
Projeto 2 - Centralidade e correlação do grau em redes complexas

Aluno:

Gil Barbosa Reis Nº 8532248

Professor:

Francisco Aparecido Rodrigues

INSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO
UNIVERSIDADE DE SÃO PAULO

São Carlos, SP
20 de novembro de 2016

Sumário

1	Introdução	3
2	Redes utilizadas	3
3	Métodos	3
3.1	Correlação dos graus	3
3.1.1	$Knn(i)$ e $Knn(k)$	3
3.1.2	Assortatividade	4
3.2	Centralidade	4
3.2.1	Grau	4
3.2.2	Closeness Centrality	4
3.2.3	Betweenness Centrality	4
3.2.4	Eigenvector Centrality	5
3.2.5	Page Rank	5
3.3	Caminhada aleatória	5
4	Resultados	5
4.1	Correlação entre graus	5
4.2	Centralidade	6
4.3	Correlação entre as medidas de centralidade	9
4.4	Caminhadas aleatórias	11
5	Conclusão	12

Lista de Tabelas

1	Correlação entre graus para cada rede	6
---	---	---

Lista de Figuras

1	Correlação entre graus	6
2	Distribuição de probabilidade do Grau	7
3	Distribuição de probabilidade do Betweenness Centrality	7
4	Distribuição de probabilidade do Eigenvector Centrality	8
5	Distribuição de probabilidade do Closeness Centrality	8
6	Distribuição de probabilidade do Page Rank	9
7	Correlação entre Grau e Betweenness Centrality	10
8	Correlação entre Eigenvector Centrality e Page Rank	10
9	Correlação entre Grau e Caminhada aleatória	11
10	Correlação entre Eigenvector Centrality e Caminhada aleatória	12

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/> para efetuar as medidas e testes:

- Hamster friendships[ham]: Amizades entre usuários do website hamsterster.com.
- US airports[air]: Voos diretos entre aeroportos dos Estados Unidos em 2010.
- U. Rovira i Virgili[are]: Comunicação direta por email entre usuários na Universidade Rovira i Virgili no sul da Catalonia na Espanha.

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] e Matplotlib[plo].

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 Correlação dos graus

A correlação entre os graus relaciona a tendência de vértices com grau similar se conectarem.

3.1.1 $Knn(i)$ e $Knn(k)$

Para medir essa correlação para um vértice, pode-se utilizar a fórmula 1, que mede o grau médio dos vizinhos de i .

$$Knn(i) = \sum_j \frac{A_{ij} \cdot K_j}{K_i} \quad (1)$$

É possível realizar uma média dessa correlação para cada grau, a partir da equação 2, onde N_k é o número de vértices de grau k .

$$Knn(k) = \sum_{i|K_i=k} \frac{Knn(i)}{N_k} \quad (2)$$

3.1.2 Assortatividade

Outro meio de correlacionar os graus de uma rede é através do Coeficiente de Assortatividade r , calculado a partir das fórmulas 3 e 4.

$$\sum_{ij} e_{ij} = 1, \quad \sum_j e_{ij} = a_i, \quad \sum_i e_{ij} = b_j \quad (3)$$

$$r = \frac{\sum_i e_{ij} - \sum_i a_i \cdot b_i}{1 - \sum_i a_i \cdot b_i} \quad (4)$$

Se o valor de r for próximo de 1, a rede é dita assortativa, ou seja, nós de grau próximo tendem a se conectar entre si. Caso esse valor seja próximo de -1, a rede é dita desassortativa, ou seja, nós de grau maior tendem a conectar-se com nós de grau baixo e vice-versa. Quando o valor é mais próximo de 0, a rede é dita não assortativa, ou seja, não apresenta tendência entre a conexão dos vértices em relação a seu grau.

3.2 Centralidade

O nível de importância de um vértice dentro da rede pode ser medido pela sua centralidade, pois vértices mais centrais propagam uma informação de maneira mais eficiente.

3.2.1 Grau

Podemos considerar que um vértice altamente conectado é central.

$$K_i = \sum_j A_{ij} \quad (5)$$

O problema do uso dessa medida é que podem haver vértices altamente conectados na periferia da rede.

3.2.2 Closeness Centrality

Mede a centralidade baseado na distância mínima média entre um vértice e todos os outros vértices. Essa medida pode ser encontrada pela fórmula 6, onde N é o número de vértices da rede e d_{ij} é a menor distância entre os vértices i e j .

$$CC(i) = \frac{N}{\sum_j d_{ij}} \quad (6)$$

O problema do uso dessa medida é que quando há mais de um componente na rede, a medida $d_{ij} \rightarrow \infty$, fazendo com que $CC(i) = 0, \forall i$.

