Clustering - II

Prof. Dan A. Simovici

UMB
1 Hierarchies

2 Dendrograms
Definition

Let $S$ be a set. A hierarchy on the set $S$ is a collection of sets $\mathcal{H} \subseteq \mathcal{P}(S)$ that satisfies the following conditions:

1. the members of $\mathcal{H}$ are nonempty sets;
2. $S \in \mathcal{H}$;
3. for every $x \in S$, we have $\{x\} \in \mathcal{H}$;
4. if $H, H' \in \mathcal{H}$ and $H \cap H' \neq \emptyset$, then we have either $H \subseteq H'$ or $H' \subseteq H$. 
A standard technique for constructing a hierarchy on a set $S$ starts with a rooted tree $(\mathcal{T}, v_0)$ whose nodes are labeled by subsets of the set $S$. Let $V$ be the set of vertices of the tree $\mathcal{T}$. The function $\mu : V \rightarrow \mathcal{P}(S)$, which gives the label $\mu(v)$ of each node $v \in V$, is defined as follows:

1. the tree $\mathcal{T}$ has $|S|$ leaves, and each leaf $v$ is labeled by a distinct singleton $\mu(v) = \{x\}$ for $x \in S$;
2. if an interior vertex $v$ of the tree has the descendants $v_1, v_2, \ldots, v_n$, then $\mu(v) = \bigcup_{i=1}^{n} \mu(v_i)$. 
The set of labels $\mathcal{H}_T$ of the rooted tree $(\mathcal{T}, v_0)$ forms a hierarchy on $S$.

- Each singleton $\{x\}$ is a label of a leaf.
- Every vertex is labeled by the set of labels of the leaves that descend from that vertex.
- The root $v_0$ of the tree is labeled by $S$.

Suppose that $H, H'$ are labels of the nodes $u, v$ of $\mathcal{T}$, respectively. If $H \cap H' \neq \emptyset$, then the vertices $u, v$ have a common descendant. In a tree, this can take place only if $u$ is a descendant of $v$ or $v$ is a descendant of $u$; that is, only if $H \subseteq H'$, or $H' \subseteq H$, respectively.
Example

Let $S = \{s, t, u, v, w, x, y\}$ and let $T$ be a tree. It is easy to verify that the family of subsets of $S$ that label the nodes of $T$ is a hierarchy on the set $S$.

$$
H = \{\{s\}, \{t\}, \{u\}, \{v\}, \{w\}, \{x\}, \{y\}, \\
\{s, t, u\}, \{w, x\}, \{s, t, u, v\}, \{w, x, y\}, \{s, t, u, v, w, x, y\}\}
$$

Tree labeled by subsets of $S$. 
Chains of partitions defined on a set generate hierarchies, as we show next.

**Theorem**

Let $S$ be a set and let $C = (\pi_1, \pi_2, \ldots, \pi_n)$ be an increasing chain of partitions $(\text{PART}(S), \leq)$ such that $\pi_1 = \alpha_S$ and $\pi_n = \omega_S$. Then, the collection $\mathcal{H}_C = \bigcup_{i=1}^{n} \pi_i$ that consists of the blocks of all partitions in the chain is a hierarchy on $S$. 


Proof

The blocks of any of the partitions are nonempty sets, so $\mathcal{H}_C$ satisfies the first condition of the Definition on Slide 3.

We have $S \in \mathcal{H}_C$ because $S$ is the unique block of $\pi_n = \omega_S$. Also, since all singletons $\{x\}$ are blocks of $\alpha_S = \pi_1$, it follows that $\mathcal{H}_C$ satisfies the second and the third conditions of Definition on Slide 3.

Finally, let $H$ and $H'$ be two sets of $\mathcal{H}_C$ such that $H \cap H' \neq \emptyset$. It is clear that these two sets cannot be blocks of the same partition. Thus, there exist two partitions $\pi_i$ and $\pi_j$ in the chain such that $H \in \pi_i$ and $H' \in \pi_j$.

Suppose that $i < j$. Since every block of $\pi_j$ is a union of blocks of $\pi_i$, $H'$ is a union of blocks of $\pi_i$ and $H \cap H' \neq \emptyset$ means that $H$ is one of these blocks. Thus, $H \subseteq H'$. 
Theorem on Slide 7 can be stated in terms of chains of equivalences; we give the following alternative formulation for convenience.

**Theorem**

Let $S$ be a finite set and let $(\rho_1, \ldots, \rho_n)$ be a chain of equivalence relations on $S$ such that $\rho_1 = \iota_S$ and $\rho_n = \theta_S$. Then, the collection of blocks of the equivalence relations $\rho_r$ (that is, the set $\bigcup_{1 \leq r \leq n} S/\rho_r$) is a hierarchy on $S$. 
Define the relation “≺” on a hierarchy $\mathcal{H}$ on $S$ by $H \prec K$ if $H, K \in \mathcal{H}$, $H \subset K$, and there is no set $L \in \mathcal{H}$ such that $H \subset L \subset K$.

