Hierarchical Clustering Agreggation criteria

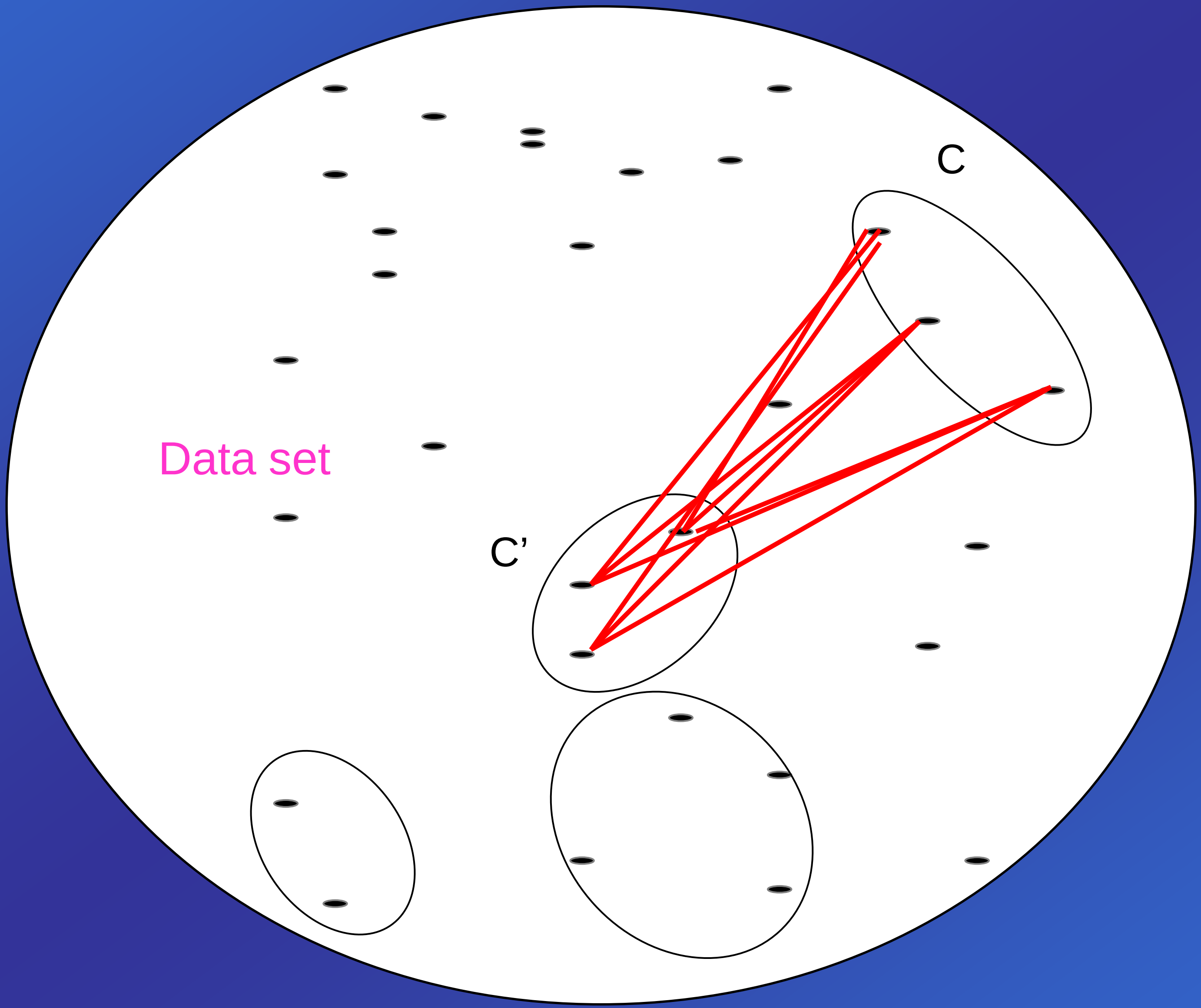
K. Gibert[1]

