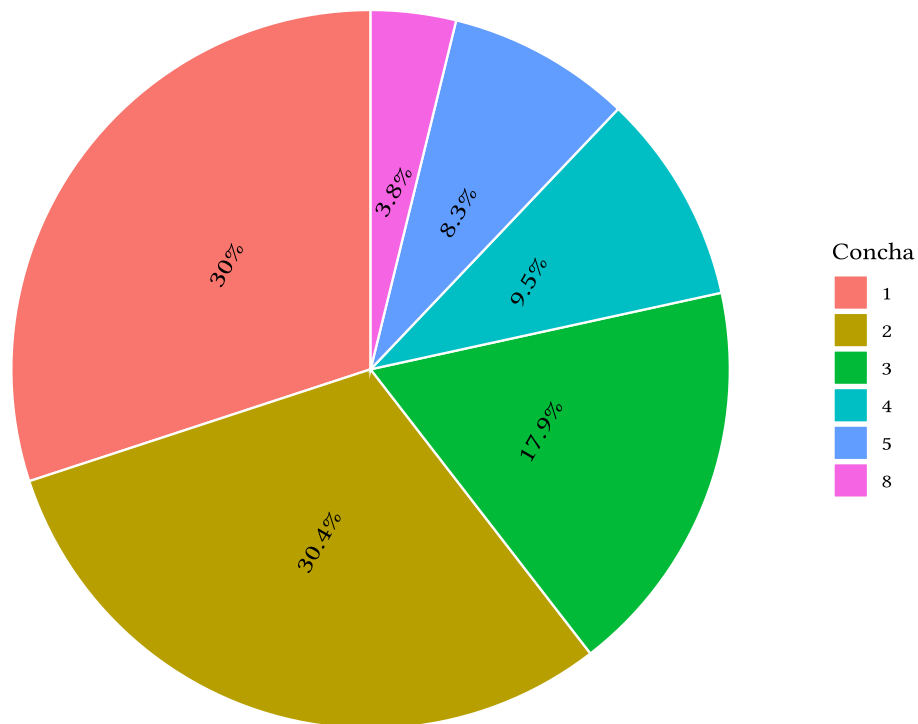


Figura 8: Gráfico circular com a percentagem do número de nós por concha na componente gigante

Ao observar o número de nós em cada concha, podemos constatar que a maioria dos nós está nas conchas de menor número (1 e 2), que têm praticamente o mesmo número de nodos em cada concha (149 e 151, respetivamente), representando cerca de 30% cada concha, pelo que estas duas conchas reúnem cerca de 60% do número de nós da componente gigante. As restantes conchas possuem percentagens menores, indicando uma diminuição gradual do número de nós conforme avançamos para conchas de números mais altos.

Q3: Comparação entre a rede e a componente gigante

Em seguida, apresenta-se uma tabela com os resultados obtidos para a rede e para a componente gigante, de forma à comparação ser mais fácil:

Tabela 1: Tabela com alguns dos resultados obtidos para a rede e para a componente gigante

Medidas	Rede	Componente Gigante
Nº nodos (N)	787	496
Nº arestas (L)	1197	984
Densidade (d)	0.0039	0.008
Grau médio ($\langle k \rangle$)	3.04	3.97
Associação / Coeficiente de Pearson (ρ)	0.48	0.35
Diâmetro	21	21
Distância média	7.91	7.33
$\log_{10}(N)$	2.9	2.7
Coeficiente de <i>clustering</i>	0.4199126	0.419774
Heterogeneidade	1.84	1.61
Nº conchas	8	8

Comparando o **tamanho** e a **densidade**, a rede original tem todos os nós e conexões, enquanto a componente gigante é a maior porção conectada na rede. Por esse motivo, a densidade da componente gigante é maior, já que consiste na parte mais conectada entre os habitantes nessa parte específica da rede.

Relativamente à **distribuição de graus**, a rede e a componente gigante possuem concentração em graus mais baixos.

Tanto para a rede como para a componente gigante, a **associação** é positiva e significativa, pelo que nodos com características semelhantes têm tendência a se conectarem, isto é, pessoas com características semelhantes têm uma chance maior de estabelecer contatos sociais diretos entre si.

Quanto ao **diâmetro** e à **distância média**, o diâmetro é o mesmo para ambas, contudo, a distância média é mais pequena na componente gigante, refletindo a maior coesão e conectividade nessa subrede. É de realçar que tanto para a rede como a subrede, a distância média não é pequena.