**Lemma**

Let $\mathcal{H}$ be a hierarchy on a finite set $S$ and let $L \in \mathcal{H}$. The collection $\mathcal{P}_L = \{ H \in \mathcal{H} \mid H \prec L \}$ is a partition of the set $L$.

**Proof:** We claim that $L = \bigcup \mathcal{P}_L$. Indeed, it is clear that $\bigcup \mathcal{P}_L \subseteq L$. Conversely, suppose that $z \in L$ but $z \notin \bigcup \mathcal{P}_L$. Since $\{z\} \in \mathcal{H}$ and there is no $K \in \mathcal{P}_L$ such that $z \in K$, it follows that $\{z\} \in \mathcal{P}_L$, which contradicts the assumption that $z \notin \bigcup \mathcal{P}_L$. This means that $L = \bigcup \mathcal{P}_L$.

Let $K_0, K_1 \in \mathcal{P}_L$ be two distinct sets. These sets are disjoint since otherwise we would have either $K_0 \subset K_1$ or $K_1 \subset K_0$, and this would contradict the definition of $\mathcal{P}_L$. 


Theorem

Let $\mathcal{H}$ be a hierarchy on a set $S$. The graph of the relation $\prec$ on $\mathcal{H}$ is a tree whose root is $S$; its leaves are the singletons $\{x\}$ for every $x \in S$.

Proof.

Since $\prec$ is an antisymmetric relation on $\mathcal{H}$, it is clear that the graph $(\mathcal{H}, \prec)$ is acyclic. Moreover, for each set $K \in \mathcal{H}$, there is a unique path that joins $K$ to $S$, so the graph is indeed a rooted tree.
Definition

Let $\mathcal{H}$ be a hierarchy on a set $S$. A grading function for $\mathcal{H}$ is a function $h : \mathcal{H} \rightarrow \mathbb{R}$ that satisfies the following conditions:

1. $h(\{x\}) = 0$ for every $x \in S$, and
2. if $H, K \in \mathcal{H}$ and $H \subset K$, then $h(H) < h(K)$.

If $h$ is a grading function for a hierarchy $\mathcal{H}$, the pair $(\mathcal{H}, h)$ is a graded hierarchy.
Example

For the hierarchy $\mathcal{H}$ defined in Example on Slide 6 on the set $S = \{s, t, u, v, w, x, y\}$, the function $h : \mathcal{H} \rightarrow \mathbb{R}$ given by

$h(\{s\}) = h(\{t\}) = h(\{u\}) = h(\{v\}) = h(\{w\}) = h(\{x\}) = h(\{y\}) = 0,$
$h(\{s, t, u\}) = 3, h(\{w, x\}) = 4, h(\{s, t, u, v\}) = 5, h(\{w, x, y\}) = 6,$
$h(\{s, t, u, v, w, x, y\}) = 7,$

is a grading function and the pair $(\mathcal{H}, h)$ is a graded hierarchy on $S$. 
Theorem

Let $S$ be a finite set and let $C = (\pi_1, \pi_2, \ldots, \pi_n)$ be an increasing chain of partitions $(\text{PART}(S), \subseteq)$ such that $\pi_1 = \alpha_S$ and $\pi_n = \omega_S$.

If $f : \{1, \ldots, n\} \rightarrow \mathbb{R}_{\geq 0}$ is a function such that $f(1) = 0$, then the function $h : \mathcal{H}_C \rightarrow \mathbb{R}_{\geq 0}$ given by $h(K) = f(\min\{j \mid K \in \pi_j\})$ is a grading function for the hierarchy $\mathcal{H}_C$.

Proof: Since $\{x\} \in \pi_1 = \alpha_S$, it follows that $h(\{x\}) = 0$.

Suppose that $H, K \in \mathcal{H}_C$ and $H \subset K$. If $\ell = \min\{j \mid H \in \pi_j\}$ it is impossible for $K$ to be a block of a partition that precedes $\pi_\ell$. Therefore, $\ell < \min\{j \mid K \in \pi_j\}$, so $h(H) < h(K)$, and $(\mathcal{H}_C, h)$ is indeed a graded hierarchy.
A graded hierarchy defines an ultrametric, as shown next.

**Theorem**

Let \((\mathcal{H}, h)\) be a graded hierarchy on a finite set \(S\). Define the function \(d : S^2 \rightarrow \mathbb{R}\) as

\[d(x, y) = \min\{h(U) \mid U \in \mathcal{H} \text{ and } \{x, y\} \subseteq U\}\]

for \(x, y \in S\). The mapping \(d\) is an ultrametric on \(S\).
Observe that for every $x, y \in S$ there exists a set $H \in \mathcal{H}$ such that \{x, y\} $\subseteq H$ because $S \in \mathcal{H}$.

It is immediate that $d(x, x) = 0$. Conversely, suppose that $d(x, y) = 0$. Then, there exists $H \in \mathcal{H}$ such that \{x, y\} $\subseteq H$ and $h(H) = 0$. If $x \neq y$, then \{x\} $\subset H$, hence $0 = h(\{x\}) < h(H)$, which contradicts the fact that $h(H) = 0$. Thus, $x = y$.

The symmetry of $d$ is immediate.

