2. Enron Email Network

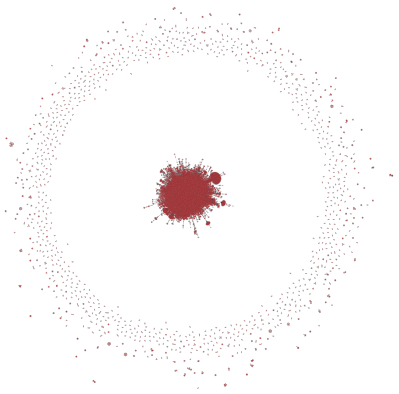

Figure 1: Grafo Enron Email

A rede de comunicação de e-mails da Enron dentro de um conjunto de dados de cerca de meio milhão de e-mails. Os vértices da rede são os endereços de e-mail e se o endereço i enviou pelo menos um e-mail para o endereço de e-mail j, forma-se uma aresta unidirecional entre os vértices i e j. O grafo contem 366692 vértices e 183831 arestas.

Métricas

Email	degrees	distance	clustering	betweenness edges	betweeness vertex	components	closeness
min	1	1	0	1,49E-09	0		0,114172
max	1383	13	1	0,016142	0,064851	33696	1
mean	10,020222	4,025143	0,496982	1,85E-05	6,95E-05	34,45258	0,307050
std	36,100004		0,002247	0,000073	0,000879	1.031,96	0,190206

- Clusterização Global : 0,08531080
- Quantidade de Componentes: 1065

Como pode ser visto na imagem do grafo (figura 1), o grafo não é fortemente conexo pois existem diversos vértices que não conseguem ser alcançados a partir de um vertice qualquer. Além disso, se observarmos os dados da métrica de grau, veremos que a distribuição de grau é bem discrepante, contendo muitos vértices com um grau baixo e poucos vértices com grau alto, mantendo assim a média baixa e a variação alta.

O grafo tem como maior distância entre dois vértices 13 arestas, e uma média de 4 arestas. Porém, vale a pena ressaltar que como o grafo não é fortemente

conexo, existem pequenos componentes que não podem ser alcançados, gerando assim uma distancia infinita que foi desconsiderada para a realização do calculo.

O gráfico (figura 2) pode-se avaliar a distribuição dos índices de clusterização e a sua probabilidade. Podem perceber que existem diversos vértices com índice próximos de 0, porém uma boa parte está em uma componente em que muitos vértices se relacionam entre si, aumentando assim a media do índice de clusterização.

O gráfico (figura 3) é a representação do índice de *betweeness* do vertice pela probabilidade. Podemos concluir que a maioria dos vértices não fazem parte do menor caminho entre outros dois e que existem um numero baixo de vértices que funcionam como vertice de ligação entre outros dois formando assim, essa distribuição desequilibrada.

Como comprovado pelos dados de clusterização e de *betweeness*, os dados da métrica de componentes afirmam q existência de uma componente com mais de 92 por cento dados vértices e diversas componentes que não estão conectadas a componente principal. Dessa forma, a media do tamanho dos componentes é baixa, entretanto, o desvio padrão é elevado.

Por último, o gráfico do índice *closeness* dos vértices pela sua probabilidade (figura 4). Nos indica a existem de uma componente conexa grande, visto que a maior parte dos vértices possui centralidade entre 0,2 e 0,4. Além disso, nos valores de 0,0 até 0,2, estão a pequena quantidade de vértices que não estão na componente maior, justificando o seu índice. No fim da curva encontra-se vértices com alto grau e pertencentes a componente maior, justificando assim, seu alto índice de *closeness*.

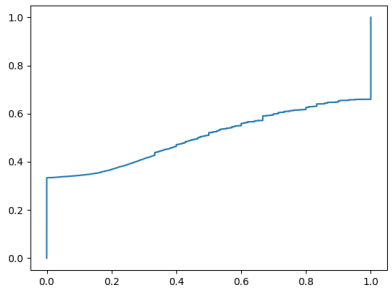

Figure 2: Clusterização x Probabilidade

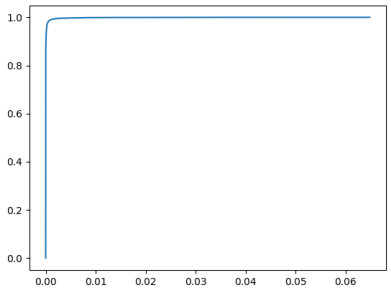

Figure 3: *Betweeness* x Probabilidade

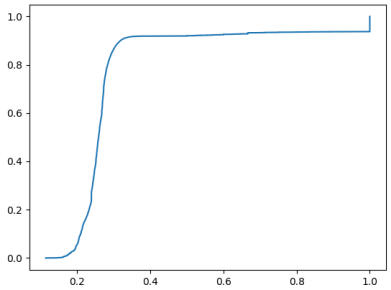

Figure 4: *Closeness* x Probabilidade