

## SME0130 - REDES COMPLEXAS

# Projeto 1 - Caracterização de redes complexas: grau, clustering e caminho

*Aluno:*

Gil Barbosa Reis Nº 8532248

*Professor:*

Francisco Aparecido Rodrigues

INSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO  
UNIVERSIDADE DE SÃO PAULO

São Carlos, SP  
30 de setembro de 2016

## Sumário

<b>1</b>	<b>Introdução</b>	<b>4</b>
<b>2</b>	<b>Redes utilizadas</b>	<b>4</b>
<b>3</b>	<b>Métodos</b>	<b>4</b>
3.1	Distribuição de Lei de Potência . . . . .	5
3.2	Distribuição de probabilidade do Grau . . . . .	5
3.3	Distribuição de probabilidade acumulada complementar do Grau	5
3.4	Grau médio . . . . .	5
3.5	Segundo momento do grau . . . . .	6
3.6	Entropia de Shannon . . . . .	6
3.7	Coefficiente de aglomeração local . . . . .	6
3.8	Transitividade . . . . .	6
3.9	Média dos menores caminhos . . . . .	7
3.10	Eficiência . . . . .	7
3.11	Diâmetro . . . . .	7
<b>4</b>	<b>Resultados</b>	<b>7</b>
4.1	Medidas . . . . .	7
4.2	Distribuição do grau . . . . .	8
4.3	Coefficiente de aglomeração . . . . .	8
4.4	$k(i)$ vs $cc(i)$ . . . . .	9
4.5	Lei de Potência . . . . .	9
<b>5</b>	<b>Conclusão</b>	<b>10</b>

## Lista de Tabelas

1	Resultados das medidas para cada rede . . . . .	8
2	Coefficiente de Pearson entre Grau e Coefficiente de Aglomeração em escala log-log . . . . .	9
3	Parâmetros da Lei de Potência para a distribuição do grau das redes . . . . .	9

## Lista de Figuras

1	Distribuição de probabilidade do grau e distribuição acumu- lada complementar . . . . .	12
---	--	----

2	Distribuição acumulada do coeficiente de aglomeração . . . . .	13
3	Grau x Aglomeração local de cada rede . . . . .	14

# 1 Introdução

O estudo da estruturas de redes reais são importantes para podermos entender mais sobre o mundo em que habitamos. A partir desse estudo, podemos identificar vários comportamentos apresentados pelos indivíduos e suas interações, possibilitando por exemplo previsões comportamentos a acontecer.

## 2 Redes utilizadas

Foram utilizadas as seguintes redes reais disponibilizadas pelo site [http://konect.uni-koblenz.de/\[kon\]](http://konect.uni-koblenz.de/[kon]) para efetuar as medidas e testes:

- Hamster friendships[ham ]: Amizades entre usuários do website `hamsterster.com`.
- Les Misérables[mis ]: Co-ocorrências de personagens na obra "Os miseráveis" de Victor Hugo, onde os nós representam os personagens e arestas representam que os personagens aparecem no mesmo capítulo do livro.
- US power grid[gri ]: Rede de linhas de energia dos estados do leste dos Estados Unidos.
- Euroroad[eur ]: Rede das estradas da E-road, estradas situadas principalmente na Europa, onde cada nó é uma cidade, e arestas são as estradas.
- Air traffic control[air ]: Rotas preferenciais entre aeroportos e centros de serviço, baseados nas rotas recomendadas pela National Flight Data Center (NFDC) dos Estados Unidos.
- Human protein (Figeys)[pro ]: Interações entre proteínas em humanos (*Homo Sapiens*).

## 3 Métodos

Várias informações foram extraídas de cada rede para sua caracterização. Foi utilizada a linguagem de programação Python[py ] em conjunto com as bibliotecas Scipy[sci ], NetworkX[nx ], Matplotlib[plo ] e Powerlaw[pow ].

Alguns dos conceitos foram tirados dos artigos [SHANNON 1948] e [BARABASI 1999], outros de anotações do material de aula.

Os conceitos apresentados a seguir foram usados no experimento.

### 3.1 Distribuição de Lei de Potência

Grandes redes tendem a seguir a distribuição de Lei de Potência, que apresenta as seguintes características: a rede aumenta pela adição de novos vértices; novos vértices se relacionam com vértices já conectados.