[1]Department of Statistics and Operation Research

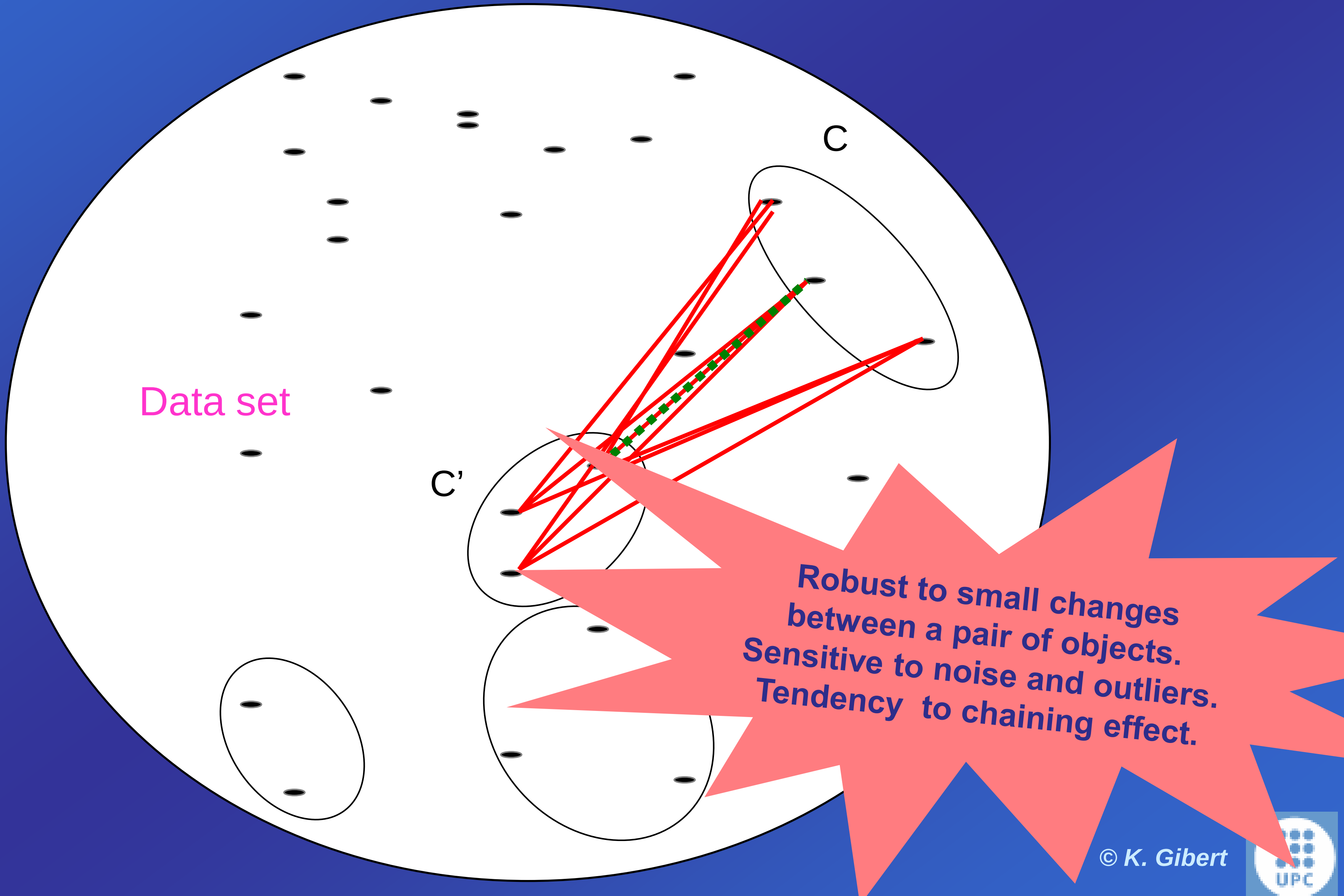
Knowledge Engineering and Machine Learning group
Universitat Politècnica de Catalunya, Barcelona