3.2.3 Betweenness Centrality

Mede a centralidade baseado na carga que cada vértice recebe, calculada a partir do número de menores caminhos que passam pelo vértice. A fórmula 7 demonstra como calcular o Betweenness, onde $u_{st}(i)$ é o número de menores caminhos entre s e t que passam por i .

$$B_i = \sum_s \sum_t u_{st}(i) \quad (7)$$

O problema dessa medida é que para o Betweenness ser alto, o caminho mínimo entre s e t deve ser exclusivo. Uma solução para isso é normalizar o número total de caminhos que passam entre cada par de vértices (g_{st}):

$$B_i = \frac{1}{N} \cdot \sum_s \sum_t \frac{u_{st}(i)}{g_{st}} \quad (8)$$

3.2.4 Eigenvector Centrality

Mede a centralidade levando em consideração se um vértice é importante, ou se ele está conectado a outro vértice importante.

Definindo a centralidade inicial de cada vértice $X_o(i) = 1$, podemos calcular a centralidade através da fórmula 9, que pode ser escrita em notação matricial (fórmulas 10 e 11).

$$X_{t+1}(i) = \sum_j A_{ij} \cdot X_t(i) \quad (9)$$

$$X_{t+1} = A \cdot X_t \quad (10)$$

$$\vec{X} = A \cdot \vec{X} \quad (11)$$

O Eigenvector Centrality é calculado pelo autovetor da matriz de adjacência A associado ao maior autovalor da mesma.

3.2.5 Page Rank

Esse algoritmo simula alguém navegando na internet clicando nos links de cada página, e redirecione pela barra de endereços ao cair em uma página sem links. Páginas com mais links, tanto de entrada como saída, são consideradas mais importantes.

O algoritmo Page Rank é uma variante do Eigenvector Centrality, onde a matriz P representa a matriz de probabilidades de ir para uma determinada página a partir de outra. A equação 12 demonstra como montar essa matriz, considerando uma rede não direcionada.

$$P_{ij} = \frac{A_{ij}}{K_i} \quad (12)$$

A importância de cada nó é calculado pelo autovetor da matriz P associado ao maior autovalor da mesma.

3.3 Caminhada aleatória

Foi implementada uma caminhada aleatória de 1000 passos entre vértices das redes, começando de cada vértice. Essa medida é usada como comparação para outras medidas de centralidade, para ver se ocorre na prática o que era esperado pelo cálculo das outras medidas.

4 Resultados

4.1 Correlação entre graus

A correlação entre os graus das redes foi medido através da Assortatividade e da comparação entre $K(i)$ e $Knn(k)$ para cada rede, apresentadas na tabela 1. A figura 1 mostra essa relação nas redes

