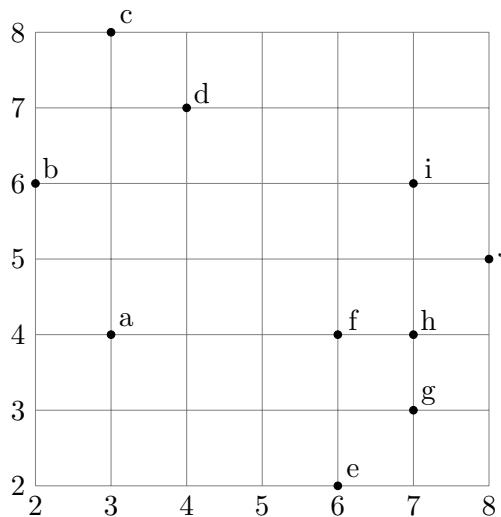


## Cvičení 11

**Příklad 1:** Pro níže uvedená data vytvořte shluky pomocí algoritmu k-means (k-průměrů) a PAM (k-medoidů) pro počet shluků k=2 a 3. Jako počáteční reprezentanty shluků vemte pro k=2 prvky  $r_1 = g$  a  $r_2 = h$  a pro k=3 prvky  $r_1 = g$ ,  $r_2 = h$  a  $r_3 = e$ . Určete nové reprezentanty shluků po první a druhé iteraci obou algoritmů pro k=2 a 3. Pro stejná data použijte algoritmus DBSCAN. Jaké budou parametry tohoto algoritmu, aby ve výsledku byly detekovány dva shluky?



**Příklad 2:** Ověrte kvalitu jednotlivých níže zadaných shluků a výsledného shlukování nad daty v příkladu 1 pomocí SSE, Silhouette indexu a Dunn indexu. Pokud můžete použít různé typy Minkowského metriky ( $k = 1, 2, \infty$ ) pro interní evaluaci, spočtete je.

- a)  $C_1 = \{a, b, c, d\}$ ,  $C_2 = \{e, f, g, h, i, j\}$ ,
- b)  $C_1 = \{a, b\}$ ,  $C_2 = \{c, d\}$ ,  $C_3 = \{e, f, g, h\}$ ,  $C_4 = \{i, j\}$ .

**Příklad 3:** Pomocí matice vzdáleností (pro vrcholy  $u$  a  $v$  v grafu je metrikou nejkratší cesta:  $d(u, v) = sp(u, v)$ ) nalezněte hierarchii shluků (komunit) v grafu (cvičení 10/příklad 6) vytvořeném nad daty v příkladu 1.

- a) Porovnejte metody single linkage, complete linkage a average linkage.
- b) Zamyslete se, zda můžete použít metodu PAM. Pokud ano, vytvořte 3 shluky.
- c) Lze pro určení komunit v grafu použít kliky? Vytvořte komunity pomocí algoritmu Clique-percolation, který ”překlápi” kliky velikosti  $k$  přes podkliku velikosti  $(k-1)$ , kde  $k \geq 3$ .
- d) Jaké vlastnosti budou mít nalezené komunity pomocí Clique-percolation? Budou nalezené komunity pokryvat všechny vrcholy?