# CSE 20
# DISCRETE MATH

Fall 2020

http://cseweb.ucsd.edu/classes/fa20/cse20-a/

# Today's learning goals and updates

- Identify and prove properties of relations
- Evaluate whether a given relation is an appropriate model for a given application

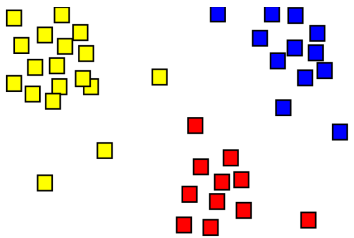
# Clustering

**Scenario**: Good morning! You're a user experience engineer at Netflix. A product goal is to design customized home pages for groups of users who have similar interests. Your manager tasks you with designing an algorithm for producing a clustering of users based on their movie interests, with the following constraints:

**Definition**: The set of movie ratings over $n$ movies is $R_n$, where each element of $R_n$ is a $n$-tuple with each entry in the tuple one of $\{-1, 0, 1\}$. The distance between two ratings is defined by $d$:

$$d(\ (x_1, \ldots, x_n), (y_1, \ldots, y_n)\ ) = \sum_{1 \leq i \leq n} |x_i - y_i|$$

$U = \{r_1, r_2, \cdots, r_t\}$ is a set of user ratings, and $U \subseteq R_5$. Assume that each user represented by an element of $U$ has a unique ratings tuple. A candidate clustering is $C_1, \cdots, C_m$ that is a **partition** of $U$: set of non-empty, disjoint subsets of $U$ whose union equals $U$. We compare candidate clusterings by computing a metric, e.g. min cluster density or average cluster density, where density relates number of ratings in a cluster with the maximum distance between them.



# Partitions from relations

# Partitions from relations

**Definition**: A binary relation $E$ on $U$ is an **equivalence relation** means it is reflexive, symmetric, and transitive.

$\forall x \in U($__________$)$ , $\forall x \in U\, \forall y \in U$ (________________________) , and $\forall x \in U\, \forall y \in U\, \forall z \in U$ (______________________________________)

An **equivalence class** of an element $x \in U$ for an equivalence relation $E$ on the set $U$ is the set

$$[x]_E = \{s \in U | (x, s) \in E\}$$

The set of equivalence classes of $E$ is $\{[x]_E \mid x \in U\}$.

A. $\{[x]_E \mid x \in U\} \in U$

B. $\{[x]_E \mid x \in U\} \subseteq U$

C. $\{[x]_E \mid x \in U\} \in \mathcal{P}(U)$

D. $\{[x]_E \mid x \in U\} \subseteq \mathcal{P}(U)$

E. None of the above.

# Partitions from relations

**Theorem**: Given an equivalence relation $E$ on set $U$, $\{[x]_E \mid x \in U\}$ is a partition of $U$.

Set of nonempty disjoint subsets of U whose union is U

# Partitions from relations

**Theorem**: Given an equivalence relation $E$ on set $U$, $\{[x]_E \mid x \in U\}$ is a partition of $U$.

- To show: For each $a \in U$, $[a]_E \neq \emptyset$, and for each $a \in U$, there is some $b \in U$ such that $a \in [b]_E$.
- To show: For each $a, b \in U$ , ( $(a,b) \in E$ ) $\rightarrow$ ( $[a]_E = [b]_E$ )
- To show: For each $a, b \in U$ , ( $(a,b) \notin E$ ) $\rightarrow$ ( $[a]_E \cap [b]_E = \emptyset$ )

# Partitions from relations

- To show: For each $a \in U$, $[a]_E \neq \emptyset$, and for each $a \in U$, there is some $b \in U$ such that $a \in [b]_E$.

# Partitions from relations

- To show: For each $a, b \in U$ , $((a,b) \in E) \to ([a]_E = [b]_E)$

# Partitions from relations

- To show: For each $a, b \in U$ , ( $(a, b) \notin E$ ) $\rightarrow$ ( $[a]_E \cap [b]_E = \emptyset$ )

# Relations on U

$$E_{proj} = \{(\ (x_1, x_2, x_3, x_4, x_5), (y_1, y_2, y_3, y_4, y_5)\ ) \in U \times U \ \mid\ (x_1 = y_1) \wedge (x_2 = y_2) \wedge (x_3 = y_3)\}$$

$$E_{dist} = \{(u, v) \in U \times U \ \mid\ d(u, v) \leq 2\}$$

$$E_{circ} = \{(u, v) \in U \times U \ \mid\ d((0,0,0,0,0), u) = d((0,0,0,0,0), v)\}$$

$$d(\ (x_1, \ldots, x_n), (y_1, \ldots, y_n)\ ) = \sum_{1 \leq i \leq n} |x_i - y_i|$$

A. $(\ (0,1,-1,0,1), (1,1,-1,0,0)\ ) \in E_{proj}$

B. $(\ (0,1,-1,0,1), (1,1,-1,0,0)\ ) \in E_{dist}$

C. $(\ (0,1,-1,0,1), (1,1,-1,0,0)\ ) \in E_{circ}$

D. More than one of the above.

E. None of the above.

# Relations on U

$$E_{proj} = \{(\ (x_1, x_2, x_3, x_4, x_5), (y_1, y_2, y_3, y_4, y_5)\ ) \in U \times U \ \mid \ (x_1 = y_1) \wedge (x_2 = y_2) \wedge (x_3 = y_3)\}$$
$$E_{dist} = \{(u, v) \in U \times U \ \mid \ d(u, v) \leq 2\}$$
$$E_{circ} = \{(u, v) \in U \times U \ \mid \ d((0,0,0,0,0), u) = d((0,0,0,0,0), v)\}$$

Which of these relations is **not** an equivalence relation?

A. $E_{proj}$

B. $E_{dist}$

C. $E_{circ}$

D. More than one of the above.

E. None of the above.

# Relations on U

$$E_{proj} = \{(\ (x_1, x_2, x_3, x_4, x_5), (y_1, y_2, y_3, y_4, y_5)\ ) \in U \times U \ \mid\ (x_1 = y_1) \land (x_2 = y_2) \land (x_3 = y_3)\}$$

$$E_{dist} = \{(u, v) \in U \times U \ \mid\ d(u, v) \leq 2\}$$

$$E_{circ} = \{(u, v) \in U \times U \ \mid\ d((0, 0, 0, 0, 0), u) = d((0, 0, 0, 0, 0), v)\}$$

The partition of $U$ defined by _____ is:

# Relations on U

$$E_{proj} = \{(\ (x_1, x_2, x_3, x_4, x_5), (y_1, y_2, y_3, y_4, y_5)\ ) \in U \times U \ \mid \ (x_1 = y_1) \wedge (x_2 = y_2) \wedge (x_3 = y_3)\}$$
$$E_{dist} = \{(u, v) \in U \times U \ \mid \ d(u, v) \leq 2\}$$
$$E_{circ} = \{(u, v) \in U \times U \ \mid \ d((0,0,0,0,0), u) = d((0,0,0,0,0), v)\}$$

The partition of $U$ defined by _____ is:

# Generating Clusters (Efficiently)

- CSE 150 series (AI & Machine Learning)
- https://en.wikipedia.org/wiki/K-means_clustering
- https://en.wikipedia.org/wiki/Hierarchical_clustering
- This is a big, active research area!

# Clustering