



universidade
de aveiro

Universidade de Aveiro
Mestrado em Engenharia Computacional
Modelação de Sistemas Complexos
40780

PROJECT 5
COMPLEX NETWORKS
ERDŐS-RÉNYI NETWORK

Vasco Costa - 97746

08 de junho de 2023

O estudo de redes é algo transversal a vários ramos da ciência uma vez que a base em questão é em muitos aspetos comum. Um destes é a caracterização da forma de como se descreve o diagrama esquemático com as suas ligações e como estas se dão numa dinâmica individual ou coletiva. As áreas da biologia, da química, de redes neuronais, da medicina, da engenharia e da Internet são exemplos de sistemas compostos por um grande número de unidades dinamicamente ligadas. A primeira abordagem à modelação de redes complexas passa tipicamente pela utilização de grafo com os nodos a representar as unidades dinâmicas e as arestas as interações. Para isto é necessário atender às especificidades estruturais e como estas podem crescer e reproduzir [5, 1]. O entendimento é no entanto limitado pelo controlo que se consegue obter nestes. Apesar de possível, através das ferramentas matemáticas da teoria de controlo, direccionar sistemas naturais para um desejado estado, o mesmo nem sempre é possível e existem estudos que procuram analisar quais os aspetos mais influenciadores [4].

Uma rede de Erdős-Rényi é um modelo de criação de grafos de forma aleatória proposto por Paul Erdős e Alfréd Rényi. Este modelo permite a inserção de arestas com probabilidades independentes, que podem ou não ser diferentes, sendo dos mais utilizados na teoria de grafos [3].

Assim com este projeto pretendem-se responder às questões propostas no enunciado de forma a se conseguirem estudar alguns aspetos de redes complexas e em particular de Erdős-Rényi. Esta será gerada partindo de N nodos isolados e inserindo $L = \frac{N}{2}c$ arestas entre pares destes escolhidos aleatoriamente não permitindo laços e mais do que uma aresta entre o mesmo par. A utilização da matriz de adjacência, A , é a que mais facilmente permite representar esta rede assim como responder às restantes questões propostas.

O grau de um vértice i é definido como sendo $q_i = \sum_{j=1}^N A_{ij}$ e assim a distribuição dos graus de vértices será $P(q) = \frac{N(q)}{N}$ onde $N(q)$ é o número de vértices de grau q [2].

Para as tarefas que se seguem se nada for dito em contrário estas foram concretizadas com os parâmetros sugeridos no enunciado, $N = 10.000$, $c = 50$ e $m = 100$, com m a retratar o número de repetições efetuadas.

Task 1

Com esta tarefa é procurado calcularem-se o grau médio, $\langle q \rangle^1$, o fator de ramificação, B^2 , e o segundo³ e terceiro momentos⁴. Repetindo-se m vezes para se poderem comparar os valores obtidos com os que se podem expectar teoricamente, (1), ou seja, uma distribuição binomial podendo variar e exibir características de uma distribuição de Poisson ou Gaussiana em algumas situações. Neste caso será esperado que seja mais aproximado a uma Poisson [2].

¹ $\langle q \rangle = \frac{1}{N} \sum_{i=1}^N q_i = \sum_{q=0}^{N-1} qP(q)$

² $B = \frac{1}{N\langle q \rangle} \sum_{i=1}^N q_i(q_i - 1) = \frac{1}{\langle q \rangle} \sum_{q=0}^{N-1} P(q)q(q-1) = \frac{\langle q(q-1) \rangle}{\langle q \rangle}$

³ $\langle q^2 \rangle = \frac{1}{N} \sum_{i=1}^N q_i^2$

⁴ $\langle q^3 \rangle = \frac{1}{N} \sum_{i=1}^N q_i^3$

$$\begin{aligned}
P(q) &= e^{-c} \frac{c^q}{q!}, \\
\langle q \rangle &= c, \\
B &= c
\end{aligned} \tag{1}$$

O quociente $\frac{B}{\langle q \rangle}$ é um fator que é também sugerido para analisar que deve ser próximo de 1 dado que ambos devem ser iguais a c . Na Tabela 1 são apresentados os resultados médios com o respetivo desvio padrão σ para 100 repetições. De notar que o valor de $c = 50$ pelo que tanto $\langle q \rangle$ e B devem ficar muito próximos deste valor assim como o seu quociente de 1. Analisando os resultados obtidos verifica-se que estes seguem o que se poderia esperar teoricamente sendo que em média o valor de B é inferior a $\langle q \rangle$ e assim o seu coeficiente também é inferior a 1, como se confirmou.

