

Relational Neighbor Classifier

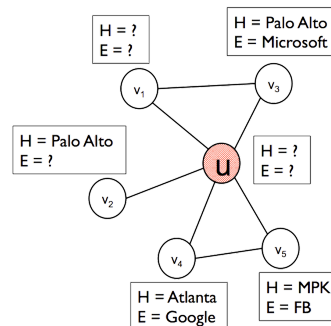
- Relational neighbor (RN) classifier: Simple relational probabilistic model that makes predictions for a given node based on class labels of neighbors
- Weighted-vote relational neighbor classifier (wvRN)
- Case study at Facebook: Inferring missing node attributes

Facebook: Profile Inference



Profile:

- Hometown: Palo Alto
- High school: ?
- College: ?
- Employer: Microsoft
- Current city: ?
- Hobbies, politics, music



- Use the social network.
- Assume homophily.
 - Friendships form between "similar" people.
 - Infer missing labels to maximize similarity.

Markov Assumption

- The label of one node depends on that of its immediate neighbors in the graph.
- Relational models are built based on the labels of neighbors.
- Predictions are made using collective inference.

A Relational Neighbor Classifier

Definition. The relational-neighbor classifier estimates $P(c|e)$, the class-membership probability of an entity e belonging to class c , as the (weighted) proportion of entities in D_e that belong to class c . We define D_e as the set of entities that are linked to e . Thus,

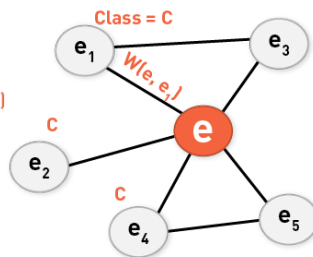
$$P(c|e) = \frac{1}{Z} \sum_{\{e_j \in D_e | \text{label}(e_j)=c\}} w(e, e_j), \quad (1)$$

where $Z = \sum_{e_i \in D_e} w(e, e_i)$, and $w(e, e_i)$ is the weight of the link² between entities e and e_i . Entities in D_e that are not of the same type as e are ignored. If D_e is empty or has no entities with known class labels, then the RN will estimate e based on the class prior (of the known labels).

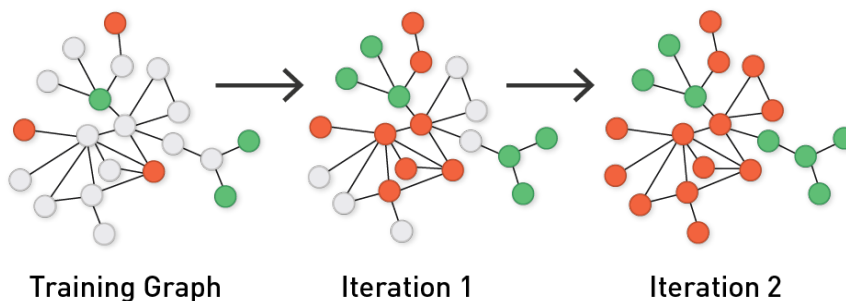
$$P(c|e) = 3/5$$

#3 neighbors with class C

(assume edge between neighbors of e is 1 for all edges)



Iterative Relational-Neighbor Classifier (RNâ€”â€”â€”)



- The relational neighbor (RN) classifier proposed by Macskassy and Provost (2003) is a simple relational probabilistic model that makes predictions for a given node based solely on the class labels of its neighbors.
- E.g., hometown (multinomial)

Weighted-Vote Relational Neighbor Classifier (wvRN)

Weighted-vote relational neighbor classifier estimates prediction probability as:

$$P(L_i=c|v_i) = \frac{1}{z} \sum_{v_j \in N_i} w(v_i, v_j) \times P(L_j=c|N_j)$$

\uparrow
 Networks studied use {0,1}
 to represent connections
 between actors

\uparrow
 Class
 probability of
 its neighbors

Here z is the usual normalization factor, and $w(v_i, v_j)$ is the weight of the link between node v_i and v_j .

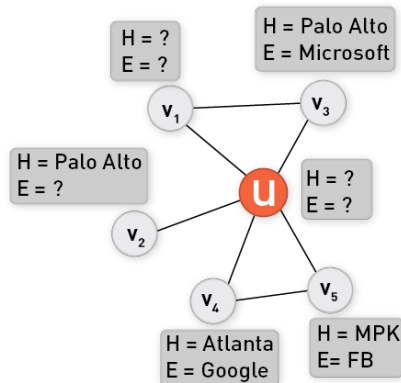
Mappers-Reducers: Join

NODE: Attributes; graph neighbors

V1: Hometown? Links: 3 and u (you)

V4: H= Atlanta; Links: v5 and you

Etc.



- Mapper
 - Send Hometown label and probability to its neighbors.
 - Send graph also.

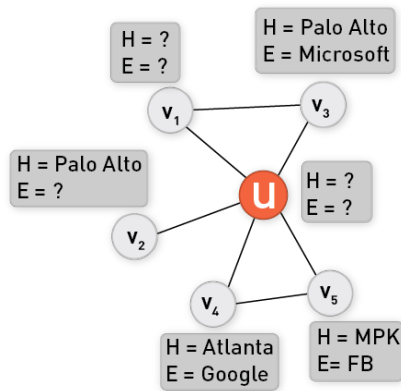
Mappers-Reducers: Join

NODE: Attributes; graph neighbors

V1: Hometown? Links: 3 and u (you)

V4: H= Atlanta; Links: v5 and you

Etc.



- Reducer
 - If a node (person) provided a label, keep as it is and yield original node.
 - Otherwise combine all neighbor candidate hometowns and select label with highest probability.
 - Yield a node with the same neighbors but with an updated hometown.