A distribuição de probabilidade  $P(k)$  representando a probabilidade de um vértice apresentar grau  $k$  para redes que seguem a Lei de Potência é representada pela fórmula 1.

$$P(k) \sim k^{-y} \quad (1)$$

### 3.2 Distribuição de probabilidade do Grau

A distribuição do grau  $k(i)$  de uma rede relaciona cada vértice  $i$  da rede com seu grau  $k$ . A partir de tal distribuição, é possível obter uma distribuição de probabilidade do grau pela fórmula 2.

$$P(k) = k \quad (2)$$

### 3.3 Distribuição de probabilidade acumulada complementar do Grau

A partir da distribuição do grau também é possível obter uma distribuição acumulada complementar, que segue a fórmula 3. Esse resultado será utilizado em TODO(referenciar o resultado)

$$ccdf(x) = P(K > k) \quad (3)$$

### 3.4 Grau médio

O grau médio de uma rede (4), que é a média dos graus de todos os vértices da rede, representa o quão conectada é a rede.

$$K_{médio} = \frac{\sum_{i=0}^N K_i}{N} \quad (4)$$

### 3.5 Segundo momento do grau

O momento de ordem 2 de uma rede é relacionado com a variância do grau, dado pela equação 5.

$$\langle k^2 \rangle = \frac{\sum_{i=0}^N K_i^2}{N} \quad (5)$$

### 3.6 Entropia de Shannon

A Entropia de Shannon, dada pela equação 6, mede a quantidade de dados necessária para se descrever uma rede. Desse modo, redes com alta entropia precisam de muita informação para serem reconstruídas.

$$H = - \sum_{k=1}^N P(k) \cdot \log(P(k)) \quad (6)$$

### 3.7 Coeficiente de aglomeração local

Esse coeficiente mostra o quanto um vértice é conectado com a rede. A fórmula 7 demonstra como calculá-lo, onde  $\Delta_i$  representa o número de triângulos em que o vértice  $i$  participa.

$$cc(i) = \frac{2 \cdot \Delta_i}{K_i \cdot (K_i - 1)} \quad (7)$$

### 3.8 Transitividade

Transitividade é uma medida de aglomeração assim como o coeficiente descrito na subseção 3.7, mas calculado para a rede inteira. Essa medida representa a probabilidade de vizinhos estarem conectados entre si a partir da quantidade de triângulos  $\Delta$  e de triplas conectadas *triplas*.

A fórmula 8 demonstra como calcular a transitividade.

Note que se todas as triplas conectadas formam triângulos, a transitividade da rede  $T = 1$ .

$$T = \frac{3 \cdot \Delta}{triplas} \quad (8)$$

### 3.9 Média dos menores caminhos

Essa medida representa a média entre todos os menores caminhos entre todos os vértices de uma rede. Se uma rede tem esse valor alto, ela é menos conectada, pois vértices se encontram afastados, em geral.

A fórmula 9 demonstra como calcular essa medida, sendo  $d_{ij}$  o comprimento do menor caminho entre os nós  $i$  e  $j$ .

$$l = \frac{1}{N \cdot (N - 1)} \sum_{i=1}^N \sum_{j=1, j \neq i}^N d_{ij} \quad (9)$$

### 3.10 Eficiência

A eficiência mede o quão fácil é trafegar uma informação pela rede, e está relacionada com o inverso das distâncias mínimas. Ela é utilizada em redes que não são formadas por somente um componente conexo, onde existe um  $d_{ij} \rightarrow \infty$  entre vértices de componentes diferentes.

A fórmula 10 mostra como calcular a eficiência. Note que para um  $d_{ij} \rightarrow \infty$ ,  $\frac{1}{d_{ij}} = 0$ , possibilitando a eficiência ser calculada para a rede inteira independente da quantidade de componentes conexos.

$$E = \frac{1}{N \cdot (N - 1)} \sum_{i=1}^N \sum_{j=1, j \neq i}^N \frac{1}{d_{ij}} \quad (10)$$

### 3.11 Diâmetro

Diâmetro é a maior distância entre as menores distâncias entre vértices de uma rede, calculada por 11. Note que essa medida só pode ser calculada para componentes conexos.

$$D = \max(d_{ij}) \quad (11)$$

## 4 Resultados

### 4.1 Medidas

Várias medidas foram tomadas de cada rede, apresentadas na tabela 1.