Tabela 1 – Valores médios e respetivo desvio padrão para diferentes parâmetros em análise em 100 repetições para $c = 50$.

	$\langle q \rangle$	B	$\langle q^2 \rangle$	$\langle q^3 \rangle$	$\frac{B}{\langle q \rangle}$
Valor médio	50	49,9958	2549,7922	132513,7699	0,99992
σ	0	0,015505	0,77527	120,3196	0,00031011

Na Figura 1 comparam-se os resultados das distribuições de probabilidade média obtida dos graus dos vértices, $P_{obtained}$, com a que seria de esperar, P_{theory} . Como se pode verificar estes resultados são muito semelhantes existindo uma sobreposição evidente sendo tanto a forma geral seguida assim como os valores numéricos que são bastante próximos.

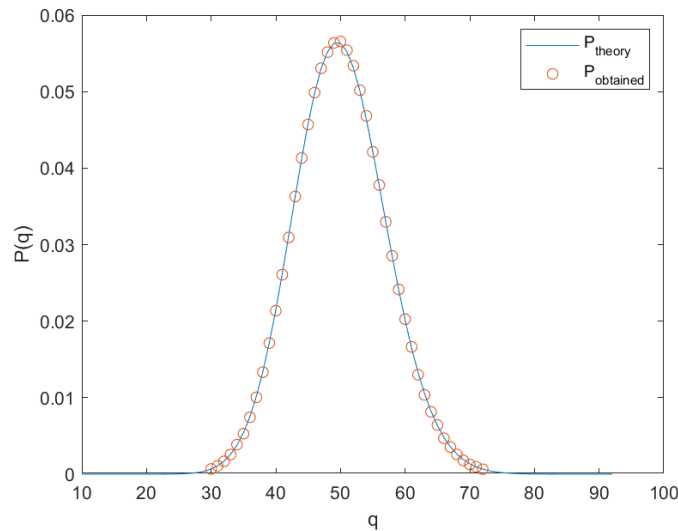


Figura 1 – Gráfico da distribuição dos graus dos vértices para $c = 50$.

Para analisar de que forma o número de arestas presente no grafo é um fator influenciador também se concretizaram experiências para $c = 100$ e $c = 500$. Na Tabela 2 são apresentados

os valores numéricos obtidos com os respectivos desvios padrão em 100 repetições. Estes aqui também vão de encontro ao que já se poderia esperar de (1). Da mesma maneira o quociente $\frac{B}{\langle q \rangle}$ é próximo de 1 em ambas as situações e inferior a 1 assim como o valor de B que é em média inferior a $\langle q \rangle$ como aconteceu para $c = 50$.

Tabela 2 – Valores médios e respectivo desvio padrão para diferentes parâmetros em análise em 100 repetições para $c = 100$ e $c = 500$.

c		$\langle q \rangle$	B	$\langle q^2 \rangle$	$\langle q^3 \rangle$	$\frac{B}{\langle q \rangle}$
100	Valor médio	100	99,9942	10099,4162	1029896,8478	0,99994
	σ	0	0,014173	1,4173	428,301	0,00014173
500	Valor médio	500	499,9755	250487,7395	125729150,4551	0,99995
	σ	0	0,013158	6,5791	9844,916	2,6316e-05

Na Figura 2 são apresentados os gráficos de probabilidade de grau dos vértices para ambos os valores de c . No gráfico da Figura 2b a linha teórica não é possível de representar devido ao cálculo de $q!$ que necessitaria de ser efetuado, $P(q) = e^{-c} \frac{c^q}{q!}$. Com o crescimento de q o seu fatorial é muitas vezes superior. Ainda assim, a tendência geral destes é semelhante com os resultados a acompanharem a linha teórica, 2a. Daqui visualiza-se também que o maior valor da probabilidade é para o valor de $q = c$ como acontece na Figura 1 sendo que para um valor superior de c este máximo é menor.