Relativamente à **análise e caracterização de triângulos**, ambos os valores do coeficiente de *clustering* são muito próximos, existindo, tal como mencionado anteriormente, 2307 triângulos na rede e 2209 triângulos na componente gigante. Desta forma, a componente gigante possui cerca de 96.6% dos triângulos existentes na rede.

Comparando a **decomposição de cores**, de forma a fornecer uma visão hierárquica da estrutura da rede e da subrede, a maior diferença está na concentração percentual dos nós entre as conchas 1 e 2. Na rede completa, a concha 1 tem uma percentagem significativamente maior (48.5%) comparativamente com a componente gigante (30.4%), que também tem uma percentagem idêntica de nodos na concha 2 (30%). Além disso, é de realçar que a concha 3 é mais proeminente na componente gigante com 17.9% em comparação com a rede completa com 12.3%.

Conclusão

Entender as diferenças entre a rede completa e a sua componente gigante revela detalhes importantes sobre como nos relacionamos. Enquanto a rede completa abarca todos os habitantes e as suas ligações, a componente gigante destaca-se como a parte onde estamos mais próximos e ligados. Nesse núcleo específico, sente-se uma proximidade maior entre as pessoas, refletida numa menor distância média entre todos nós. A presença significativa de triângulos nesse grupo reforça a formação de pequenos grupos unidos, que compõem a maior parte das estruturas sociais.

Essas diferenças mostram a importância crucial da componente gigante para compreender como nos relacionamos. A sua maior proximidade e coesão indicam uma comunidade mais unida, sendo este grupo fundamental na formação de pequenas comunidades dentro da comunidade maior. Esta análise dá-nos um vislumbre valioso sobre como nos conectamos, oferecendo *insights* preciosos sobre a ligação e dinâmica social entre todos nós numa comunidade.

Apêndices

Rede

Apêndice A1 - Importação das bibliotecas necessárias

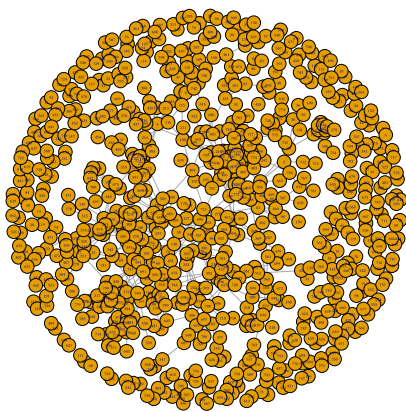
```
library(igraph) # para manipular as redes  
library(ggplot2) # para desenhar alguns gráficos
```

Apêndice B1 - Importação da rede através do ficheiro “trab_links.txt”

```
rede ← read_graph("trab_links.txt", format = c("edgelist"), directed=F)
```

Apêndice C1 - Representação visual

```
plot(rede, vertex.size = 7, vertex.label.cex = .35) # com estes args para reduzir tamanho do grafo
```



Apêndice D1 - Dimensão e número de ligações

```
# dimensao da rede  
  
# vertices  
vcount(rede)  
787  
  
# arestas  
ecount(rede)  
1197
```

Apêndice E1 - Densidade

```
# densidade
edge_density(rede, loops = FALSE)
0.003870142
```

