Fault tolerant subspace clustering for dealing with missing ratings

\[
\hat{r}(u, i) = \frac{\sum_{u' \in F_u} \text{sim}(u, u') \cdot r(u', i)}{\sum_{u' \in F_u} \text{sim}(u, u')}
\]

Recommendations for a user \( u \) are based on the ratings of his/her similar users (Friends \( F_u \)).

**NEAREST NEIGHBORS**

\( F_u = \{ u' \in U : \text{sim}(u, u') \geq \delta \} \)

- \( \delta \) is the user similarity threshold
- Friends are defined in the full dimensional feature space
- Linear scan of the DB to compute \( F_u \)

**EXISTING APPROACHES ON COMPUTING THE SET OF FRIENDS \( F_u \)**

- To speed up friends computation, users are grouped into clusters of similar users \( \Theta = \{ \Theta_1, \Theta_2, ..., \Theta_k \} \), \( \Theta_i \cap \Theta_j = \emptyset \).
  \( F_{u_\Theta} = \{ u' \in \Theta_i : u \in \Theta_i \} \)
  - Faster than nearest neighbors approach
  - For each \( u, u' \in \Theta_i \), \( \text{sim}(u, u') \geq \delta \) (correctness)
  - \( F_{u_\Theta} \subseteq F_u \) (incompleteness)
    - For small clusters, \( F_{u_\Theta} \) too narrow

**DIVERSIFICATION OF THE SET OF FRIENDS**

For both cases, user similarity is evaluated in the full (high) dimensional feature space
- Its difficult to find similar users when so many dimensions are considered
- Its more probable for users to exhibit similarity in some subspace of the feature space
- e.g., similar taste in comedies but not in dramas

**SUBSPACE CLUSTERING BASED RECOMMENDATIONS**

- Clusters are defined in subspaces of the original feature space: \( \Theta = \{ U_i, I_i \} \)
- Subspace clustering: \( \Theta = \{ \Theta_1, \Theta_2, ..., \Theta_k \} \)
  - \( \Theta_i \) might overlap w.r.t. both users and items
- Fault tolerant subspace clustering for dealing with missing ratings
  - Missing values are tolerated but bounded within a cluster per user \( \varepsilon_{i,j} \), per item \( \sigma_{i} \), and in total \( \delta_{i} \).

**FULL DIMENSIONAL CLUSTERING**

**WEIGHTED FULL DIMENSIONAL RANKING**

- Rank the users in \( F_{u_\Theta} \) based on their full dimensional distance to \( u \)
  - Weight by #shared dimensions
- Select the most prominent ones

The friends of a user \( u \) are the members of all subspace clusters where \( u \) belongs to:

\[
F_{\text{subclu}_u} = \{ u' \in \Theta_i : u \in \Theta_i \}
\]