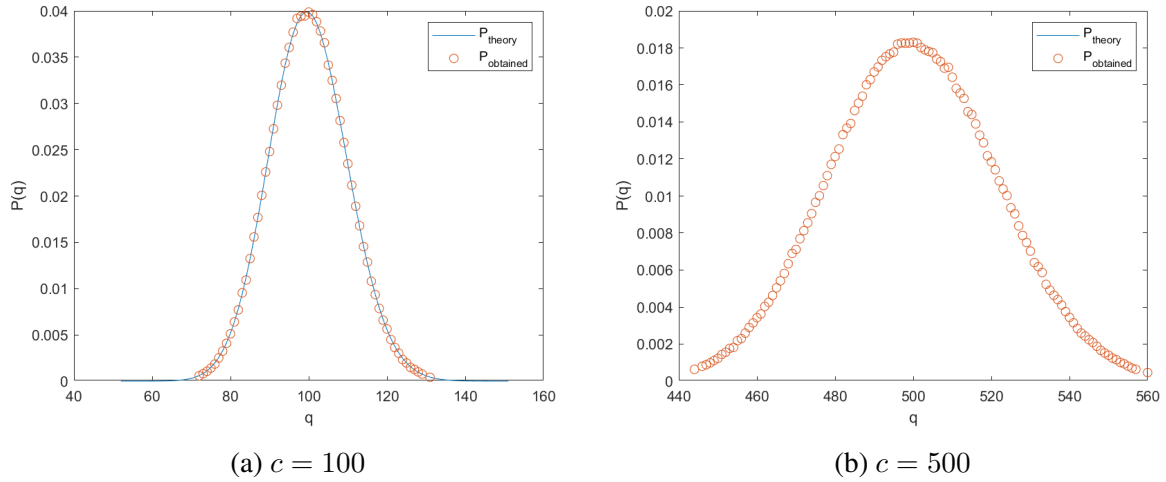


Figura 2 – Gráficos da distribuição dos graus dos vértices para diferentes valores de c .

Task 2

Neste exercício pretende-se calcular o coeficiente de agrupamento $C = \frac{n_{tr}}{n_{pt}}$ em que o número de triângulos possíveis, $n_{pt} = \frac{1}{6}N\langle q(q-1) \rangle$, e o número de triângulos na rede, $n_{tr} = \frac{1}{6} \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N A_{ij} A_{jk} A_{ki}$. Contudo, esta tarefa consegue ser computacionalmente pesada, $\mathcal{O}(N^3)$, e por isso pode-se fazer uso das propriedades da matriz de adjacência e constatar que ao elevar a matriz de adjacência A a um grau l , A^l , obtém-se o número de cami-

nhos de tamanho l . Apesar do produto de matrizes também apresentar uma complexidade elevada esta consegue ser menor e o tempo de computação foi confirmado como o sendo, `timer_method.m`. Assim, com o traço da matriz A^3 consegue-se obter o número de triângulos na rede. O valor do traço de A^3 tem no entanto que ser dividido por $3!$ uma vez que são calculadas todas as permutações possíveis dos pontos ijk sem serem no entanto triângulos distintos.

Das redes de Erdős-Rényi é também possível concluir que como traço de A^3 é igual a $\sum_{i \neq j \neq k} A_{i,j} A_{j,k} A_{k,i} = \sum_{i \neq j \neq k} p^3 = N(N-1)(N-2)q^3$ que é aproximadamente $\langle q^3 \rangle$, $p = \frac{\langle q \rangle}{N-1}$ pelo que se pode definir $C_{approx} \approx \frac{\langle q \rangle}{N}$. Na Tabela 3 é apresentado o gráfico dos valores de C obtidos em cada iteração e a sua comparação a C_{approx} . Efetuando a média destes valores e do seu desvio padrão chega-se a que $C = 0,0050019$ com $\sigma_C = 3,5692e-05$ e $C_{approx} = 0,005$ com $\sigma_{C_{approx}} = 8,7173e-19$ assim estes dois são bastante próximos. O desvio de $\sigma_{C_{approx}}$ é no entanto pouco relevante uma vez que os valores de $\langle q \rangle$ nunca apresentaram diferença e por isso C_{approx} foi sempre constante.