Apêndice F1 - Grau médio e distribuição de grau

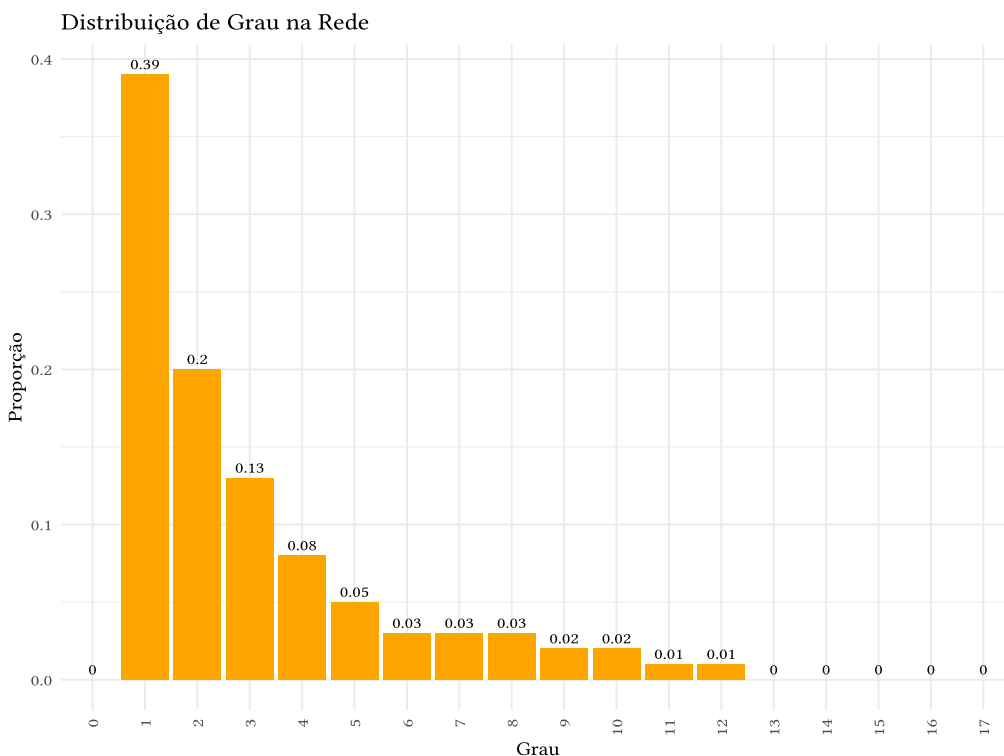
```
# grau medio
mean(degree(rede))
3.041931

# distribuicao de graus
round(degree_distribution(rede),2)
0.00 0.39 0.20 0.13 0.08 0.05 0.03 0.03 0.03 0.02 0.02 0.01 0.01 0.00 0.00 0.00 0.00 0.00

# Cria um data frame com os graus e suas proporções
(degree_df <- data.frame(Grau = 0:(length(degree_dist)-1), Proporcao = degree_dist))

# Cria o gráfico de barras com a distribuição de grau na rede
ggplot(degree_df, aes(x = factor(Grau), y = Proporcao)) +
  geom_bar(stat = "identity", fill = "orange") +
  geom_text(aes(label = Proporcao), vjust = -0.5, size = 3) +
  labs(x = "Grau", y = "Proporção", title = "Distribuição de Grau na Rede") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))

ggsave("Imagens/distribuicao_grau_rede.svg") # guarda o gráfico numa imagem svg
```



Apêndice G1 - Verificação se a rede é conexa e, caso não o seja, indicação do número componentes e as dimensões mínima e máxima das componentes conexas

```
# verificação se a rede é conexa
(is_connected <- is.connected(rede))
FALSE

# identificação do do número de componentes, tamanho mínimo e máximo da componente conexa
if (!is_connected) {

  num_components <- length(components(rede))
  print(paste("Número de componentes:", num_components))

  component_sizes <- sizes(components(rede))
  min_size <- min(component_sizes)
  max_size <- max(component_sizes)
  print(paste("Tamanho mínimo da componente conexa:", min_size))
  print(paste("Tamanho máximo da componente conexa:", max_size))
}

"Número de componentes: 3"
"Tamanho mínimo da componente conexa: 2"
"Tamanho máximo da componente conexa: 496"
```

Apêndice H1 - Associação de grau

```
# associação de grau
assortativity_degree(rede)
0.4765607
```

Apêndice J1 - Diâmetro da rede

```
diameter(rede)
21
```

Apêndice K1 - Média dos comprimentos dos caminhos mais curtos e averiguação se a distância média é pequena

```
# média dos caminhos mais curtos (distância média)
mean_distance(rede)
7.914034

# logaritmo para averiguar se a distância média é pequena
log10(vcount(rede))
2.895975
```

Apêndice L1 - Estudo e caracterização de triângulos

```
# coeficiente de clustering da rede
transitivity(rede, type = "global")
0.4199126

# calcula o número de triângulos fechados
(numero_triangulos_rede ← sum(count_triangles(rede)))
2307
```

Apêndice M1 - Cálculo do parâmetro de heterogeneidade

```
deg ← degree(rede, mode="all")
(ht ← mean(deg^2) / mean(deg)^2)
1.837585
```

