A Characterization of Linkage-Based Algorithms

David Loker
Joint work with
Margareta Ackerman and Shai Ben-David
Motivation

• There are a wide variety of clustering algorithms, which often produce very different clusterings.

• How should a user decide which algorithm to use for a given application?
Our approach

- Identify properties that distinguish between the results of different clustering paradigms
- The properties should be:
  1) Intuitive and “user-friendly”
  2) Useful for classifying clustering algorithms
- Clustering users can utilize prior knowledge to determine which properties make sense for their application
- Then use these properties to sort out clustering algorithms
Previous work

- Kleinberg proposes abstract properties ("Axioms") of clustering functions (NIPS, 2002)
- Bosagh Zadeh and Ben-David provide a set of properties that characterize single linkage clustering (UAI, 2009)
Our contributions

• Propose a couple of properties that uniquely indentify linkage-based clustering algorithms

• Construct a taxonomy of clustering algorithms based on the properties
Outline

• Define linkage-based clustering
• Our new clustering properties
• Main result
• Sketch of proof
• A taxonomy of common clustering algorithms using clustering properties
• Conclusions
For a finite domain set $X$, a *distance function* $d$ over the members of $X$.

A Clustering Function $F$ maps

**Input:** $(X,d)$ and $k>0$

to

**Output:** a $k$-partition (clustering) of $X$
Linkage-based algorithm: An informal definition

- Start with the clustering of singletons
- Merge the closest pair of clusters
- Repeat until only $k$ clusters remain.

Informal definition of between-cluster distance

"An extension of the between-point distance that applies to subsets of the domain"

- The definition of between-cluster distances is what distinguishes between linkage-based algorithms.
• Define linkage-based clustering

• Our new clustering properties
  • Main result
  • Sketch of proof

• A taxonomy of common clustering algorithms using our properties

• Conclusions
Hierarchical clustering

- A clustering $C$ is a *refinement* of clustering $C'$ if every cluster $c'$ in $C'$ is a union of some clusters in $C$.
- A clustering function is *hierarchical* if for every $k' \leq |X|,$

$$F(X,d,k')$$ is a refinement of $F(X,d,k).$
Locality

\[ F \text{ is local if for any } C \subseteq F(X, d, k), \]
\[ C = F(X' = \bigcup_{c \in C} c, d / X', \vert C \vert) \]
Which paradigms satisfy locality?

- Many clustering algorithms are local
  - K-means
  - K-median
  - Single-linkage
  - Average-linkage
  - Complete-linkage

- Notably, not all clustering algorithms satisfy locality
  - Ratio cut
  - Normalized cut
If \( d' \) equals \( d \), except for increasing between-cluster distances, then \( F(X,d,k) = F(X,d',k) \).
Many clustering algorithms satisfy outer-consistency:

- K-means
- K-median
- Single-linkage
- Average-linkage
- Complete-linkage
- Ratio cut
- Normalized cut
Outline

• Define linkage-based clustering
• Our new clustering properties
• Main result
• Sketch of proof
• A taxonomy of common clustering algorithms using our properties
• Conclusions
Our main result

Theorem:
An outer-consistent clustering function is linkage based iff it is hierarchical and local.
Any linkage based clustering function is hierarchical and local

The proof is quite straightforward.
Interesting direction of proof

If $F$ is outer-consistent, local, and hierarchical then $F$ is linkage-based.

To prove this direction we first need to formalize linkage-based clustering, by formally defining between-cluster distance.
What do we expect from the between-cluster distance?

1) Extends the point-wise distance: $\Delta: X^2 \rightarrow R^+$

2) The distance between subsets $A$ and $B$ is independent of data outside of these two clusters

3) Usually assumes no ties
Recall direction:
If $F$ is outer-consistent, local, and hierarchical then $F$ is linkage-based.

Goal:
Given $d$ over $X$, find a between-cluster distance $d : P(X)^2 \to \mathbb{R}^+$, such that the following algorithm outputs $F(X,d,k)$:
- Start with the clustering of singletons
- Merge the two clusters that minimize
- Repeat until $k$ clusters remain
Sketch of proof (continued…)

• For \( A, B, C, D \subseteq X \), \((A, B) < (C, D)\) if there exists a dataset such that when applying the clustering function to , \( A \) and \( B \) are merged before \( C \) and \( D \)

• Prove that \(<\) is an ordering (anti-symmetric and transitive)

• Use the ordering to construct...
Outline

• Define linkage-based clustering
• Our new clustering properties
• Main result
• Sketch of proof

• A taxonomy of common clustering algorithms using our properties
• Conclusions
## Taxonomy of clustering algorithms

<table>
<thead>
<tr>
<th></th>
<th>local</th>
<th>Outer consistent</th>
<th>Inner consistent</th>
<th>hierarchical</th>
<th>Order Invariant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single linkage</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Average linkage</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>Complete linkage</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>K-means</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>K-median</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Ratio-cut</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Normalized-cut</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
</tbody>
</table>
## Taxonomy of clustering algorithms

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>local</th>
<th>Outer consistent</th>
<th>Inner consistent</th>
<th>hierarchical</th>
<th>Order Invariant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single linkage</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Average linkage</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>Complete linkage</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>K-means</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>K-median</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Ratio-cut</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Normalized-cut</td>
<td>×</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
</tbody>
</table>
## Taxonomy of clustering algorithms

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>local</th>
<th>Outer consistent</th>
<th>Inner consistent</th>
<th>hierarchical</th>
<th>Order Invariant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single linkage</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Average linkage</td>
<td>✓</td>
<td>✓</td>
<td>❌</td>
<td>✓</td>
<td>❌</td>
</tr>
<tr>
<td>Complete linkage</td>
<td>✓</td>
<td>✓</td>
<td>❌</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>K-means</td>
<td>✓</td>
<td>✓</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>K-median</td>
<td>✓</td>
<td>✓</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>Ratio-cut</td>
<td>❌</td>
<td>✓</td>
<td>✓</td>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>Normalized-cut</td>
<td>❌</td>
<td>✓</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
</tr>
</tbody>
</table>
# Taxonomy of clustering algorithms

<table>
<thead>
<tr>
<th></th>
<th>local</th>
<th>Outer consistent</th>
<th>Inner consistent</th>
<th>hierarchical</th>
<th>Order Invariant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single linkage</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Average linkage</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>Complete linkage</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>K-means</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>K-median</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Ratio-cut</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Normalized-cut</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
</tbody>
</table>
## Taxonomy of clustering algorithms

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Local</th>
<th>Outer consistent</th>
<th>Inner consistent</th>
<th>Hierarchical</th>
<th>Order Invariant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single linkage</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Average linkage</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>Complete linkage</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>K-means</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>K-median</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Ratio-cut</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Normalized-cut</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
</tbody>
</table>
Conclusions

• We introduced new properties of clustering algorithms.

• We showed that an outer-consistent clustering algorithm is linkage-based iff it is hierarchical and local.

• We classified common clustering algorithms using these properties.
An extension operator is an injection $\bar{d} : P(X)^2 \rightarrow R^+$ such that for all $A, B$

$$\bar{d}(A, B) = d/(A, B)(A, B)$$

Defining cluster distance