Aggregation Criteria

$d(C,C') = f(d(i,i')),\ i \in C,\ i' \in C'$



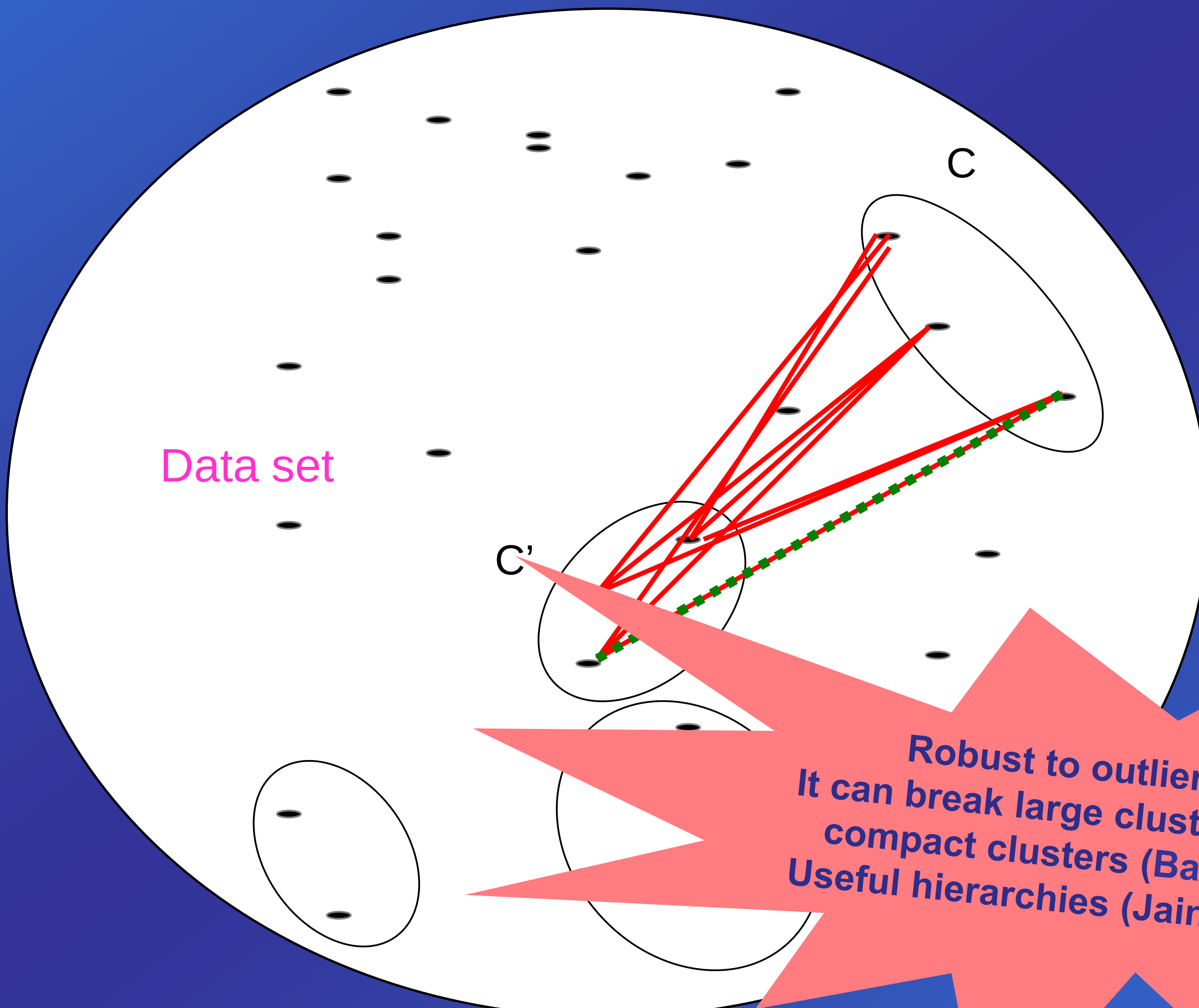
Single linkage (Sneath 1957) $d(C,C') = \min(d(i,i')), i \in C, i' \in C'$



Complete linkage

(Sorensen 1948)