A partir dos resultados, percebe-se que a rede com menor média de menores caminhos e maior transitividade e eficiência é a Les Misérables, o que significa que ela é muito conectada.

	Air traffic control	Human protein	Les Misérables	US power grid	Hamster fri-endships	Euroroad
Grau médio	3.9315	5.7454	6.5974	2.6691	13.4919	2.4140
Segundo momento	29.0196	321.7606	79.5325	10.3327	611.8149	7.2402
Entropia de Shannon	2.5780	11.3339	4.8676	0.7642	18.7655	0.1439
Média do coeficiente de aglomeração	0.0675	0.03995	0.5731	0.0801	0.1414	0.0167
Transitividade	0.0639	0.0076	0.4989	0.1032	0.0904	0.0339
Média dos menores caminhos	5.9290	3.8436	2.6411	18.9892	3.4526	18.3951
Eficiência	0.00020	0.00014	0.00421	0.00002	0.00018	0.00001
Diâmetro	17	10	5	46	14	62

Tabela 1: Resultados das medidas para cada rede

As redes de rodovias (Euroroad) e aeroportos (Air traffic control) não são muito parecidas, pois são dois transportes que funcionam muito diferente. A rede de aeroportos é bem mais conectada e complexa, mas com maior eficiência.

## 4.2 Distribuição do grau

Uma das medidas mais simples e influentes em uma rede é o grau dos vértices. A figura 1 apresenta a distribuição de probabilidade dos graus de cada rede e a distribuição da probabilidade acumulada complementar. Nota-se que a última apresenta uma curva bem mais suave que a primeira.

## 4.3 Coeficiente de aglomeração

Foi calculada a distribuição de probabilidade acumulada do coeficiente de aglomeração das redes. A figura 2 mostra esse resultado. Percebe-se que as redes que apresentam menor conectividade (calculada pela fórmula da transitividade, por exemplo) apresentam uma baixa variação do coeficiente de aglomeração, com os vértices com baixa aglomeração sendo predominantes.



#### 4.4 $k(i)$ vs $cc(i)$

Quando comparadas a distribuição de graus e a distribuição do coeficiente de aglomeração, é possível encontrar uma correlação entre os valores em escala log-log. A figura 3 mostra os valores para cada rede.

A correlação entre os valores do grau e do coeficiente de aglomeração local pode ser vista calculando o Coeficiente de Pearson. Os valores mais próximos do valor  $-1$  mostram uma correlação linear decrescente entre as medidas. A tabela 2 mostra que várias das redes apresentam tal correlação.

Rede	Coeficiente de Pearson
Air traffic control	-0.831544181845
Human protein	-0.855787531257
Les Misérables	-0.502859456827
US power grid	-0.806952987955
Hamsterster friendships	-0.612813215518
Euroroad	-0.940909479633

Tabela 2: Coeficiente de Pearson entre Grau e Coeficiente de Aglomeração em escala log-log

#### 4.5 Lei de Potência

Para cada rede, foram calculados o expoente da distribuição de Lei de Potência e o  $k_{min}$  onde a distribuição de grau encaixa na Lei de Potência. A tabela 3 mostra tais valores.

Rede	Expoente	$K_{min}$
Air traffic control	4.14216526849	9
Human protein	2.52926193941	8
Les Misérables	4.67644606843	10
US power grid	10.4358481583	10
Hamsterster friendships	4.3992005075	71
Euroroad	27.8885207063	8

Tabela 3: Parâmetros da Lei de Potência para a distribuição do grau das redes

## 5 Conclusão

A partir desse experimento foi possível ver como redes reais, apesar de parecerem muito diferentes, apresentam características comuns. Além disso, é possível retificar alguns padrões de comportamento das redes, como por exemplo a maior distância média entre pontos da rodovia (Euroroad [eur ]) em relação às dos aeroportos (Air traffic control [air ]).

É interessante também notar que o estudo das redes pode ajudar na compreensão do comportamento real do mundo em que vivemos.

