



O trabalho pode ser feito em grupos de até 04 alunos. A entrega deve ser realizada pelo Canvas até às 23:59 do dia 15/05/2022.

## Descrição do Problema

O **problema dos  $k$ -centros** representa uma tarefa clássica da análise de dados sendo sua solução relacionada às técnicas de *clustering*. Neste problema, dado um grafo completo com custos nas arestas (geralmente, respeitando a desigualdade triangular) e um inteiro positivo  $k$ , deseja-se encontrar um conjunto de  $k$  vértices (chamados centros) que minimize a maior distância de um vértice qualquer do grafo ao conjunto de centros (isto é, a maior distância de um vértice até um dos centros selecionados). Essa distância é, por vezes, denominada de **raio** da solução.

Dessa forma, essa tarefa lida com a questão de se particionar elementos em conjuntos (ou *clusters*) de acordo com suas (dis)similaridades (representadas pelas distâncias entre vértices). Trata-se, portanto, de um problema bastante genérico que aparece em diversas aplicações (além de existirem muitas variantes dele na literatura). A principal diferença entre essas variantes é a função objetivo (ou melhor, o critério utilizado para se determinar a solução ótima).

Uma definição formal do problema dos  $k$ -centros pode ser dada da seguinte forma:

**Problema dos  $k$ -centros:** Dada uma coleção de  $V$  pontos (ou vértices) com distâncias definidas para cada par de pontos pela função  $d : V \times V \mapsto \mathbb{R}^+$ , encontrar um conjunto  $C \subseteq V$ , com  $|C| \leq k$ , cujos elementos são chamados de centros, de forma que a maior distância de um ponto (ou vértice)  $v \in V$  ao centro mais próximo seja mínima.

Vale notar que nessa formulação, fixada uma instância  $(V, d, k)$ , uma escolha de centros  $C = \{c_1, c_2, \dots, c_p\} \subseteq V$  fornece implicitamente uma partição de  $V$  em  $p \leq k$  conjuntos  $C_1, C_2, \dots, C_p$  em que  $C_i = \{v \in V \mid d(v, c_i) \leq d(v, c_j), \forall j \neq i\}$  para  $i = 1, 2, \dots, p$ . Se para um dado ponto (ou vértice)  $v$  houver mais de um centro que atinge a distância mínima, escolhe-se um deles arbitrariamente para se determinar o *cluster* de  $v$ .

Uma possível interpretação desse problema seria a automatização de uma processo de divisão em categorias. Pode-se imaginar, como exemplo, o uso de dados de compras de consumidores. De acordo com os tipos e quantidades de produtos comprados, poderia ser elaborada uma maneira de se medir a “distância” entre dois consumidores. Assim, o problema dos  $k$ -centros pode ser visto como o problema de se estabelecer  $k$  “perfis de consumidor” de forma que os consumidores de um dado perfil (*cluster*) formem um grupo o mais homogêneo possível. Essa mesma ideia pode ser explorada para se categorizar alunos de um curso, usuários de uma rede social, deputados de uma casa legislativa, entre outros cenários, utilizando critérios de comparação que modelem mais convenientemente cada situação.

Outra interpretação é como uma tarefa de localização de facilidades. Os pontos (ou vértices) representam um conjunto de locais que precisam ser supridos por certo tipo de instalação e suas distâncias são definidas pela rota de menor custo de um ponto a outro – em que o custo pode ser a distância no espaço, o tempo, ou uma combinação desses e outros fatores. Deseja-se, assim, “abrir” tais instalações em até  $k$  pontos de maneira que nenhum ponto esteja muito longe da instalação mais próxima. Por exemplo, imagine que se tem um conjunto de hospitais públicos e que se dispõe de recursos para abrir centros para tratamento de uma doença (como, um câncer) em no máximo  $k$  deles. Deseja-se determinar em quais hospitais devem ser instalados tais centros de



modo a minimizar o maior tempo de transporte entre um hospital “comum” e seu centro de tratamento mais próximo, a fim de que a transferência mais remorada de um paciente seja tão rápida quanto possível.

A resolução do problema dos  $k$ -centros de forma exata pode ser difícil (ou mesmo, inviável para grandes instância), contudo ela pode ser feita por diferentes abordagens aproximadas.

## Descrição do Trabalho

Neste trabalho, cada grupo deverá implementar e comparar duas formas distintas de resolução deste problema: (i) uma exata (capaz de obter a solução mínima para instâncias pequenas); e (ii) outra aproximada (capaz de obter soluções “razoáveis” mesmo quando a instância for grande demais para o método exato).

A implementações deverão ser obrigatoriamente testadas com as instâncias disponíveis para **OR-Library**<sup>1</sup>. Mais especificamente, devem ser usadas as 40 instâncias para o problema (não capacitado) das  $p$ -medianas (que é bem semelhante ao problema dos  $k$ -centros exceto pela função objetivo)<sup>2</sup>.

Os arquivos contendo de dados devem ser processados<sup>3</sup> para se obter cada uma das instâncias de teste. A Tabela 1 apresenta para cada instância seu tamanho (em número de vértices, isto é,  $|V|$ ), o número máximo de centros a serem encontrados (isto é,  $k$ ) e o valor de raio da solução ótima<sup>4</sup>.

Além das implementações, você deverá realizar uma análise comparativa entre elas, visando determinar diferenças no desempenho (eficácia e eficiência) das mesmas para resolução do problema, principalmente na medida que o tamanho da instância aumentar.

## Descrição da Entrega

Todos os códigos implementados devem ser entregues, incluindo os códigos utilizados na preparação das instâncias.

Além dos códigos implementados, cada grupo deverá entregar um relatório (em formato PDF e também os fontes em TeX) descrevendo detalhes das implementações feitas, dos experimentos realizados e dos resultados obtidos.

---

<sup>1</sup><http://people.brunel.ac.uk/~mastjjb/jeb/info.html>

<sup>2</sup>Os arquivos com as 40 instâncias são denominados “pmed1.txt”, “pmed2.txt”, ..., “pmed40.txt” e estão disponíveis na seguinte URL: <http://people.brunel.ac.uk/~mastjjb/jeb/orlib/files>

<sup>3</sup>De acordo com a descrição detalhada para se criar uma instância do problema disponível em <http://people.brunel.ac.uk/~mastjjb/jeb/orlib/pmedinfo.html>

<sup>4</sup>Não devem ser utilizados os valores ótimos disponíveis no arquivo <http://people.brunel.ac.uk/~mastjjb/jeb/orlib/files/pmedopt.txt> pois representam a solução do problema de  $p$ -medianas.



Tabela 1: Lista de instâncias com seus tamanhos e valores de raio.

Instância	$ V $	$k$	Raio
01	100	5	127
02	100	10	98
03	100	10	93
04	100	20	74
05	100	33	48
06	200	5	84
07	200	10	64
08	200	20	55
09	200	40	37
10	200	67	20
11	300	5	59
12	300	10	51
13	300	30	35
14	300	60	26
15	300	100	18
16	400	5	47
17	400	10	39
18	400	40	28
19	400	80	18
20	400	133	13
21	500	5	40
22	500	10	38
23	500	50	22
24	500	100	15
25	500	167	11
26	600	5	38
27	600	10	32
28	600	60	18
29	600	120	13
30	600	200	9
31	700	5	30
32	700	10	29
33	700	70	15
34	700	140	11
35	800	5	30
36	800	10	27
37	800	80	15
38	900	5	29
39	900	10	23
40	900	90	13