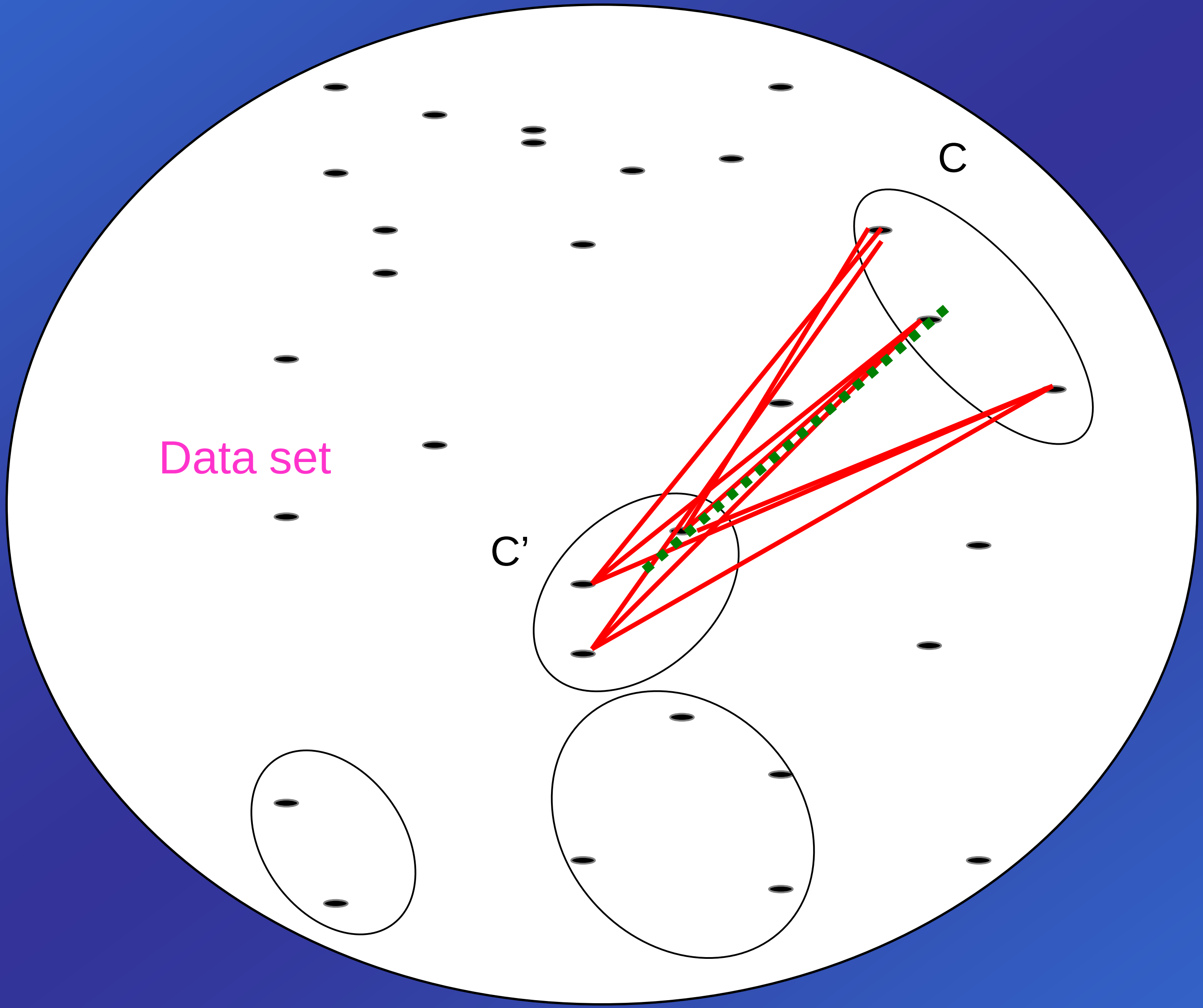
$d(C,C') = \max(d(i,i')),\ i \in C,\ i' \in C'$



Robust to outliers and noise.
It can break large clusters Conservative, compact clusters (Baeza-Yates 1992).
Useful hierarchies (Jain and Dubes 1998)