Apêndice N1 - Decomposição de *core* da rede e identificação do número de conchas (*shells*) existentes e a dimensão de cada uma

```
# calcular a coreness para cada nó na rede
coreness_values ← coreness(rede)

# encontrar o número total de conchas (shells) na rede
(num_conchas ← max(coreness_values))
8

# contagem de nós em cada concha (shell)
for (i in 1:num_conchas) {
  dimensao_concha ← sum(coreness_values == i)
  print(paste("Número de nós na concha", i, ":", dimensao_concha))
}

"Número de nós na concha 1 : 382"
"Número de nós na concha 2 : 201"
"Número de nós na concha 3 : 97"
"Número de nós na concha 4 : 47"
"Número de nós na concha 5 : 41"
"Número de nós na concha 6 : 0"
```



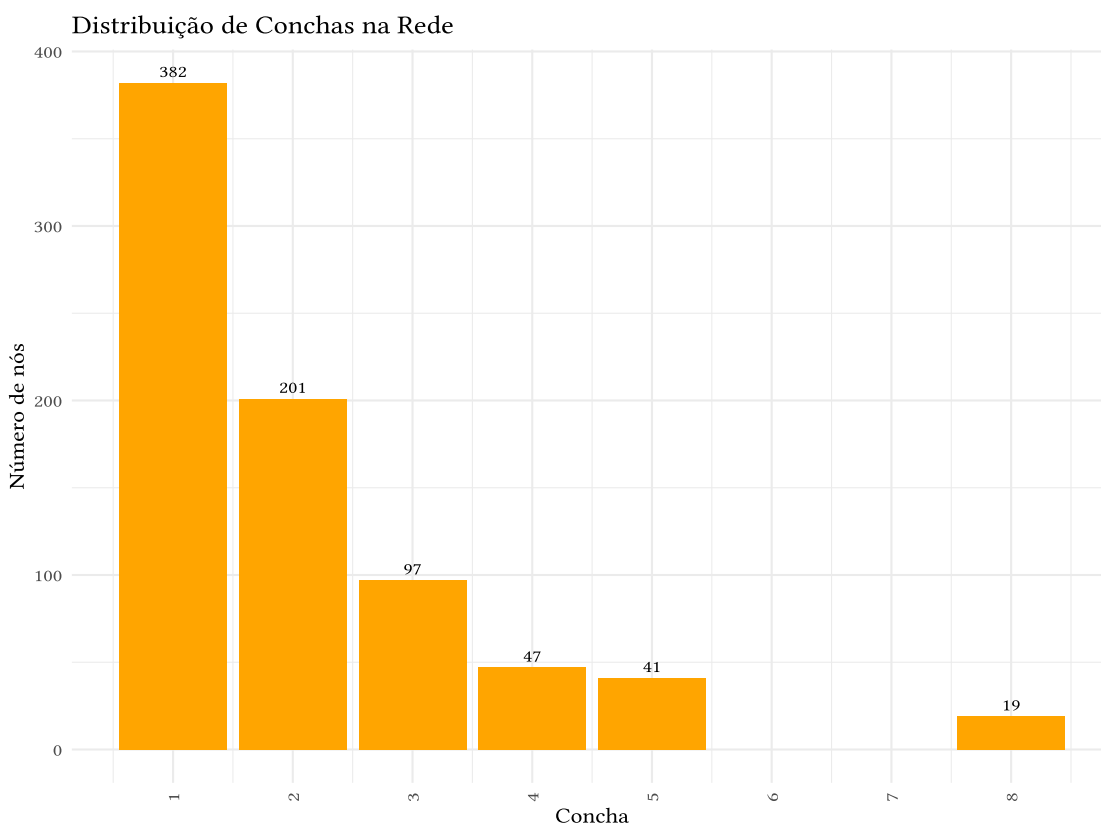
```

"Número de nós na concha 7 : 0"
"Número de nós na concha 8 : 19"

# gráfico de barras com a distribuição de conchas na rede

ggplot(data.frame(coreness_values), aes(x = coreness_values)) +
  geom_bar(stat = "count", fill = "orange") +
  geom_text(stat = "count", aes(label = ..count..), vjust = -0.5, size = 3) +
  labs(x = "Concha", y = "Número de nós", title = "Distribuição de Conchas na Rede") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1)) +
  scale_x_continuous(breaks = 1:num_conchas)
ggsave("Imagens/distribuicao_conchas_rede.svg") # guardar gráfico num svg