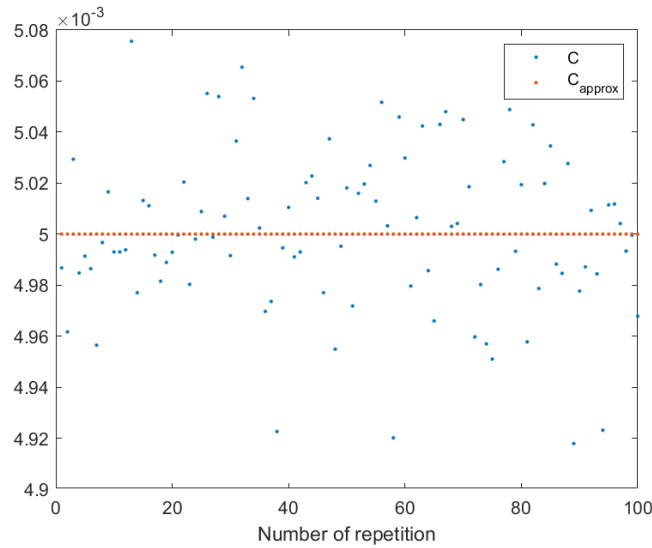


Figura 3 – Valores de C e do quociente $C_{approx} = \frac{\langle q \rangle}{N}$ para as diferentes repetições com $c = 50$.

Para as experiências em que $c = 100$ e $c = 500$ é também possível concretizar a mesma análise e na Figura 4 apresentam-se os gráficos de dispersão. O valor médio para $c = 100$ foi de $C_{100} = 0,010006$ com $\sigma_{C_{100}} = 2,3063e-05$ e $C_{approx100} = 0,01$ com $\sigma_{C_{approx100}} = 1,7435e-18$. Para $c = 500$ o valor médio foi de $C_{500} = 0,05001$ com $\sigma_{C_{500}} = 1,0759e-05$ e $C_{approx500} = 0,05$ com $\sigma_{C_{approx500}} = 1,3948e-17$. Também nesta situação se verifica a proximidade dos resultados, sendo o valor de C_{500} dez vezes superior a C , ($c = 50$), como se poderia esperar uma vez que o valor de N é o mesmo mas de $\langle q_{500} \rangle$ é dez vezes superior.

Este cálculo apresentou um valor baixo como se esperaria e pretendia uma vez que este mede o quanto os nodos numa rede tendem a agrupar-se. Dado que aqui esta é gerada de forma aleatória e com as arestas a serem independentes o valor reduzido seria expectável. Este cálculo