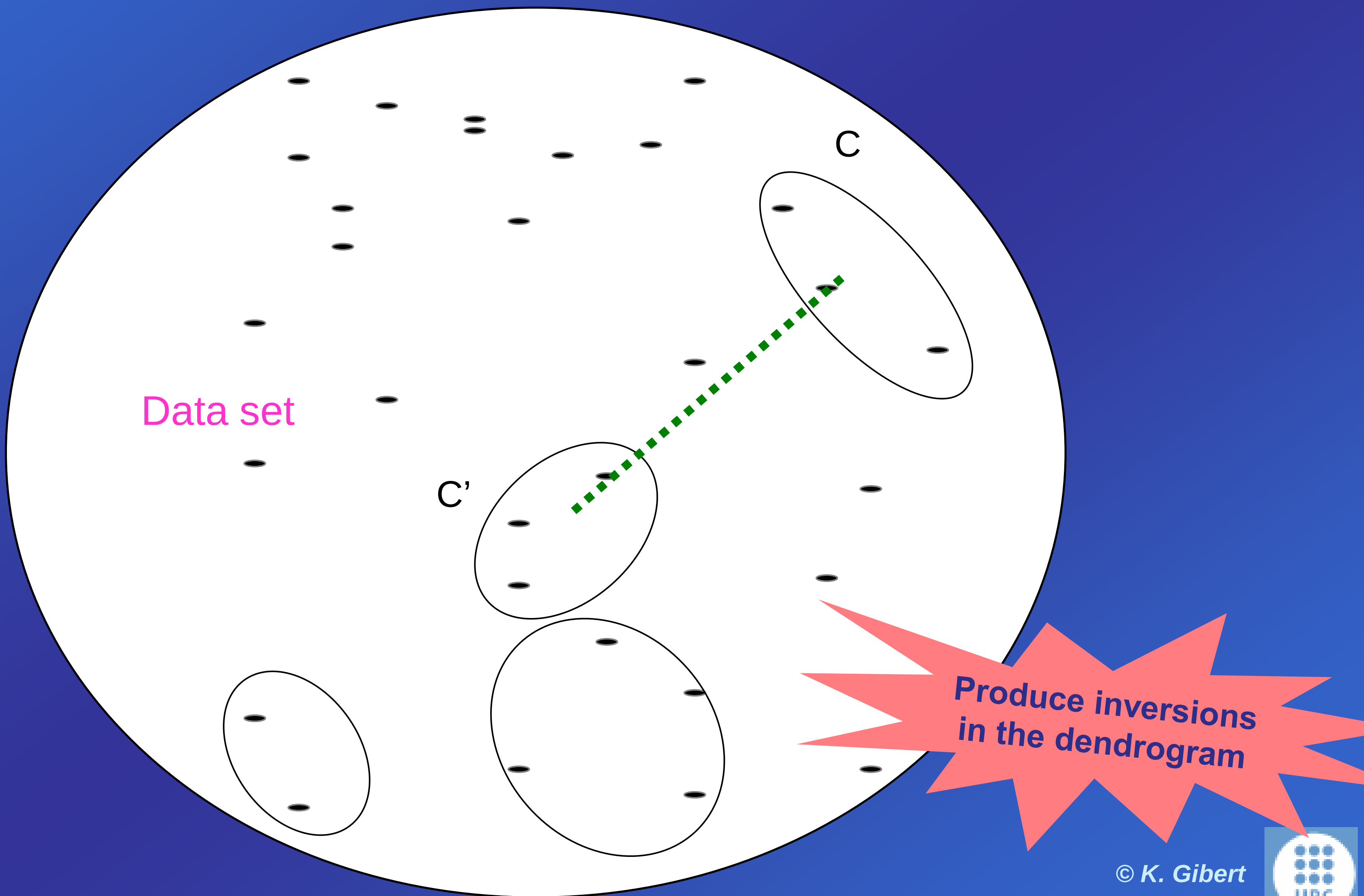
Average linkage

$d(C,C') = \text{mean}(d(i,i')),\ i \in C,\ i' \in C'$



Centroid

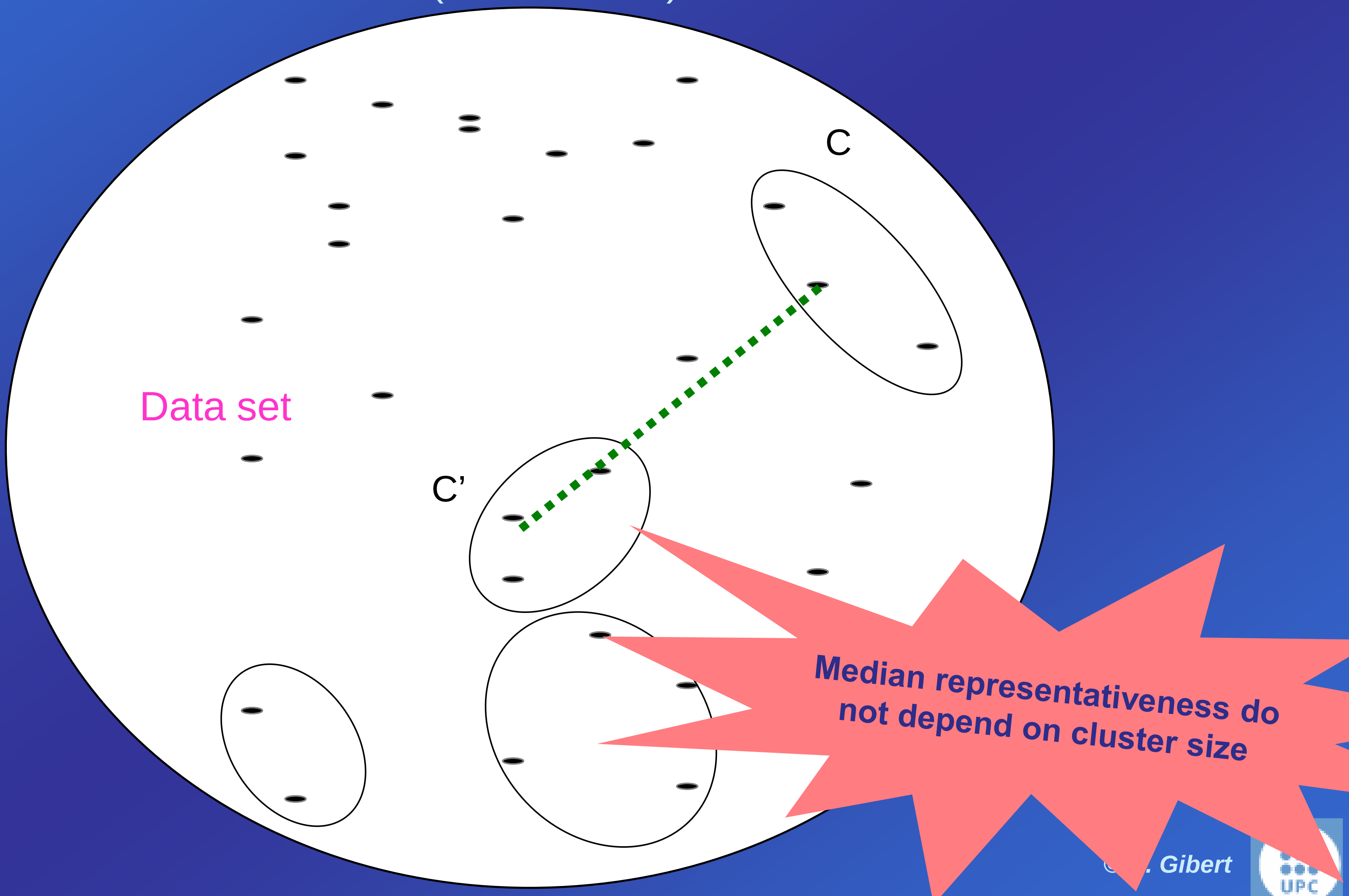
d(C,C’) = d(c,c’)), c,c’ centroids



Median linkage

(Gower 1967)

d(C,C') = d(c,c')), c,c' centroids



Ward's method (Ward 1963)

- Ascendant hierarchical method
- Group the two classes giving minimal inter-class inertia loss

- Inertia (physical concept) $$M^2(\mathcal{I}/\bar{\imath}) = \sum_{i \in \mathcal{I}} m_i\, d^2(i, \bar{\imath})$$

- Huygens theorem $$M^2(\mathcal{I}/\bar{\imath}) = Mi^2_{\mathcal{P}}(\mathcal{I}/\bar{\imath}) + Me^2_{\mathcal{P}}(\mathcal{I}/\bar{\imath}),$$

$$Mi^2_{\mathcal{P}}(\mathcal{I}/\bar{\imath}) = \sum_{\mathcal{C} \in \mathcal{P}} M^2(\mathcal{C}/\bar{\imath}_{\mathcal{C}}) = \sum_{\mathcal{C} \in \mathcal{P}} \sum_{i \in \mathcal{C}} m_i\, d^2(i, \bar{\imath}_{\mathcal{C}})$$

$$Me^2_{\mathcal{P}}(\mathcal{I}/\bar{\imath}) = \sum_{\mathcal{C} \in \mathcal{P}} m_{\mathcal{C}}\, d^2(\bar{\imath}_{\mathcal{C}}, \bar{\imath})$$

Minimize

- Every aggregation increases intra-class inertia with

$$\Delta_\xi = Mi^2_{\mathcal{P}_{\xi+1}}(\mathcal{I}/\bar{\imath}) - Mi^2_{\mathcal{P}_\xi}(\mathcal{I}/\bar{\imath}) = m_{\mathcal{C}_e} d^2(\mathcal{C}_e, \bar{\imath}_{\mathcal{C}}) + m_{\mathcal{C}_d} d^2(\mathcal{C}_d, \bar{\imath}_{\mathcal{C}})$$

- Inertia: related with quantity of information *(Thm Benzecri 1973)*
  The more informative, more inertia

Interpretable classes

- variability
  The more variable, more inertia

Can exagerate number of Exception-classes

- Quite popular