```



```

for (i in 1:num_conchas) {
  dimensao_concha <- sum(coreness_values == i)
  print(paste("Percentagem de nós na concha", i, ":", (dimensao_concha / vcount(rede)) * 100))
}

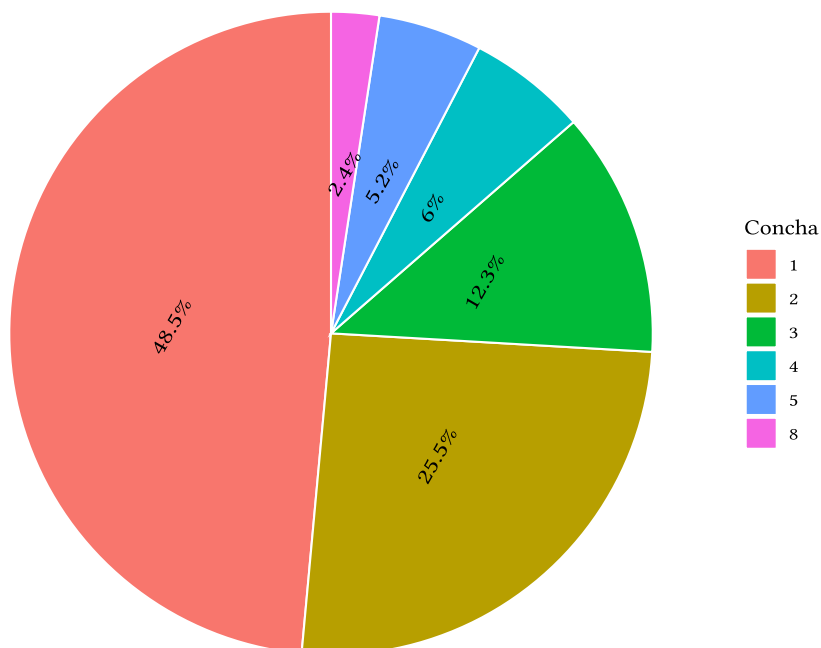
"Percentagem de nós na concha 1 : 48.5387547649301"
"Percentagem de nós na concha 2 : 25.5400254129606"
"Percentagem de nós na concha 3 : 12.3252858958069"
"Percentagem de nós na concha 4 : 5.9720457433291"
"Percentagem de nós na concha 5 : 5.20965692503177"
"Percentagem de nós na concha 6 : 0"
"Percentagem de nós na concha 7 : 0"

```

```
"Percentagem de nós na concha 8 : 2.41423125794155"
```

```
data <- data.frame(  
  concha = 1:num_conchas,  
  percentagem = NA  
)  
  
for (i in 1:num_conchas) {  
  data$percentagem[i] <- round((sum(coreness_values == i) / vcount(rede)) * 100, 2)  
}  
  
# excluir as conchas com 0% de nós  
data_filtered <- subset(data, percentagem != 0)  
  
# gráfico circular com a percentagem de nós em cada concha  
ggplot(data = data_filtered, aes(x = "", y = percentagem, fill = factor(concha))) +  
  geom_bar(stat = "identity", width = 1, color = "white") +  
  coord_polar("y", start = 0) +  
  labs(title = "Percentagem de nós em cada concha", fill = "Concha") +  
  theme_void() +  
  theme(legend.position = "right") +  
  geom_text(aes(label = paste0(round(percentagem,1), "%")), position = position_stack(vjust = 0.5),  
angle = 60)  
ggsave("Imagens/percentagem_conchas_rede.svg") # guardar gráfico num svg
```

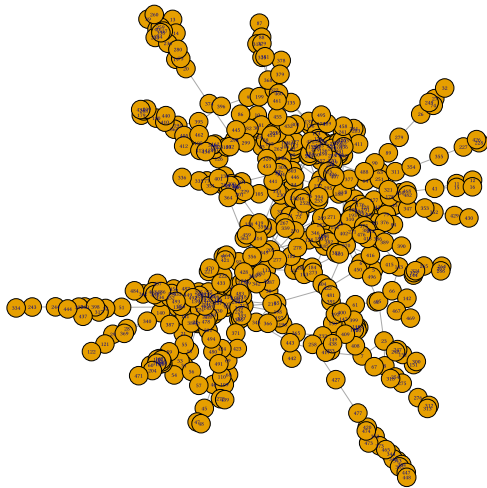
Percentagem de nós em cada concha



Componente Gigante

Apêndice A2 - Filtragem da componente gigante e representação visual

```
componentes <- components(rede)
maior_componente <- which.max(componentes$size) # indice da componente mais conectada
(componente_gigante <- induced.subgraph(rede, which(componentes$membership == maior_componente)))
plot(componente_gigante, vertex.size = 7, vertex.label.cex = .35) # representação visual
```