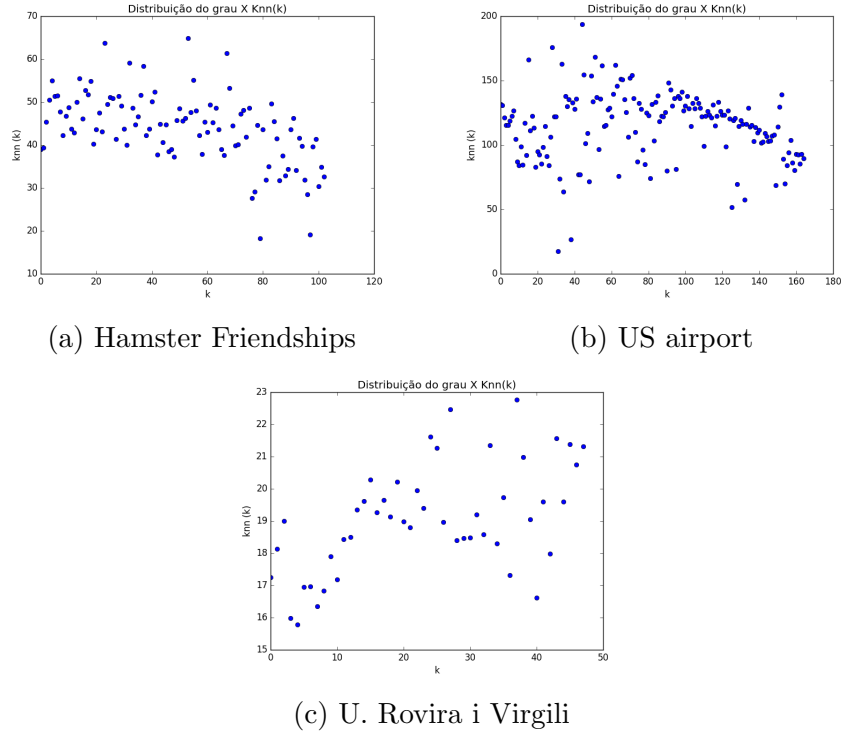


Figura 1: Correlação entre graus

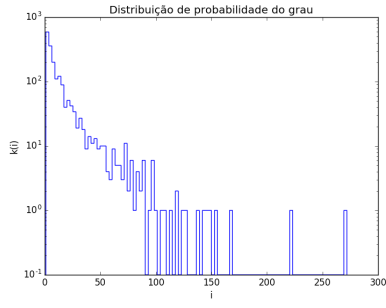
	Assortatividade	Pearson	Spearman
Hamster Friendships	-0.0847 - Não Assortativo	-0.5335	-0.5258
US airport	-0.1133 - Não Assortativo	-0.3609	-0.2848
U. Rovira i Virgili	0.0782 - Não Assortativo	0.5429	0.5226

Tabela 1: Correlação entre graus para cada rede

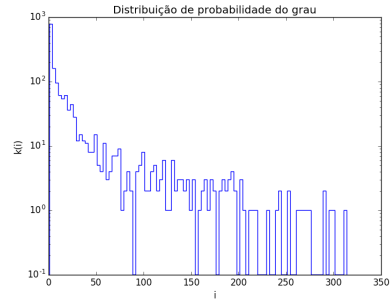
A partir dos resultados, percebe-se que as 3 redes são não assortativas, portanto não apresentam tendências entre conexão de vértices se observados os graus. O cálculo dos coeficientes de Pearson e Spearman para as redes Hamster Friendships e U. Rovira i Virgili indica uma leve tendência à de-assortatividade e assortatividade, respectivamente, mas não a ponto de contradizerem o coeficiente assortatividade.

4.2 Centralidade

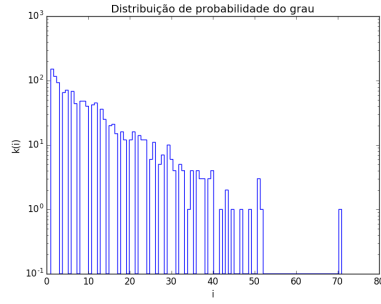
A centralidade dos nós das redes foram calculadas usando as medidas do Grau, Betweenness, Eigenvector, Closeness e Page Rank. As figuras de 2 a 6 mostram as distribuições de probabilidade de cada medida.



(a) Hamster Friendships

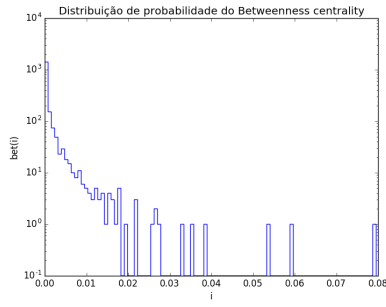


(b) US airport

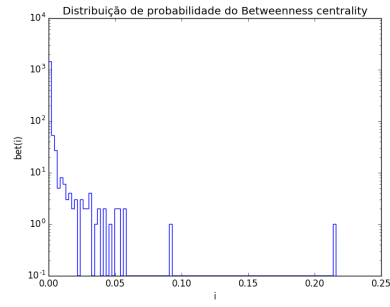


(c) U. Rovira i Virgili

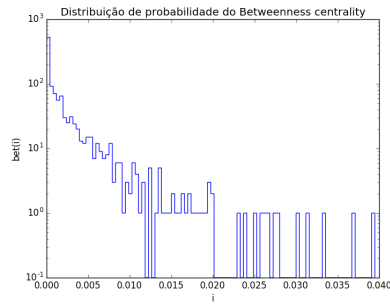
Figura 2: Distribuição de probabilidade do Grau