## Referências

- [kon ] The koblenz network collection. <http://konect.uni-koblenz.de/>.
- [plo ] Matplotlib. <http://matplotlib.org/>.
- [nx ] Networkx. <http://networkx.readthedocs.io/en/latest/>.
- [pow ] Powerlaw. <https://github.com/jeffalstott/powerlaw>.
- [py ] Python. <https://www.python.org/>.
- [air ] Rede air traffic control. <http://konect.uni-koblenz.de/networks/maayan-faa>.
- [eur ] Rede euroroad. [http://konect.uni-koblenz.de/networks/subelj\\_euroroad](http://konect.uni-koblenz.de/networks/subelj_euroroad).
- [ham ] Rede hamster friendships. <http://konect.uni-koblenz.de/networks/petster-friendships-hamster>.
- [pro ] Rede human protein (figeys). <http://konect.uni-koblenz.de/networks/maayan-figeys>.
- [mis ] Rede les misérables. [http://konect.uni-koblenz.de/networks/moreno\\_lesmis](http://konect.uni-koblenz.de/networks/moreno_lesmis).
- [gri ] Rede us power grid. <http://konect.uni-koblenz.de/networks/opsahl-powergrid>.
- [sci ] Scipy. <https://www.scipy.org/>.
- [BARABASI 1999] BARABASI, A.; ALBERT, R. (1999). Emergence of scaling in random networks. *Revista Science Volume 286*.

[SHANNON 1948] SHANNON, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*.

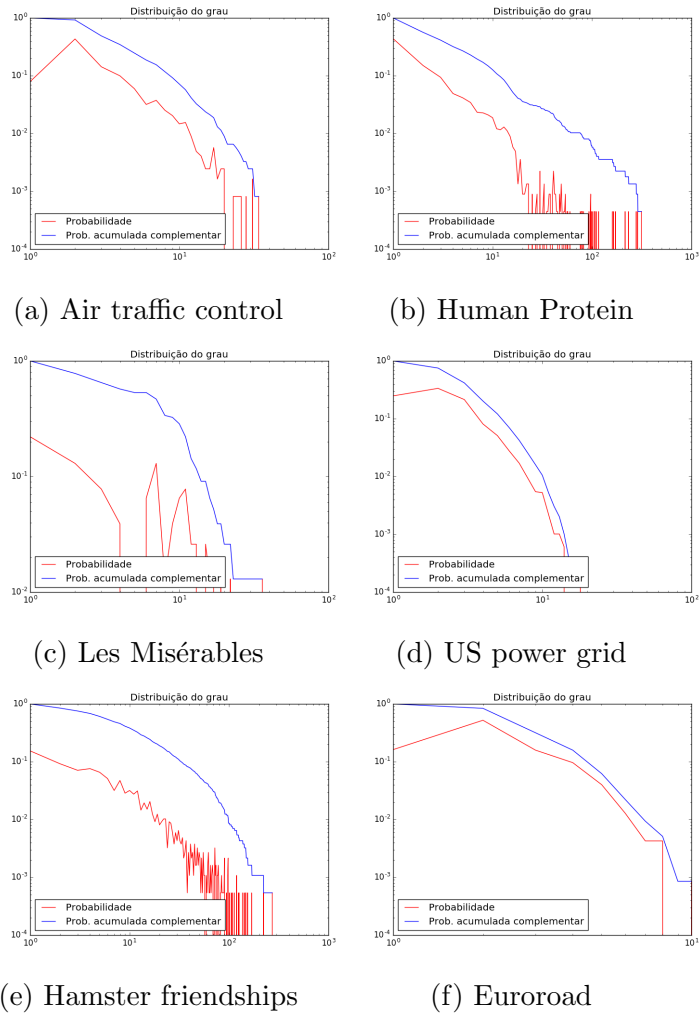


Figura 1: Distribuição de probabilidade do grau e distribuição acumulada complementar