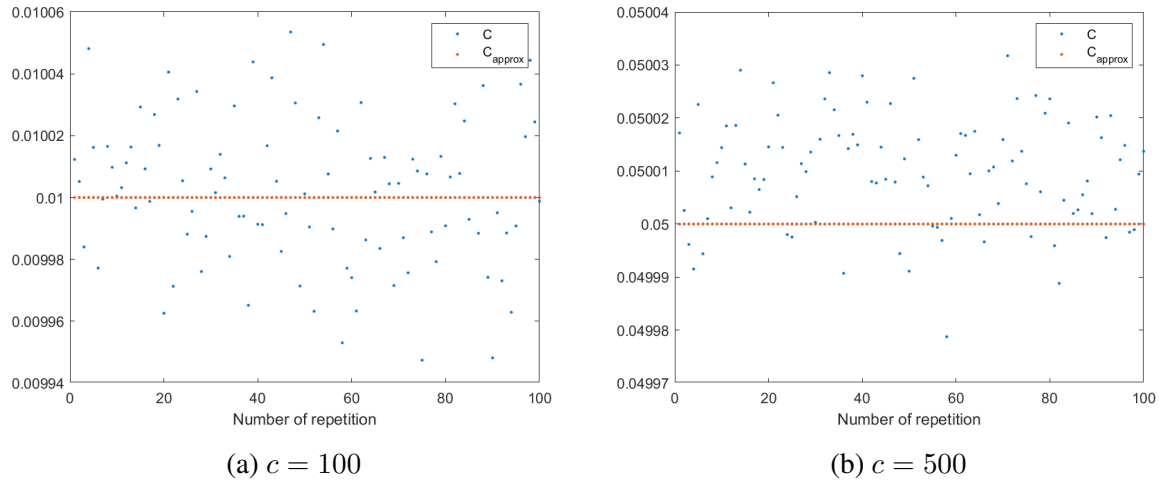


Figura 4 – Valores de C e do quociente $C_{approx} = \frac{\langle q \rangle}{N}$ para as diferentes repetições com para diferentes valores de c .

é usualmente feito na literatura em conjuntos de 3, ou seja, em triângulos como aqui foi efetuado mas existem outros comprimentos maiores que podem ser estudados.

Task 3

A análise do coeficiente de Pearson, ρ , para caracterizar a correlação linear entre os graus dos nodos na rede é sugerido e por isso pretende-se efetuar o seu estudo com esta *Task*. Este pode ser calculado com a equação 2 onde $Q = \frac{\langle q^2 \rangle}{\langle q \rangle}$ e $\sigma^2 = \frac{\langle q^3 \rangle}{\langle q \rangle} - \frac{\langle q^2 \rangle^2}{\langle q \rangle^2}$.

$$\rho = \frac{\sum_{i=1}^N \sum_{j=1}^N A_{ij} (q_i - Q)(q_j - Q)}{N \langle q \rangle \sigma^2} \quad (2)$$

Na Figura 5 apresenta-se o gráfico de dispersão obtido dos valores de ρ para cada uma das repetições. Efetuando a média obtém-se que $\rho = -0,00024882$ com $\sigma_\rho = 0,0020538$. Este resultado mostra que a correlação medida é perto de nula como se pretendia obter devido à forma como os grafos são criados, sendo esta uma das propriedades dos grafos de Erdős-Rényi [3].

Nas experiências em que $c = 100$ e $c = 500$ este parâmetro também pode ser medido e obtiveram-se os valores de $\rho_{100} = -0,0002002$ com $\sigma_{\rho_{100}} = 0,0015522$ e $\rho_{500} = -0,00016616$ com $\sigma_{\rho_{500}} = 0,00054614$, respetivamente. Na Figura 6 apresentam-se os gráficos de dispersão que também retratam a correlação perto de nula sendo o seu valor médio indicativo do mesmo.

Conclusão

Assim com esta análise conseguiram-se verificar características estruturais de uma rede de Erdős-Rényi. A distribuição de probabilidade dos graus dos vértices verificou-se que seguia

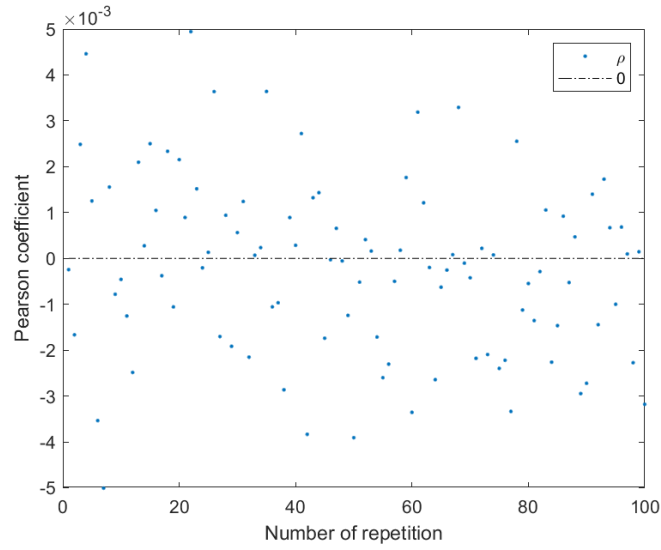


Figura 5 – Valores de ρ obtido em cada iteração para $c = 50$ com a linha em 0 para referência.