(a) Hamster Friendships

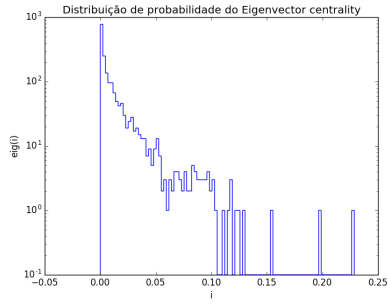


(b) US airport

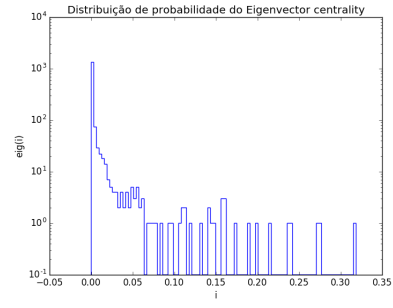


(c) U. Rovira i Virgili

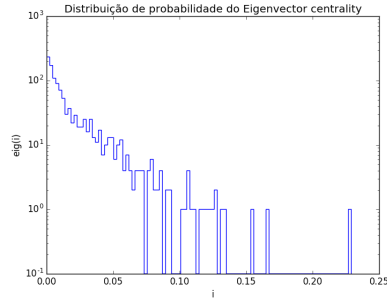
Figura 3: Distribuição de probabilidade do Betweenness Centrality



(a) Hamster Friendships

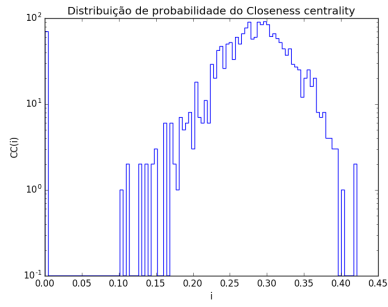


(b) US airport

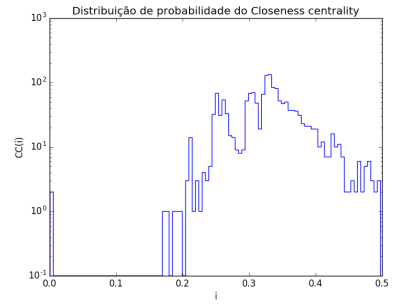


(c) U. Rovira i Virgili

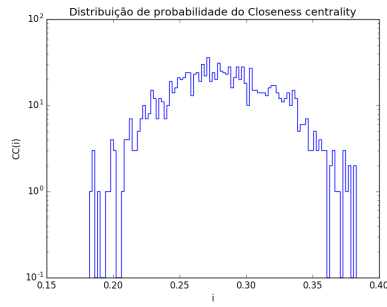
Figura 4: Distribuição de probabilidade do Eigenvector Centrality



(a) Hamster Friendships



(b) US airport



(c) U. Rovira i Virgili

Figura 5: Distribuição de probabilidade do Closeness Centrality

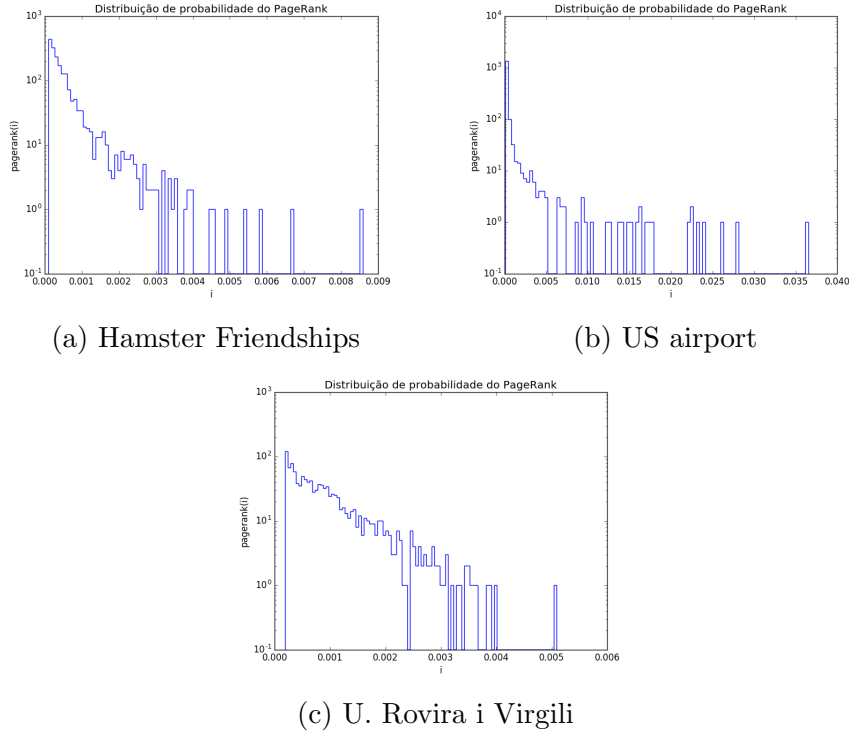
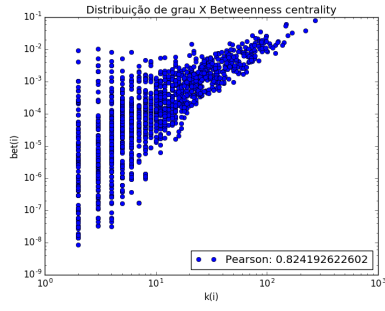


