

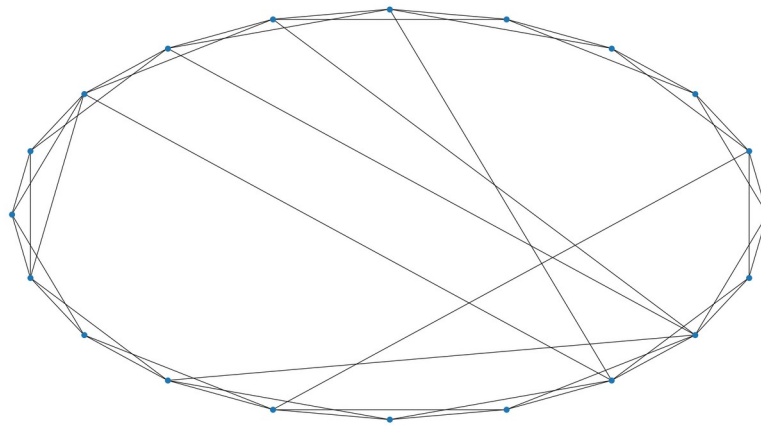
Small world networks

Matheus Farnese Lacerda Senna

Departamento de Ciência da Computação – Universidade Federal de Minas Gerais
Belo Horizonte, MG – Brasil

1. Construção da rede

O primeiro passo para a análise de uma small world network é a construção de dita rede. Para tal, foi utilizada a biblioteca python NetworkX. A seguir está ilustrada a construção de uma rede seguindo a forma explicada na especificação desse trabalho prático. Tal rede possui $N = 20$, $Z = 4$, $L = 0.2$, assim como o exemplo dado na página 2 da especificação.



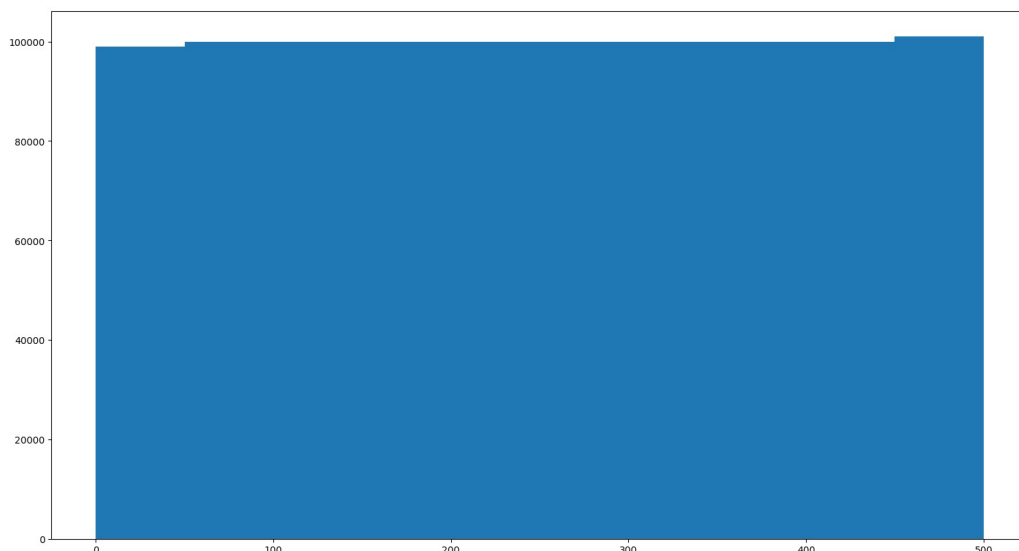
Observa-se que o grafo possui 6 arestas longas, além das 4 arestas que cada vértice possui que liga-os a seus 4 vizinhos mais próximos. O resultado está de acordo com o esperado.

