



Análise de Redes - Trabalho de Grupo n.º 2

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

Catarina Castanheira, 92478

João Martins, 93259

Joel Paula, 93392

03/01/2022

QUESTÃO 1:

Suponha que pretende gerar uma rede aleatória não orientada com 100 nodos e grau médio aproximadamente igual a 4. Qual deve ser a probabilidade utilizada na geração da rede? Gere esta rede.

O grau médio seria de aproximadamente 100, se todos os nodos estivessem ligados entre si. Neste caso teremos uma probabilidade de 4% - que cada nodo esteja ligado apenas a 4 dos 100 nodos.

É possível determinar o grau médio de uma rede conhecendo o número de nodos e a probabilidade de geração utilizada:

$$\langle k \rangle = p(N - 1)$$

Isto significa que, resolvendo a equação em função de p , temos:

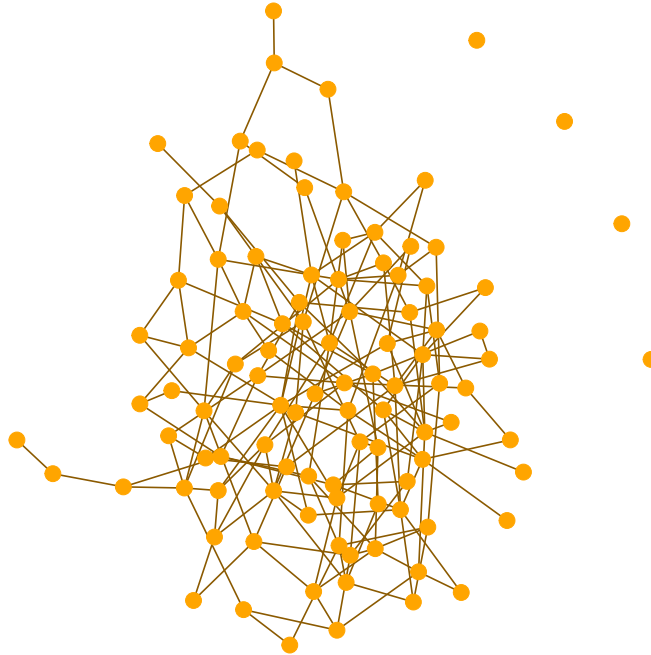
$$p = \langle k \rangle / (N - 1)$$

No nosso caso concreto, o objectivo é obtermos $\langle k \rangle = 4$, tendo $N = 100$. Logo, a probabilidade que deverá ser usada na geração da rede aleatória é:

$$p = 4 / (100 - 1) \Leftrightarrow p = 4 / 99 = 0.04 \text{ (aprox.)}$$

Ou seja, para obtermos uma rede aleatória de 100 nodos e um grau médio de 4, deveremos usar uma probabilidade de 4% na geração das suas ligações.

```
set.seed(42)
graph1 <- sample_gnp(100, 0.04)
par(mar=c(0,0,0,0))
plot(graph1, edge.color="orange4",
      , edge.width=2
      , vertex.label=NA
      , vertex.size=5
      , vertex.label.cex=0.8
      , vertex.frame.color="orange"
      , vertex.color = "orange"
      )
```



Caracterize esta rede quanto ao grau médio dos nodos, à conectividade, distância média e existência de triângulos. Aplique ainda métodos de identificação de comunidades.

grau médio

```
# grau médio
mean(degree(graph1))
```

```
## [1] 3.76
```

O grau médio é próximo de 4, tal como esperaríamos.

Conectividade

```
degree(graph1)
```

```
## [1] 2 5 2 4 4 5 3 5 5 4 5 6 6 3 9 5 2 1 4 5 2 6 4 1 1 5 0 2 3 4 2 4 3 4 7 6 1
## [38] 1 0 0 2 4 5 0 2 3 5 4 2 4 8 3 3 4 3 2 5 3 6 8 6 3 4 2 5 7 4 3 3 2 4 7 3 2
## [75] 3 6 3 4 2 9 2 5 4 5 5 5 2 3 1 3 5 4 3 5 4 3 5 4 5 4
```

Vemos que existem 4 nodos com grau zero, o que indica que não têm qualquer ligação. Neste caso estamos perante uma rede desconexa.

Portanto, existem 4 nodos e uma componente gigante.

```
components(graph1)
```

```
## $membership
## [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1
## [38] 1 3 4 1 1 1 5 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [75] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##
## $csize
## [1] 96 1 1 1 1
##
## $no
## [1] 5
```

Associação

```
# Associação de grau
assortativity_degree(graph1)
```

```
## [1] -0.08082738
```

É negativo mas muito próximo de zero, por isso não é um rede associativa mas também não se pode concluir que é Não associativa.

Vamos olhar agora para o método de medição da associação de grau com base no grau médio dos nodos adjacentes:

```
knn(graph1)$knnk
```

```
## [1] 5.166667 4.781250 4.666667 4.952381 4.740000 4.428571 4.095238 4.812500
## [9] 4.277778
```

A função tem uma tendência decrescente, mas não estritamente decrescente, oscilando. Por isso mantemos que não é uma rede associativa mas também não pode ser classificada como não associativa.

Distância média

```
# distância média
mean_distance(graph1)
```

```
## [1] 3.474123
```

```
log10(100)
```

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

```
#diâmetro
diameter(graph1)
```

```
## [1] 8
```

A distância média é grande, já que se afasta substancialmente de $\log_{10}(N)$.

A maior distância entre nodos (conectados) é de 8.

Existência de triângulos

```
# Coeficiente de clustering
transitivity(graph1, type="global")
```

```
## [1] 0.03453237
```

É um número baixo de triângulos, já que o coeficiente de *clustering* é um rácio entre o número de triângulos e o número total de ternos conexos e este é baixo.

Identificação de comunidades

TODO

QUESTÃO 2:

Utilize o programa seguinte para gerar a rede aleatória rn2:

```
rn2 <- graph(edge=c(1,2,1,3,2,3,3,4,3,5,4,5,5,6,5,7,6,7,7,8,7,9,8,9,2,4,4,6,6,8),n=100,directed=F);
x = 9;
y = 15;
for (i in 1:91) {
  new<-floor(runif(1,min=1,max=x));
  nn<-neighbors(rn2,new);
  x=x+1;
  y=y+1;
  rn2<-add_edges(rn2,c(new,x));
  newr<-runif(1);
  y=y+1;
  if (newr<0.75) {
    new1 <- floor(runif(1,min=1,max=degree(rn2,new,mode="all")));
    rn2<-add_edges(rn2,c(x,nn[new1]))}
  else {
    new2<-new;
    while (new==new2) new2<-floor(runif(1,min=1,max=x-1));
    rn2<-add_edges(rn2,c(new2,x));
  }
}
```

Qual o método utilizado nesta geração? Justifique.

Caracterize esta rede quanto ao grau médio dos nodos, à conectividade, distância média e existência de triângulos. Aplique ainda métodos de identificação de comunidades.

QUESTÃO 3:

Compare e comente os resultados obtidos nas questões anteriores.