Apêndice B2 - Dimensão e número de ligações

```
# dimensao da componente gigante
(dimensao_componente <- vcount(componente_gigante))
496

# numero de ligacoes da componente gigante
(ligacoes_componente <- ecoun(componente_gigante))
984
```

Apêndice C2 - Densidade

```
# densidade da componente gigante
edge_density(componente_gigante, loops = FALSE)
0.00801564
```

Apêndice D2 - Grau médio e distribuição de grau

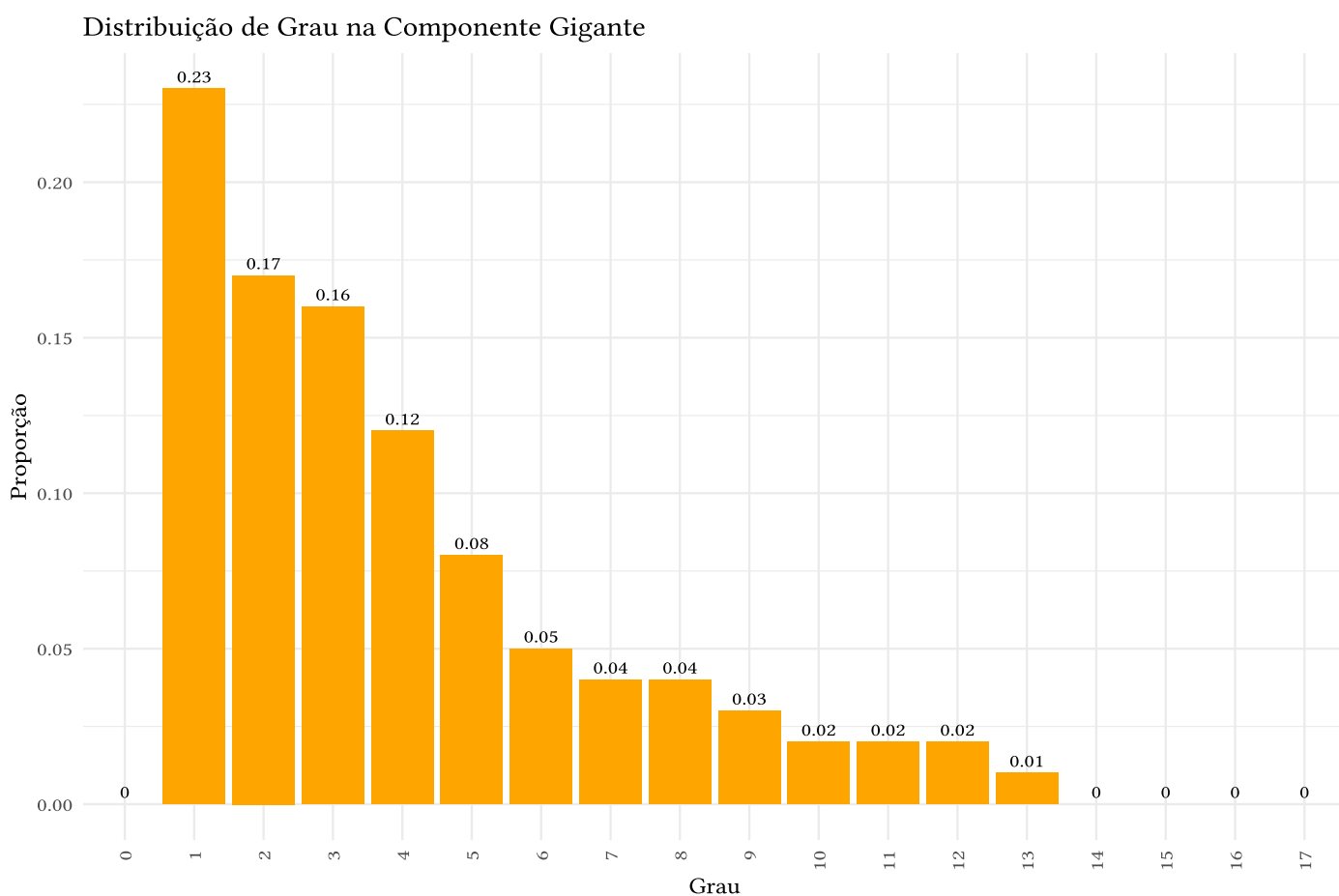
```
# grau medio da componente gigante
mean(degree(componente_gigante))
3.967742

# distribuicao de graus da componente gigante
(degree_dist_cg <- round(degree_distribution(componente_gigante), 2))
0.00 0.23 0.17 0.16 0.12 0.08 0.05 0.04 0.04 0.03 0.02 0.02 0.02 0.01 0.00 0.00 0.00 0.00
```

```
# cria um data frame com os graus e suas proporções
(degree_df_cg ← data.frame(Grau = 0:(length(degree_dist_cg)-1), Proporcao = degree_dist_cg))

# gráfico de barras com a distribuição de graus na componente gigante
ggplot(degree_df_cg, aes(x = factor(Grau), y = Proporcao)) +
  geom_bar(stat = "identity", fill = "orange") +
  geom_text(aes(label = Proporcao), vjust = -0.5, size = 3) +
  labs(x = "Grau", y = "Proporção", title = "Distribuição de Grau na Componente Gigante") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))

ggsave("Imagens/distribuicao_grau_componente_gigante.svg") # guardar o gráfico num svg
```