2. Funções para o cálculo de distâncias

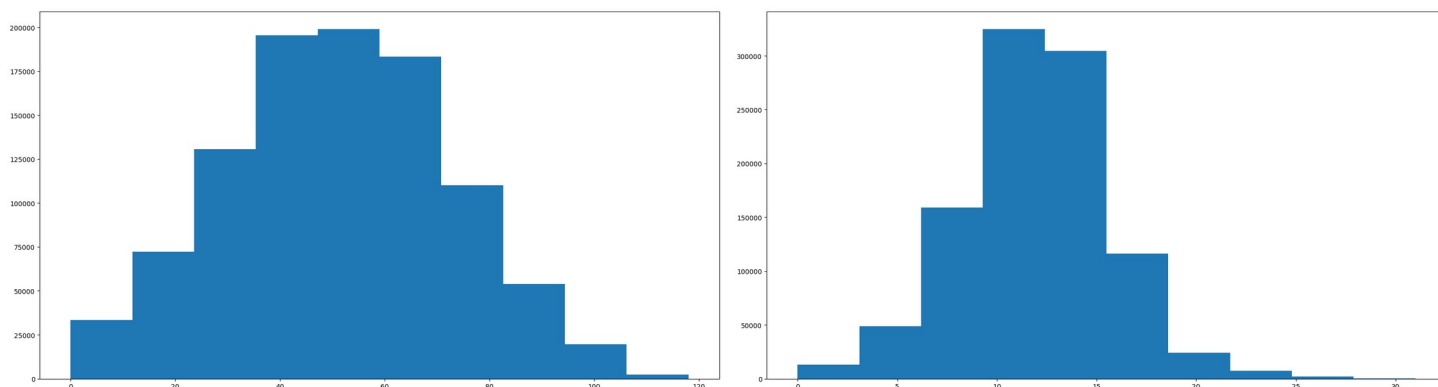
Foram implementadas 3 funções:

- 1) Calcula a menor distância entre um dado vértice e todos os outros vértices
- 2) Chama a função anterior para todos os pares de vértices, obtendo-se assim todas as menores distâncias do grafo
- 3) Calcula a média do menor caminho médio de cada vértice.

A função 2 foi testada para uma rede de $N = 1000$ e $Z = 2$. A seguir estão três histogramas com todos os menores caminhos entre cada par de vértices.



O histograma acima foi obtido com $p = 0$, verificando-se o resultado esperado: distribuição constante dos caminhos para tamanho entre 0 e 500 (N/Z).



Os histogramas acima foram obtidos com, respectivamente, $p = 0.02$ e $p = 0.2$. Observa-se que para o caso $p = 0.02$, a distribuição dos caminhos assumiu a forma de uma normal com média aproximadamente 50. Já no caso $p = 0.2$, o histograma também assumiu a forma de uma normal, mas com média aproximadamente 10. Verifica-se uma drástica redução do comprimento dos menores caminhos do grafo. No caso anterior ($p = 0$), a média era de aproximadamente 250. Esse valor caiu para 50 e depois para 10. Verifica-se que a inclusão de arestas longas (atalhos) possui um grande impacto no comprimento dos caminhos. A saber:

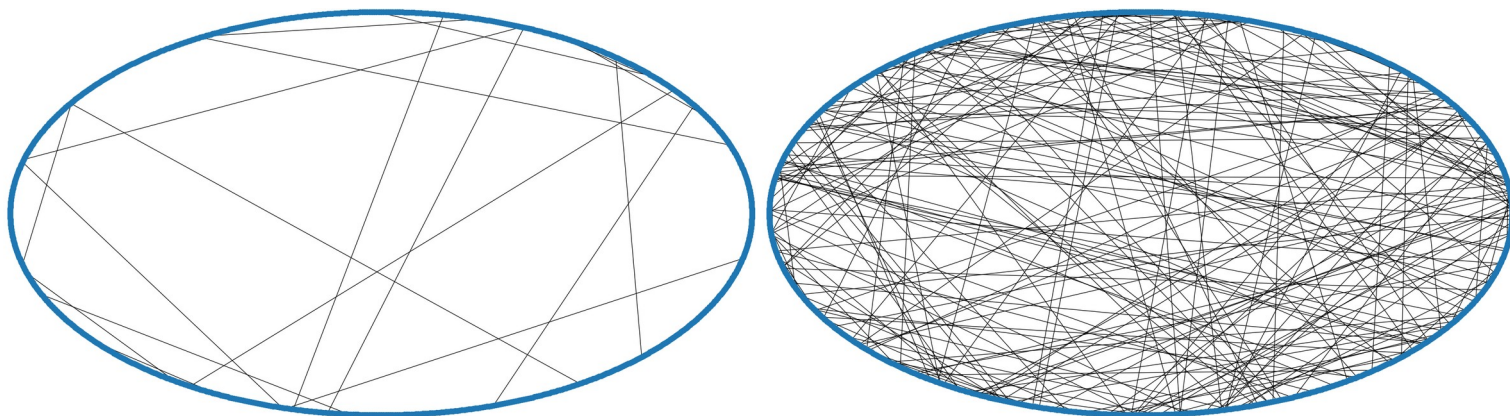
Número de atalhos para $p = 0$: 0

Número de atalhos para $p = 0.02$: 20

Número de atalhos para $p = 0.2$: 199

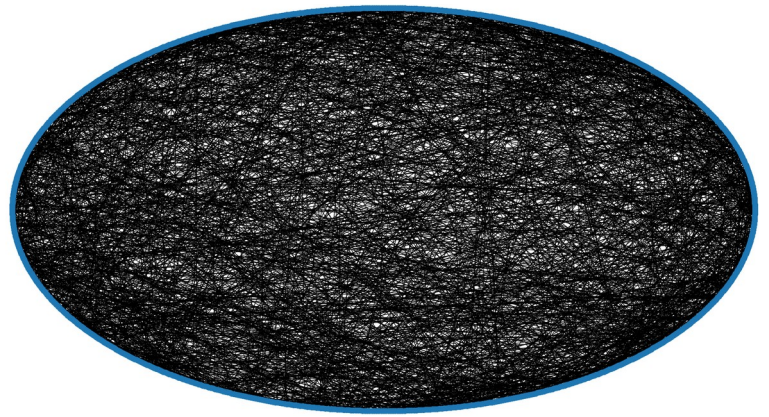
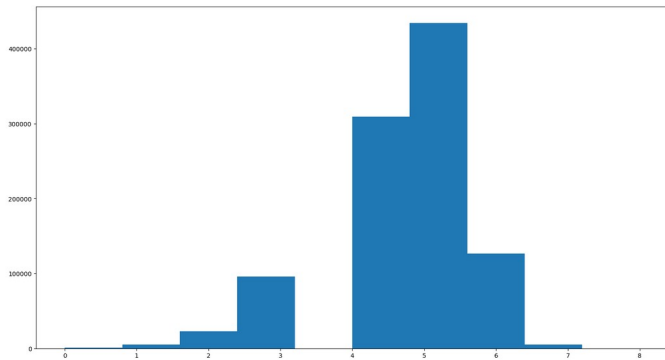
Com o aumento de p por um fator de 10, do caso 2 para o caso 3 o número de atalhos aumenta por esse mesmo fator. Já os valores dos caminhos decrescem por um fator de 5, ou seja, para a média: $50/5 = 10$.

Os grafos gerados com $p = 0.02$ e $p = 0.2$ estão expressos a seguir, respectivamente.



Observa-se que, de fato, o gráfico da direita ($p = 0.2$) é muito mais ligado do que o da esquerda.

Finalmente, para obter os “seis graus de separação” do resultado empírico bastante conhecido sobre a conectividade de redes (como dito na especificação da atividade: “vários estudos, feitos de formas pouco controladas, mostraram que quaisquer duas pessoas escolhidas aleatoriamente no mundo podem ser conectadas, uma à outra, por uma pequena cadeia de pessoas (que normalmente contém cerca de seis pessoas), sendo que cada uma destas conhece bem a próxima pessoa da cadeia”). O resultado obtido foi que é preciso um $p = 1.5$ para que consistentemente (95% das vezes) possa-se ligar quaisquer duas pessoas uma a outra em uma rede com $N = 1000$ e $Z = 2$. O histograma e o grafo associados a esse resultado estão expressos a seguir:



Percebe-se que é necessário uma alta conectividade da rede para obter o resultado dos 6 links.

