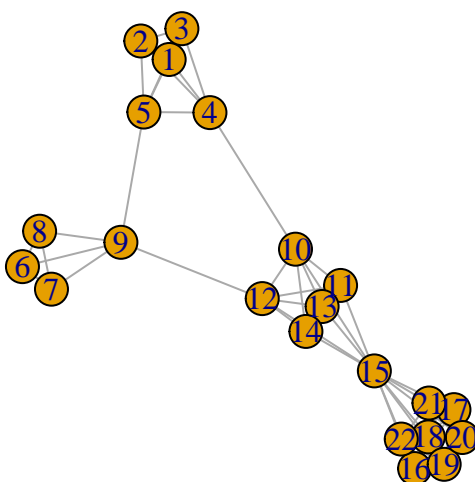


# Módulo 04 - Detecção de Comunidades

Outra aplicação muito interessante da Análise de Redes Sociais é a detecção de comunidades, que busca reunir vértices em grupos que possuam uma maior densidade de arestas entre os vértices de um mesmo grupo comparado com vértices de fora do grupo.

Observe o grafo abaixo:



É notório que existem quatro subgrupos mais densos dentro da rede, visto que, dentro de cada subgrupo (comunidade) desses, a densidade das conexões ou a quantidade de arestas entre os vértices é maior que entre os subgrupos. Assim, é possível atribuir, a cada vértice desse grafo, uma comunidade a qual ele pertence.

Existem diversos algoritmos de detecção de comunidades em grafos, muitos deles presentes na linguagem R graças ao pacote **igraph**. Cada algoritmo tem suas próprias particularidades, portanto cabe ao analista testar os diversos algoritmos e refletir sobre os resultados obtidos.

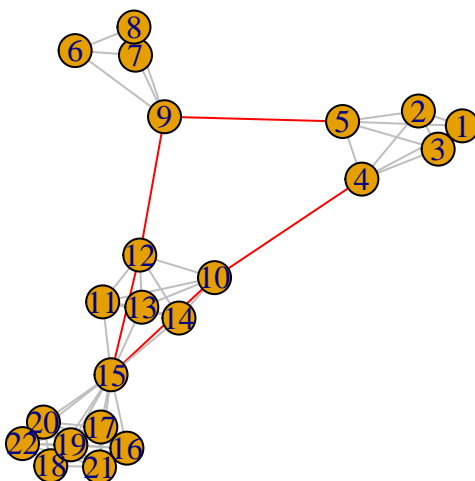
## Método edge-betweenness

Caracteriza-se por ser um algoritmo baseado em um método divisivo. Dada uma rede de interesse, a aresta que possui o maior valor da métrica de centralidade betweenness é removida. Ao fazer essa operação repetidamente, o grafo é dividido em subgrafos, ou seja, componentes menores. Após a remoção da aresta, calcula-se novamente a betweenness das arestas restantes.

A métrica betweenness foi escolhida pois favorece arestas que têm maior chance de conectar diferentes comunidades dentro de uma rede.

O gráfico abaixo destaca as 5 arestas com os maiores valores de betweenness.

## [1] 60 57 44 44 28



## Método Fast Greedy

O método edge-betweenness, conforme explicado, baseia-se na criação artificial de divisões dentro de uma rede. Mas e se não há subdivisões explícitas na rede? Para mensurar a qualidade dessas comunidades dentro de uma rede, pode-se usar o conceito de modularidade (modularity).

A modularidade de um grafo em relação a uma dada divisão em subgrafos mensura o quão significativa a divisão é. Conceitualmente, corresponde à fração de arestas que pertencem a uma comunidade menos a probabilidade de que uma aresta criada aleatoriamente pertenceria a uma comunidade.

Matematicamente, seja  $k$  o número de comunidades da rede,  $i$  uma comunidade,  $e_{ii}$  a fração de arestas na rede que pertençam à comunidade  $i$  - isto é, arestas que ligam vértices apenas de uma mesma comunidade - e  $a_i$  a fração de arestas que possuem pelo menos uma das pontas na comunidade  $i$ . A modularidade  $Q$  é definida como:

$$Q = \sum_{i=1}^k (e_{ii} - a_i^2)$$

Na equação acima,  $e_{ii}$  representa a probabilidade de que uma aresta pertence à comunidade  $i$  e  $a_i^2$  a probabilidade de que uma aresta aleatória pertença à comunidade  $i$ . Assim, quanto maior o valor da modularidade, maior a chance de que uma aresta pertença de fato à comunidade e, portanto, que a comunidade faça sentido dentro da rede. O valor da modularidade está definido no intervalo  $[-0,5, 1]$ .

A partir desse conceito de modularidade, aplica-se o algoritmo. A partir de um ponto inicial onde cada vértice da rede pertence a uma comunidade formada por apenas ele mesmo, o algoritmo cria comunidades de pares

de vértices, escolhendo em cada etapa a junção (adição de um vértice na comunidade) que resulta em um aumento de  $Q$ .

Referências <https://arxiv.org/pdf/cond-mat/0309508.pdf> <https://arxiv.org/pdf/cond-mat/0308217.pdf> <https://www.cs.cmu.edu/~ckingsf/bioinfo-lectures/modularity.pdf>