Figura 6: Distribuição de probabilidade do Page Rank

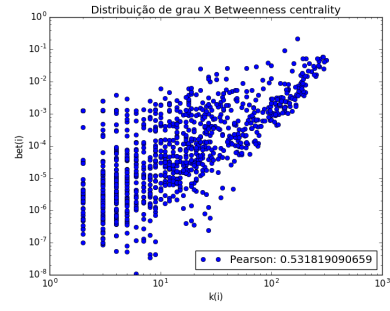
4.3 Correlação entre as medidas de centralidade

Foi verificada a correlação entre as medidas de centralidade encontradas. O coeficiente de Pearson foi encontrado para cada par de medidas. A figura 7 mostra a correlação entre Grau e Betweenness Centrality, e a figura 8 entre Eigenvector Centrality e Page Rank.

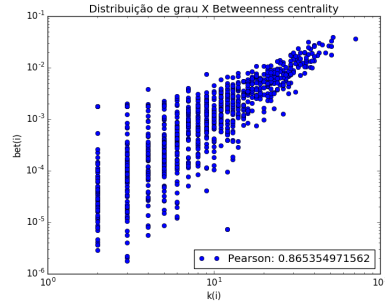
A partir do cálculo do coeficiente de Pearson entre as correlações citadas anteriormente, é possível notar que essas medidas, em geral, são condizentes.



(a) Hamster Friendships

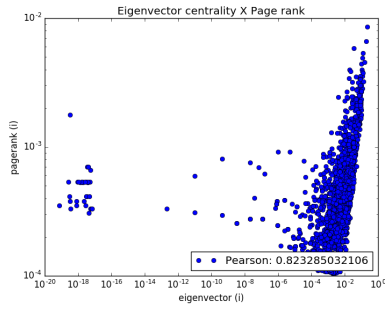


(b) US airport

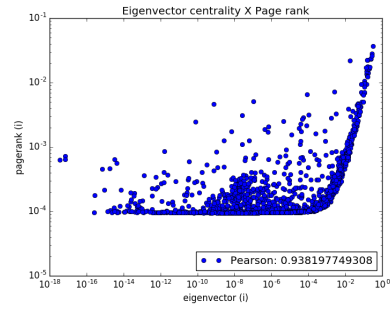


(c) U. Rovira i Virgili

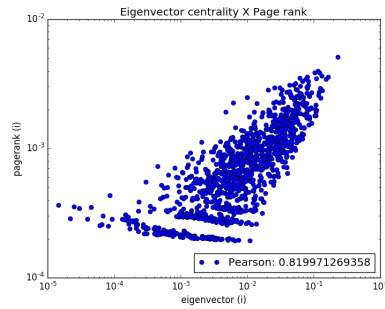
Figura 7: Correlação entre Grau e Betweenness Centrality



(a) Hamster Friendships



(b) US airport



(c) U. Rovira i Virgili

Figura 8: Correlação entre Eigenvector Centrality e Page Rank

4.4 Caminhadas aleatórias

As caminhadas aleatórias refletem o esperado pelas medidas de centralidade. O cálculo do coeficiente de Pearson entre os resultados da Caminhada aleatória e do Grau e Eigenvector centrality condizem entre si, dando suporte à teoria da centralidade dos vértices. As figuras 9 e 10 mostram essa comparação.