Finalizando as funções implementadas, testou-se a função 3 (obtenção da média do menor caminho médio de cada vértice) em uma rede de $N = 100$, $Z = 2$ e $p = 0.1$. Foram geradas 20 redes diferentes, sendo que os resultados estão expressos a seguir:

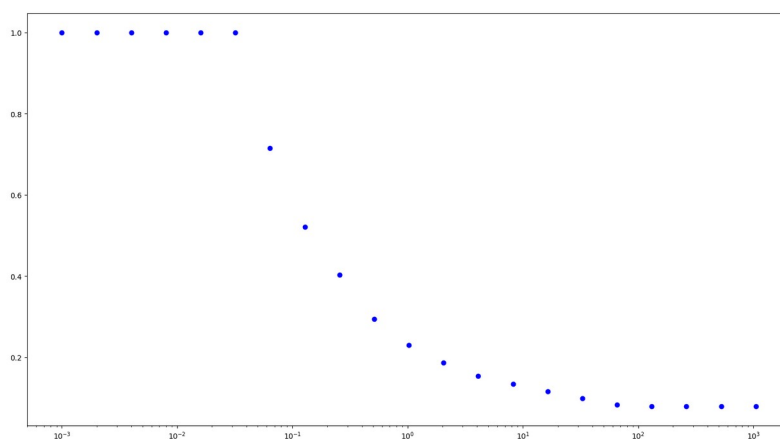
Caminho médio: 9.778 – 13.2426 – 9.5512 – 9.3546 – 9.6292 – 10.89 – 9.9138 – 8.8102 – 10.1594 – 8.9452 – 8.6178 – 9.1092 – 8.8134 – 9.0548 – 9.711 – 9.21 – 10.3456 – 9.7246 – 9.3918 – 9.088

Número de arestas longas: 10 – 9 – 10 – 9 – 10 – 10 – 10 – 10 – 10 – 10 – 10 – 10 – 10 – 10 – 9 – 9 – 10 – 10 – 10 – 10

Observa-se que os resultados dos caminhos médios estão de fato oscilando perto de 10, como era esperado. De fato, essa oscilação é esperada, pois as arestas longas podem ligar pontos mais próximos ou mais longes. No caso em que a maioria delas liguem pontos distantes da rede, o caminho médio então será um pouco menor do que no caso onde os pontos ligados por elas não são tão distantes entre si. Ademais, verifica-se também que existem, normalmente, 10 arestas longas nessa rede.

3. Comprimento médio do caminho

Foi utilizada uma rede com $N = 50$, $Z = 2$ e p variando entre 0.001 e 1000. Em seguida, calculou-se, para cada p (eixo x), o valor de $d(p) / d(p=0)$ (eixo y), onde $d(x)$ é o caminho médio obtido pela função 3 explicada na seção anterior. Os resultados estão plotados no gráfico a seguir:



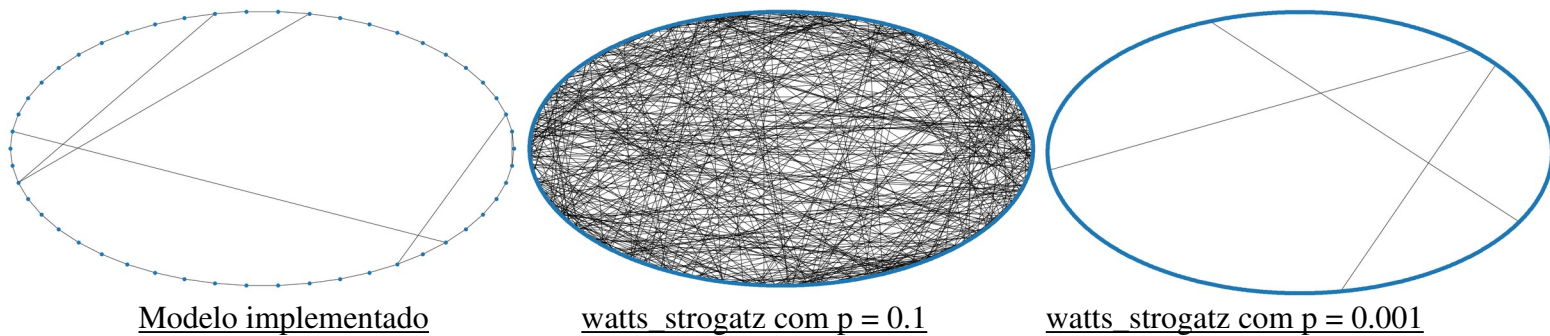
Verifica-se o comportamento esperado: decrescimento exponencial da razão $d(p) / d(p=0)$ com o aumento de p , de forma a obter uma curva parecida com a curva citada na especificação do trabalho (Watts e Strogatz; Fig.2).

Uma segunda observação é o fato de que os pontos iniciais do gráfico estão em $y = 1$, ou seja, $d(p) = d(p=0)$. Isso ocorre pois, para valores baixos de p (no caso, de $p = 0.001$ até $p = 0.032$), o número de atalhos possíveis ($pNZ / 2$) é muito baixo (0.05 para $p = 0.001$ e 1.6 para $p = 0.032$) já

para o próximo $p = 0.064$, $pNZ / 2$ passa a ser 3.2, o que acaba por introduzir pelo menos um atalho no grafo e reduzir drasticamente a razão $d(p) / d(p=0)$. Para o caso em que $p = 0.032$, o atalho gerado aleatoriamente (pois $pNZ / 2 = 1.6$, o que implica a geração de uma possível aresta longa) aconteceu de cair em uma aresta já existente, ou seja, não introduziu caminhos novos.

4. Comparação com Watts e Strogatz

Foi utilizada uma rede com $N = 50$, $Z = 2$ e $p = 0.1$, a qual foi gerada utilizando o modelo implementado. Tal rede foi posta em comparação com duas redes geradas pelo modelo de Watts e Strogatz utilizando a função `nx.watts_strogatz(N, k, p)` da biblioteca NetworkX. Ditas redes foram geradas com $N = 1000$, $k = 10$ e $p = 0.1$ para uma rede e $p = 0.001$ para a outra. Os grafos gerados estão expressos a seguir:



Observa-se que a rede do modelo `watts_strogatz` com $p = 0.001$ é mais estatisticamente semelhante ao modelo implementado do que a com $p = 0.1$. Para corroborar tal afirmação, foram calculados os $\Delta\theta = \pi Z d / N$ para as três redes, bem como o número de atalhos na rede gerada pelo modelo implementado e o valor de $M = pNZ / 2$ para as redes de `watts_strogatz`. Os resultados obtidos foram:

$$\Delta\theta = 1.2005407830134178$$

$$\Delta\theta(p=0.1) = 0.13954546160200895$$

$$\Delta\theta(p=0.001) = 1.0952217167352083$$

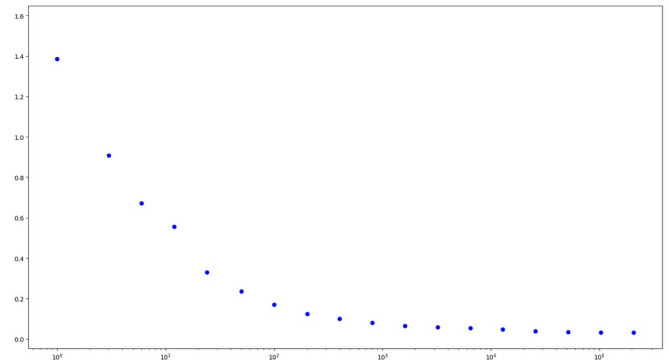
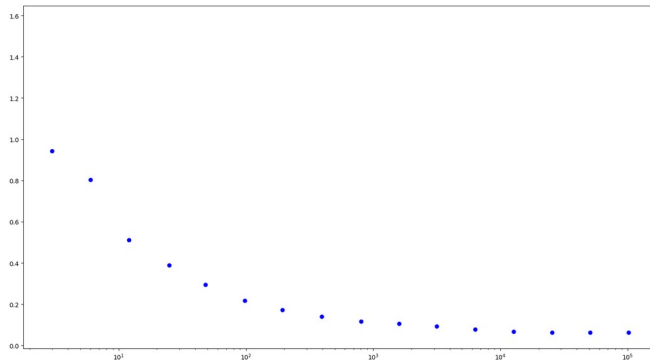
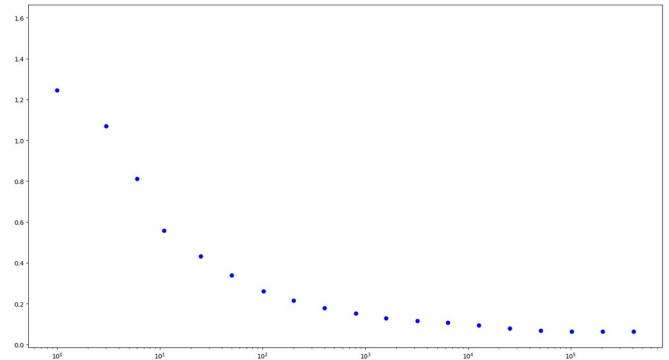
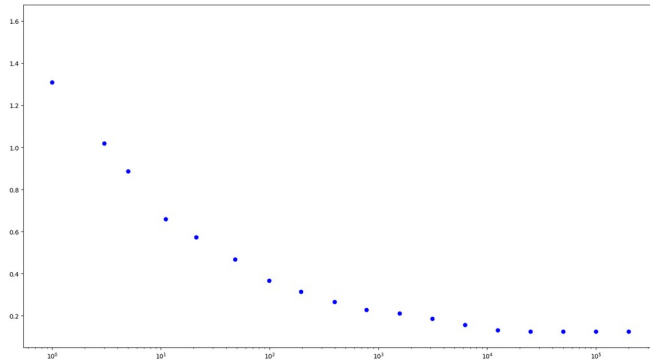
$$\text{Atalhos} = 5 \qquad M(p=0.1) = 500.0 \qquad M(p=0.001) = 5.0$$

Observa-se que o número de atalhos na rede implementada é igual ao $M(p=0.001)$ para a rede de `watts_strogatz`. Ademais, o $\Delta\theta$ do modelo implementado é bem mais próximo do $\Delta\theta$ do modelo de `watts_strogatz` com $p = 0.001$.

Assim sendo, verifica-se que, de fato, o modelo `watts_strogatz` funcionou em um valor de ZN um fator de 100 maior do que o modelo implementado e, portanto, os valores de p para esse último são um fator de 100 maior para que o número de atalhos seja igual ao valor de $M = pZN / 2$ do `watts_strogatz`.

