

UM ESTUDO DE ALGORITMOS PARA A DETECÇÃO DE COMUNIDADES EM REDES**Isabelle Alves**Escola Nacional de Ciências Estatísticas
Rua André Cavalcanti, 106 - Santa Tereza, Rio de Janeiro – RJ
belalvees@yahoo.com.br**Carla Silva Oliveira**Escola Nacional de Ciências Estatísticas
Rua André Cavalcanti, 106 - Santa Tereza, Rio de Janeiro - RJ
carla.oliveira@ibge.gov.br**José André de Moura Brito**Escola Nacional de Ciências Estatísticas
Rua André Cavalcanti, 106 - Santa Tereza, Rio de Janeiro - RJ
jambrito@gmail.com**RESUMO**

A tarefa de agrupar objetos semelhantes, no que diz respeito às suas características, é algo importante e que está presente diariamente em nosso cotidiano. Não só importante, mas também necessária, já que nos possibilita organização. Nesse sentido, podemos desenvolver essa atividade em uma rede, caracterizada por um conjunto n de objetos definidos em função de seus p atributos (características ou variáveis). Essa rede, por sua vez, possui estruturas que podem ser modeladas por um grafo, onde os vértices representam os terminais da rede, e as arestas o meio físico de comunicação entre eles. Quando objetos de uma rede têm significado que coincide com a realidade, ela é denominada complexa. Portanto, dada uma rede associada a um grafo não orientado, definimos uma comunidade como um conjunto de vértices que possuem propriedades comuns ou que desempenham funções similares dentro da rede. Considerando essa última observação, encontramos na sociedade diversos exemplos de comunidades representadas por redes complexa, quais sejam: as famílias, os grupos virtuais e as nações. Para avaliar e quantificar a semelhança entre os vértices da rede, utilizamos as medidas de similaridade que estão associadas a algum tipo de distância ou coeficiente, como por exemplo, a distância euclidiana ou o índice silhueta. Ou seja, o valor observado para a distância, ou para um coeficiente, quantifica o grau de similaridade ou dissimilaridade entre os vértices da rede. Após determinados esses coeficientes, usando uma função contida no pacote IGRAPH – pacote computacional do software livre R– foi possível detectar e analisar algumas comunidades associadas às redes representadas pelos respectivos grafos. Mais especificamente, nessas análises, observamos o número de comunidades, os vértices pertencentes a cada uma delas, a qualidade da partição e outras informações concernentes às redes. Posteriormente, essas redes foram submetidas a um algoritmo de agrupamento não hierárquico para determinação de agrupamentos. O presente algoritmo está disponível no pacote CLUSTERSIM do software R. Não obstante, neste último caso, foi necessário definir a priori o número de grupos k , estando esse associado ao número de comunidades encontradas anteriormente. E, de forma análoga ao que foi feito na detecção das comunidades, aplicamos índices sobre essa partição de forma a estabelecer uma comparação com as medidas obtidas para partições (comunidades produzidas pelo algoritmo de agrupamento) anteriores. Aplicamos as técnicas estudadas nas seguintes bases de dados: PIB dos 100 maiores municípios; IDH por município brasileiro; informações sobre recursos físicos na área de saúde pública e sistema de informação de atenção básica por municípios do Rio de Janeiro; além de dados dos alunos de graduação da ENCE, recolhidos através de um projeto para identificar o perfil socioeconômico do estudante de estatística na escola.

PALAVRAS - CHAVE: Grafos, Detecção de comunidades, Redes.