Apêndice E2 - Associação de grau

```
# associação de grau da componente gigante
assortativity_degree(componente_gigante)
0.345145
```

Apêndice F2 - Diâmetro da componente gigante

```
# diametro da componente gigante
diameter(componente_gigante)
21
```

Apêndice G2 - Média dos comprimentos dos caminhos mais curtos e averiguação se a distância média é pequena

```
# media caminhos mais curtos da componente gigante
mean_distance(componente_gigante)
7.933447

# logaritmo para averiguar se a distância média é pequena
log10(vcount(componente_gigante))
2.695482
```

Apêndice H2 - Estudo e caracterização de triângulos

```
# coeficiente de clustering da rede
transitivity(componente_gigante, type = "global")
0.419774

# calcula o número de triângulos fechados
(numero_triangulos_cg ← sum(count_triangles(componente_gigante)))
2229

# calcula a percentagem do número de triângulos da componente gigante
# em relação ao número de triângulos da rede
(percentagem_triangulos_cg ← round((numero_triangulos_cg / numero_triangulos_rede) * 100, 1))
96.6
```

Apêndice I2 - Cálculo do parâmetro de heterogeneidade

```
deg_cg ← degree(componente_gigante, mode="all")

(ht_cg ← mean(deg_cg^2) / mean(deg_cg)^2)
1.612086
```

Apêndice J2 - Decomposição de *core* da subrede e identificação do número de conchas (*shells*) existentes e a dimensão de cada uma

```
# calcular a coreness para cada nó na rede
coreness_values ← coreness(componente_gigante)

# encontrar o número total de conchas (shells) na rede
(num_conchas ← max(coreness_values))

# contagem de nós em cada concha (shell)
for (i in 1:num_conchas) {
```

```

dimensao_concha <- sum(coreness_values == i)
print(paste("Número de nós na concha", i, ":", dimensao_concha))
}