To prove the ultrametric inequality, let $x, y, z \in S$, and suppose that $d(x, y) = p$, $d(x, z) = q$, and $d(z, y) = r$. There exist $H, K, L \in \mathcal{H}$ such that \{x, y\} $\subseteq H$, $h(H) = p$, \{x, z\} $\subseteq K$, $h(K) = q$, and \{z, y\} $\subseteq L$, $h(L) = r$. Since $K \cap L \neq \emptyset$ (because both sets contain $z$), we have either $K \subseteq L$ or $L \subseteq K$, so $K \cup L$ equals either $K$ or $L$ and, in either case, $K \cup L \in \mathcal{H}$. Since \{x, y\} $\subseteq K \cup L$, it follows that

$$d(x, y) \leq h(K \cup L) = \max\{h(K), H(L)\} = \max\{d(x, z), d(z, y)\},$$

which is the ultrametric inequality.
Example

The values of the ultrametric generated by the graded hierarchy \((\mathcal{H}, h)\) on the set \(S\) introduced in Example given on Slide 13 are given in the following table:

<table>
<thead>
<tr>
<th>(d)</th>
<th>(s)</th>
<th>(t)</th>
<th>(u)</th>
<th>(v)</th>
<th>(w)</th>
<th>(x)</th>
<th>(y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(s)</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>(t)</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>(u)</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>(v)</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>(w)</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>0</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>(x)</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>4</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>(y)</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>0</td>
</tr>
</tbody>
</table>
Theorem

Let \((S, d)\) be a finite ultrametric space. There exists a graded hierarchy \((\mathcal{H}, h)\) on \(S\) such that \(d\) is the ultrametric associated to \((\mathcal{H}, h)\).
Proof

Let $\mathcal{H}$ be the collection of equivalence classes of the equivalences $\eta_r = \{(x, y) \in S^2 \mid d(x, y) \leq r\}$ defined by the ultrametric $d$ on the finite set $S$, where the index $r$ takes its values in the range $R_d$ of the ultrametric $d$. Define $h(E) = \min\{r \in R_d \mid E \in S/\eta_r\}$ for every equivalence class $E$. It is clear that $h(\{x\}) = 0$ because $\{x\}$ is an $\eta_0$-equivalence class for every $x \in S$.

Let $[x]_t$ be the equivalence class of $x$ relative to the equivalence $\eta_t$. Suppose that $E$ and $E'$ belong to the hierarchy and $E \subset E'$. We have $E = [x]_r$ and $E' = [x]_s$ for some $x \in X$. Since $E$ is strictly included in $E'$, there exists $z \in E' - E$ such that $d(x, z) \leq s$ and $d(x, z) > r$. This implies $r < s$. Therefore,

$$h(E) = \min\{r \in R_d \mid E \in S/\eta_r\} \leq \min\{s \in R_d \mid E' \in S/\eta_s\} = h(E'),$$

which proves that $(\mathcal{H}, h)$ is a graded hierarchy.
The ultrametric $e$ generated by the graded hierarchy $(\mathcal{H}, h)$ is given by

$$e(x, y) = \min\{h(B) \mid B \in \mathcal{H} \text{ and } \{x, y\} \subseteq B\}$$

$$= \min\{r \mid (x, y) \in \eta_r\} = \min\{r \mid d(x, y) \leq r\} = d(x, y),$$

for $x, y \in S$; in other words, we have $e = d$. 
Example

Starting from the ultrametric on the set $S = \{s, t, u, v, w, x, y\}$ defined by the table given in Example on Slide 17, we obtain the following quotient sets:

<table>
<thead>
<tr>
<th>Values of $r$</th>
<th>$S/\eta_r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$[0, 3)$</td>
<td>${s}, {t}, {u}, {v}, {w}, {x}, {y}$</td>
</tr>
<tr>
<td>$[3, 4)$</td>
<td>${s, t, u}, {v}, {w}, {x}, {y}$</td>
</tr>
<tr>
<td>$[4, 5)$</td>
<td>${s, t, u}, {v}, {w, x}, {y}$</td>
</tr>
<tr>
<td>$[5, 6)$</td>
<td>${s, t, u, v}, {w, x}, {y}$</td>
</tr>
<tr>
<td>$[6, 7)$</td>
<td>${s, t, u, v}, {w, x, y}$</td>
</tr>
<tr>
<td>$[7, \infty)$</td>
<td>${s, t, u, v, w, x, y}$</td>
</tr>
</tbody>
</table>
We shall draw the tree of a graded hierarchy \((\mathcal{H}, h)\) using a special representation known as a *dendrogram*. In a dendrogram, an interior vertex \(K\) of the tree is represented by a horizontal line drawn at the height \(h(K)\). For example, the dendrogram of the graded hierarchy of Example given on Slide 13 is shown next.

As we saw, the value \(d(x, y)\) of the ultrametric \(d\) generated by a hierarchy \(\mathcal{H}\) is the smallest height of a set of a hierarchy that contains both \(x\) and \(y\). This allows us to “read” the value of the ultrametric generated by \(\mathcal{H}\) directly from the dendrogram of the hierarchy.
Dendrogram of graded hierarchy of Example given on Slide 13