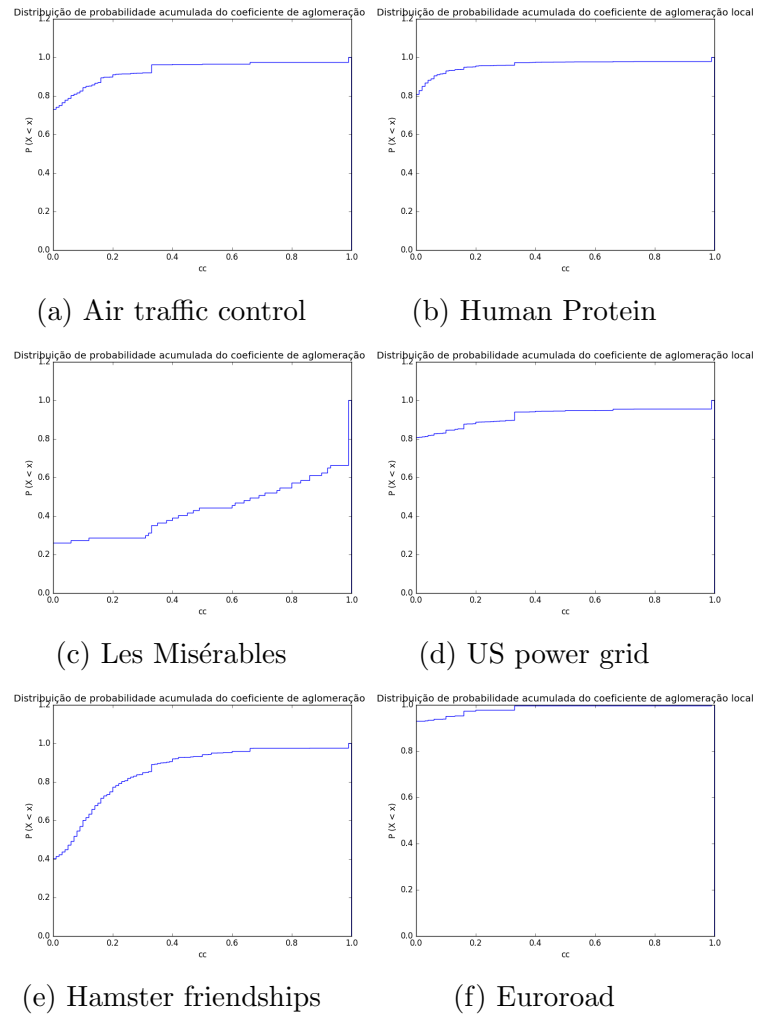


Figura 2: Distribuição acumulada do coeficiente de aglomeração

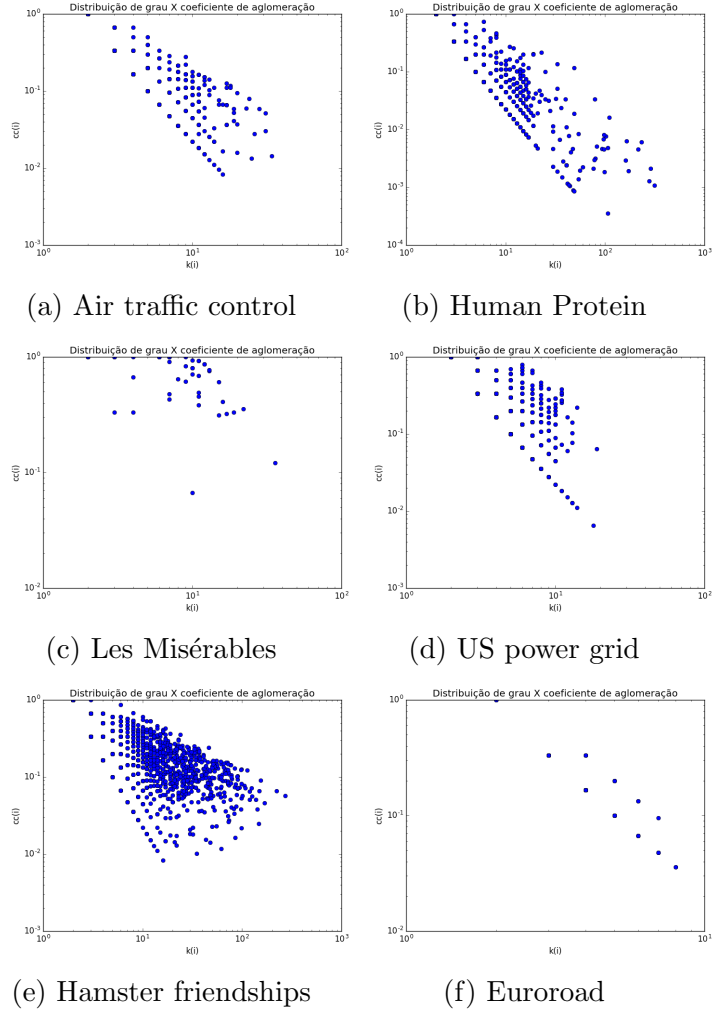


Figura 3: Grau x Aglomeração local de cada rede