```

```

"Número de nós na concha 1 : 382"
"Número de nós na concha 2 : 201"
"Número de nós na concha 3 : 97"
"Número de nós na concha 4 : 47"
"Número de nós na concha 5 : 41"
"Número de nós na concha 6 : 0"
"Número de nós na concha 7 : 0"
"Número de nós na concha 8 : 19"

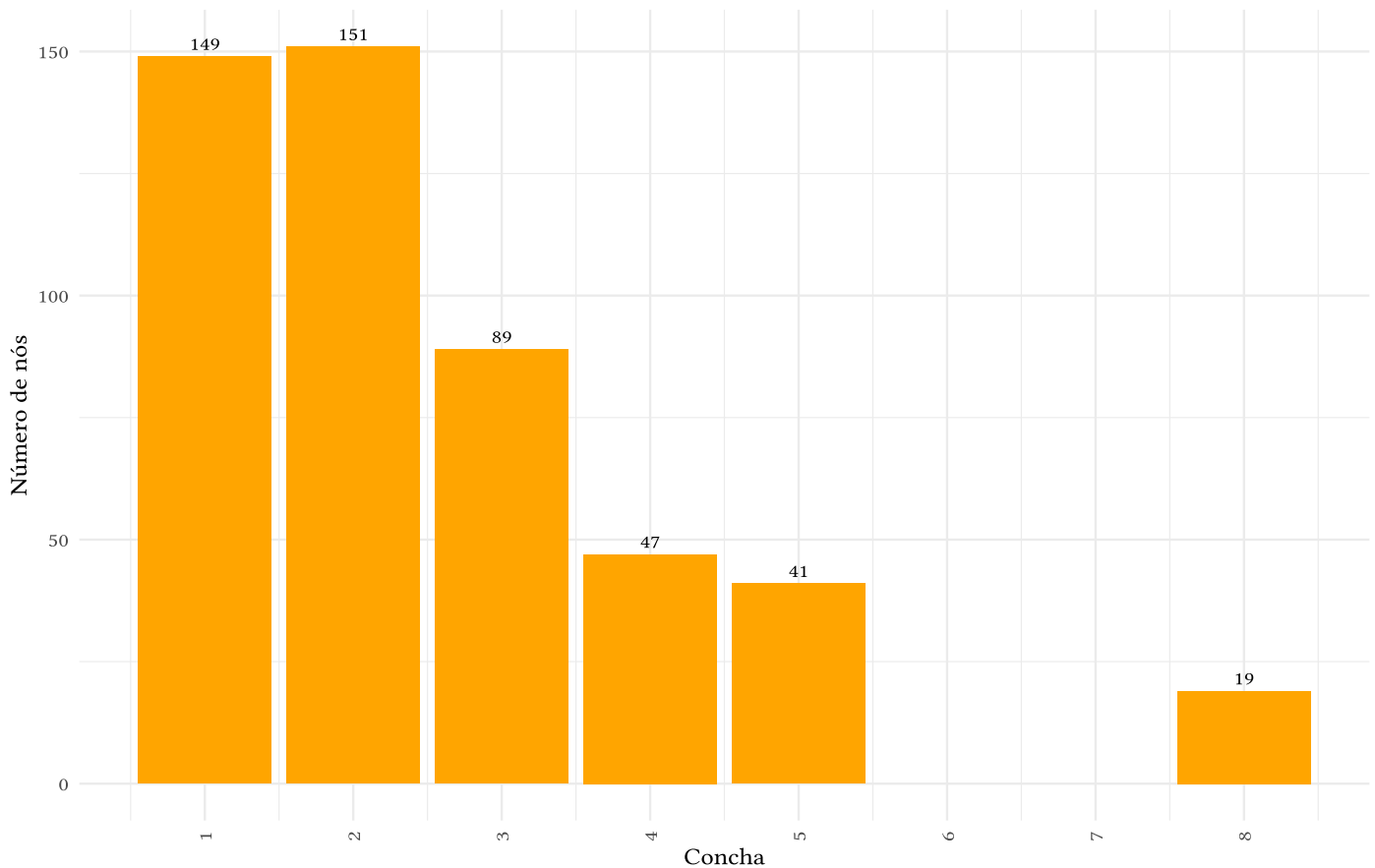
```

```

ggplot(data.frame(coreness_values), aes(x = coreness_values)) +
  geom_bar(stat = "count", fill = "orange") +
  geom_text(stat = "count", aes(label = after_stat(count)), vjust = -0.5, size = 3) +
  labs(x = "Concha", y = "Número de nós", title = "Distribuição de Conchas na Componente Gigante")
+
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1)) +
  scale_x_continuous(breaks = 1:num_conchas)
ggsave("Imagens/distribuicao_conchas_componente_gigante.svg")

```

Distribuição de Conchas na Componente Gigante



```

data_cg <- data.frame(
  concha = 1:num_conchas,
  percentagem = NA
)

for (i in 1:num_conchas) {
  data_cg$percentagem[i] <- round((sum(coreness_values == i) / vcount(componente_gigante)) * 100,
2)
}

# excluir as conchas com 0% de nós
data_filtered_cg <- subset(data_cg, percentagem != 0)

# gráfico circular com a percentagem de nós em cada concha
ggplot(data = data_filtered_cg, aes(x = "", y = percentagem, fill = factor(concha))) +
  geom_bar(stat = "identity", width = 1, color = "white") +
  coord_polar("y", start = 0) +
  labs(title = "Percentagem de nós em cada concha na Componente Gigante", fill = "Concha") +
  theme_void() +
  theme(legend.position = "right") +
  geom_text(aes(label = paste0(round(percentagem,1), "%")), position = position_stack(vjust = 0.5),
angle = 60)
ggsave("Imagens/percentagem_conchas_componente_gigante.svg")

```

Percentagem de nós em cada concha na Componente Gigante