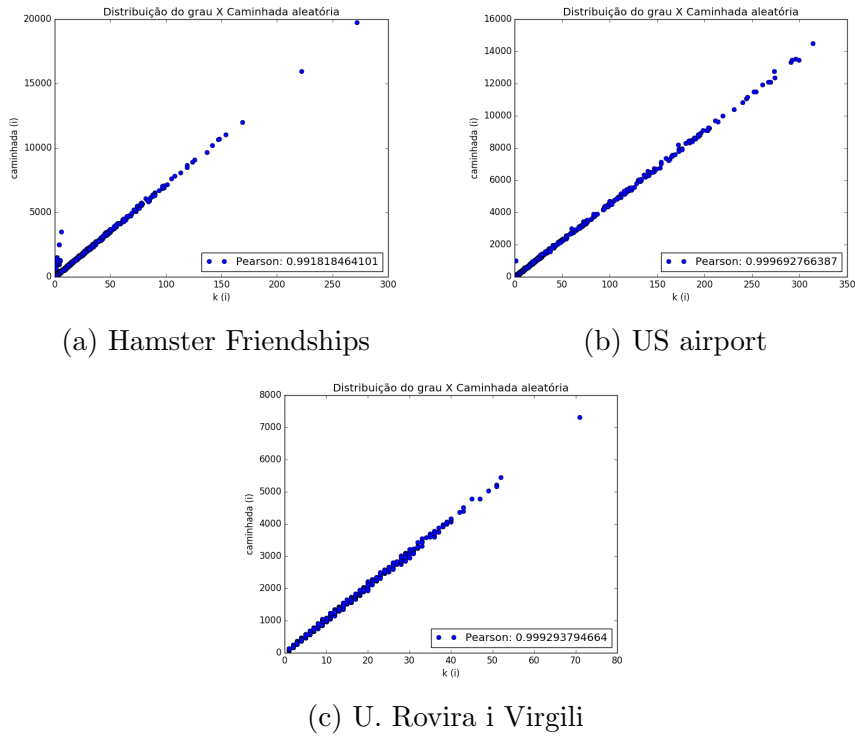


Figura 9: Correlação entre Grau e Caminhada aleatória

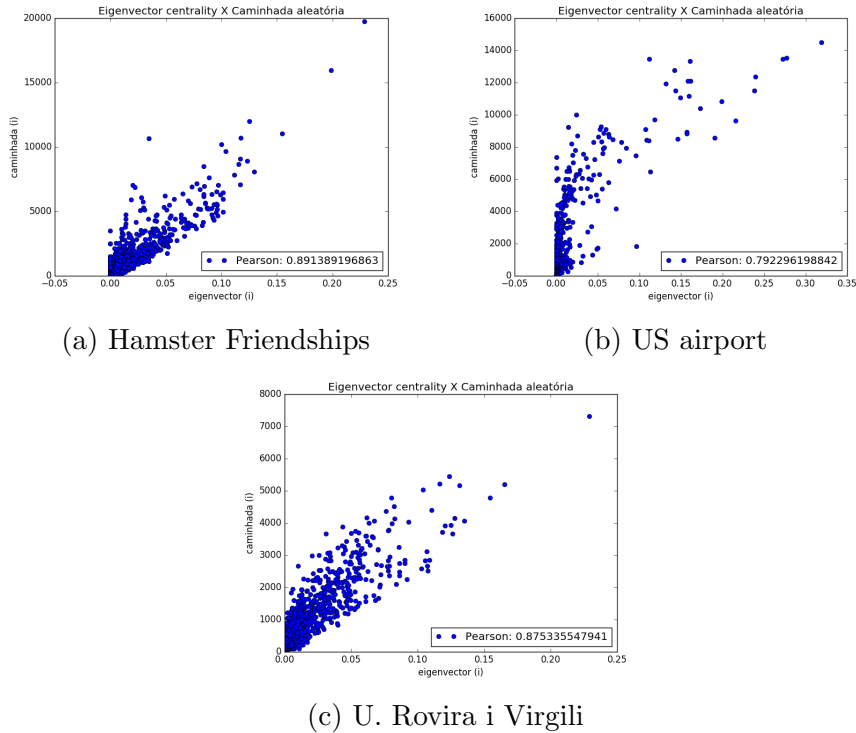


Figura 10: Correlação entre Eigenvector Centrality e Caminhada aleatória

5 Conclusão

Medidas de centralidade e correlação entre graus das redes foram tiradas e analisadas. Foi possível retificar alguns padrões de comportamento das redes, como por exemplo a proximidade entre as medidas de centralidade, e que vértices considerados mais centrais são mais visitados em caminhadas aleatórias.

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/>.
- [py] Python. <https://www.python.org/>.
- [ham] Rede hamster friendships. <http://konect.uni-koblenz.de/networks/petster-friendships-hamster>.
- [are] Rede u. rovira i virgili. <http://konect.uni-koblenz.de/networks/arenas-email>.
- [air] Rede us airports. <http://konect.uni-koblenz.de/networks/opsahl-usairport>.
- [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*.