Proposed Approach

1. a. Compute correlations of all pairs of predicates (all variabilizations)
b. Discard low correlation pairs

2. Find clusters of predicates that are highly correlated
   Express as weighted satisfiability problem
   Each pair of predicates is an atom and unit clause:
   \[ \text{Edge}(P_1, P_2) \text{ with weight } = \log(\text{correlation}(P_1, P_2)) - \text{thresh} \]
   Apply “soft” transitive closure:
   \[ \text{Edge}(P_1, P_2) \wedge \text{Edge}(P_2, P_3) \Rightarrow \text{Edge}(P_1, P_3) \] with weight \( v \)
   Higher \( v \) \( \Rightarrow \) larger clusters of predicates
   Use MaxWalkSat [Kautz et al. 1997] to solve sat. problem & select edges

3. a. Invent a predicate for each clique of predicates
    Arguments are (a subset of) the observed predicates’ arguments
   b. Model correlation among predicates in clique
    Associate a weight \( w_{ij} \) between invented predicate \( h_i \) and each of its observed predicate \( o_{ij} \)

4. a. Define a potential \( f_{ijk} \) between the \( k^{th} \) grounding of invented predicate \( h_i \) and each of its observed predicate \( o_{ij} \)
b. When the invented predicates are independent given the observables, we can sum them out and avoid using EM

5. a. Init weights \( w_{ij} \) to the average (log) correlation between \( o_{ij} \) and other observed predicates of \( h_i \)
b. Find locally optimal weights using gradient ascent

6. Iterate by treating the hidden predicates as observed predicates, and setting them to their MAP values