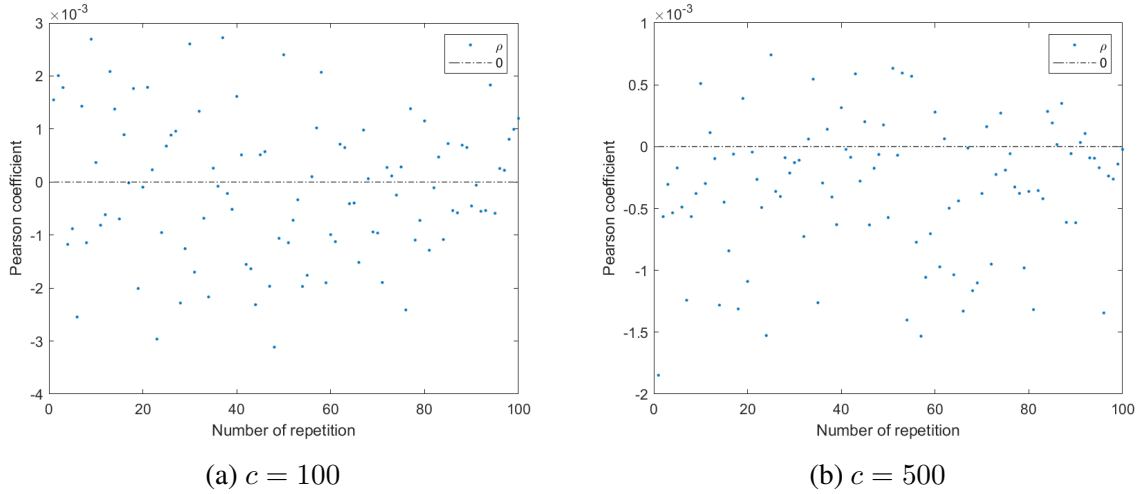


Figura 6 – Valores de ρ obtido em cada iteração para diferentes valores de c com a linha em 0 para referência.

a distribuição teórica com uma proximidade bastante grande. O coeficiente de Pearson, ρ , e do de agrupamento, C , obtidos revelam que as redes geradas apresentam um grande nível de independência entre arestas. Algo que se esperava já devido ao estilo de rede utilizado, sendo que os valores numéricos obtidos vão de encontro aos que se esperavam teoricamente. Estes elementos verificaram-se para diferentes variações à rede, experiências para diferentes valores de c .

O estudo destas métricas é relevante até para outros tipos de redes em que se pretende verificar as suas propriedades para utilizar noutras áreas como a de redes sociais, modelos epidemiológicos ou desenho de infraestruturas como a Internet. O nível de agregação implica a forma como a informação ou doenças são propagadas. Um baixo nível de correlação permite que a infraestrutura permita ser robusta e resiliente.

Referências

- [1] S. Boccaletti et al. “Complex networks: Structure and dynamics”. Em: *Physics Reports* 424.4 (2006), pp. 175–308. ISSN: 0370-1573. DOI: <https://doi.org/10.1016/j.physrep.2005.10.009>. URL: <https://www.sciencedirect.com/science/article/pii/S037015730500462X>.
- [2] Rui Américo Ferreira da Costa. *Modelling of Complex Systems - Complex networks*. eLearning. Lecture_12 & Lecture_13. Mai. de 2023.
- [3] Paul L. Erdos e Alfréd Rényi. “On the evolution of random graphs”. Em: *Transactions of the American Mathematical Society* 286 (1984), pp. 257–257.
- [4] Yang-Yu Liu, Jean-Jacques Slotine e Albert-László Barabási. “Controllability of complex networks”. Em: *Nature* 473.7346 (mai. de 2011), pp. 167–173. ISSN: 1476-4687. DOI: [10.1038/nature10011](https://doi.org/10.1038/nature10011). URL: <https://doi.org/10.1038/nature10011>.
- [5] Steven H. Strogatz. “Exploring complex networks”. Em: *Nature* 410.6825 (mar. de 2001), pp. 268–276. ISSN: 1476-4687. DOI: [10.1038/35065725](https://doi.org/10.1038/35065725). URL: <https://doi.org/10.1038/35065725>.