5. Comprimento médio de ligação

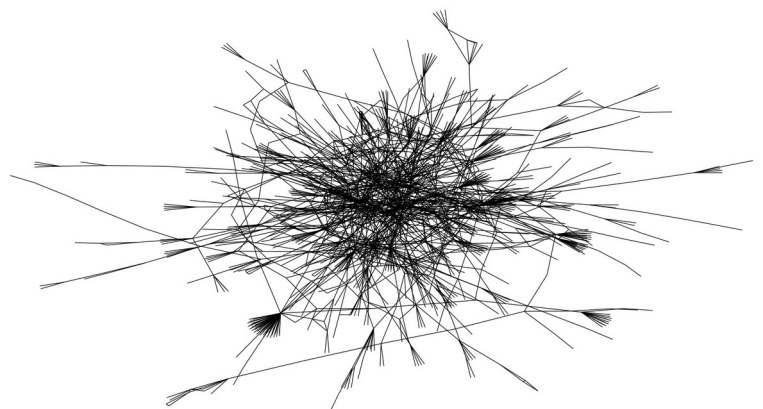
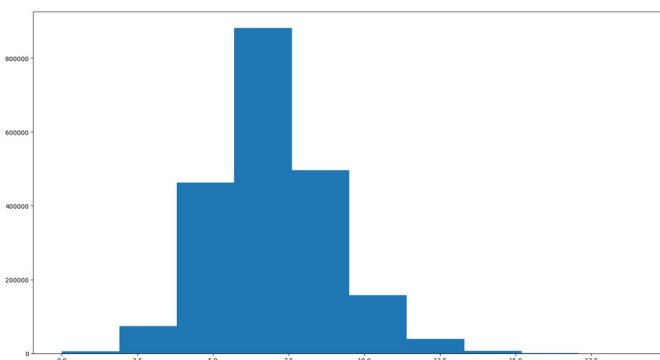
Para finalizar a análise do modelo de small world networks, foram gerados gráficos do número total de atalhos na rede (eixo x) versus o comprimento médio do caminho $\Delta\theta$ (eixo y) para valores de p entre 0.001 e 1000. As redes utilizadas foram $N = 100$ ou 200 e $Z = 2$ ou 4 . Os gráficos gerados estão expressos a seguir:

N = 100**N = 200****Z = 2****Z = 4**

Observa-se um comportamento semelhante entre os gráficos: a medida que o p vai crescendo, o número de atalhos idem e o $\Delta\theta$ diminui exponencialmente. Tal comportamento é esperado, pois, quanto maior o p , por construção, mais atalhos aparecem. Um número maior de atalhos implica em um d (distância média) menor e, conseqüentemente, $\Delta\theta = \pi Z d / N$ decresce, pois Z e N são fixos dado um gráfico.

6. Redes reais

Para finalizar a atividade, foi utilizada a rede obtida em: <https://networkrepository.com/bio-yeast-protein-inter.php> para verificar a distância média e o histograma das distâncias entre os vértices em dita rede. De forma a facilitar a análise, foi considerado apenas o maior componente conectado (que conecta 1458 dos 1870 vértices totais presentes na rede). A distância média obtida foi de 6.80769361039137. O histograma com os menores caminhos entre cada par de vértices, bem como o grafo em si, estão ilustrados a seguir:



Observa-se que, de fato, o histograma assume a forma de uma normal com média próxima de 6.8. Devido ao alto número de vértices, a visualização do grafo não é trivial, mas percebe-se a existência de um “centro”, onde os vértices estão bem conectados entre si, e uma “periferia”, onde os vértices mais externos possuem um grau médio menor.

7. Apêndice

Os códigos e arquivos usados para as análises estão disponíveis no seguinte link do google drive:

<https://drive.google.com/drive/folders/1hk0JOsdK21fYluIo62HkOui0dKB51PGI?usp=sharing>

O código em python usado para a geração e análise de small world networks está disponível no arquivo “small_world.py”. Para executá-lo, basta usar um terminal e usar a linha a seguir:

```
python3 small_world.py
```

Esse comando irá rodar todas as análises realizadas, de forma a responder tudo o que foi pedido nas letras de ‘a’ a ‘d’ da especificação. Caso queira, pode-se modificar os parâmetros das chamadas de funções nas 5 últimas linhas do arquivo.

Ademais, para a letra ‘e’ da especificação (análise de uma rede real), foi utilizado um segundo arquivo, nomeado de “real_network.py”. Para executá-lo, basta usar um terminal e usar a linha a seguir:

```
python3 real_network.py
```

Esse comando irá ler o arquivo “proteins.txt”, também disponibilizado, e irá gerar os resultados obtidos na seção 6 desse relatório.