

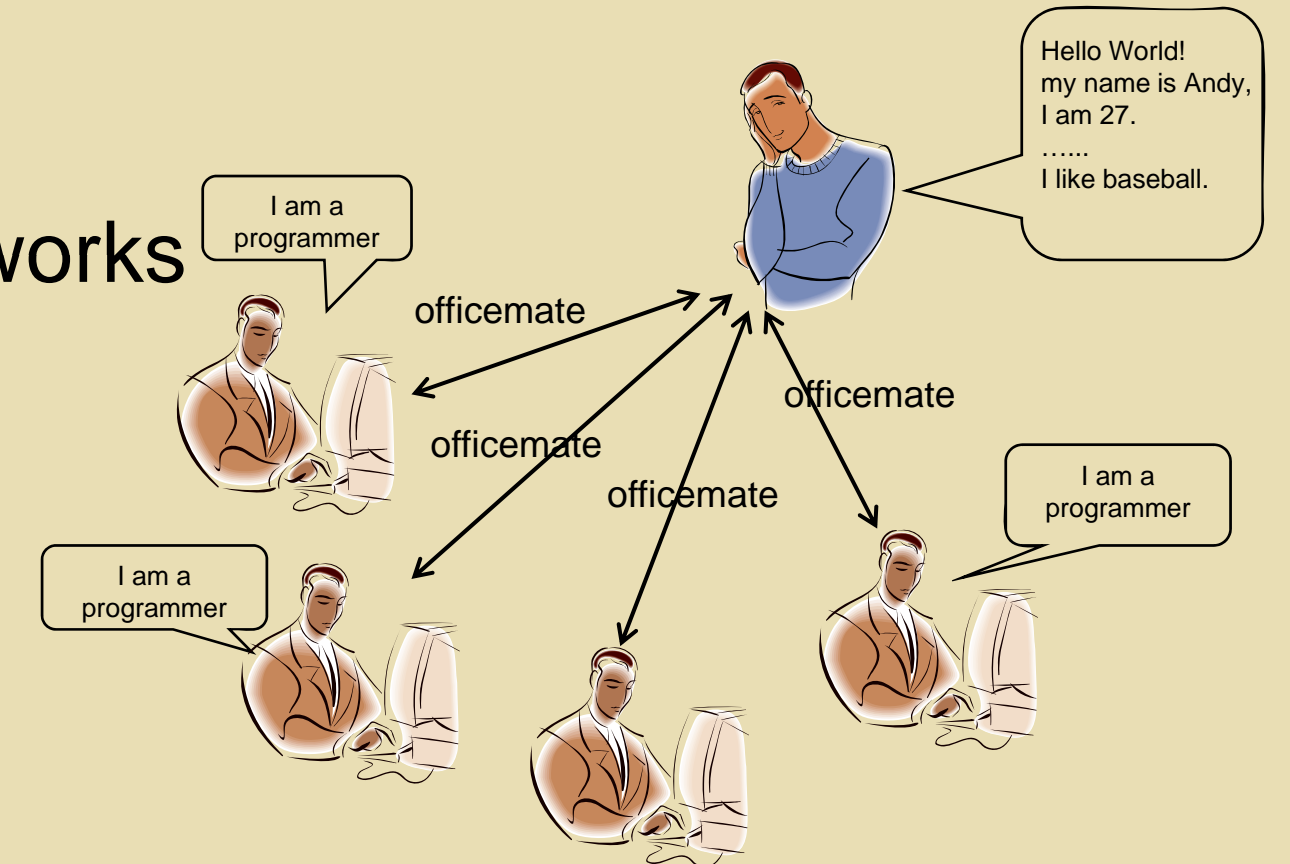
Protecting Private Information in Social Networks



Dr. Wesley W. Chu <http://www.cs.ucla.edu/~wwc>

Motivation

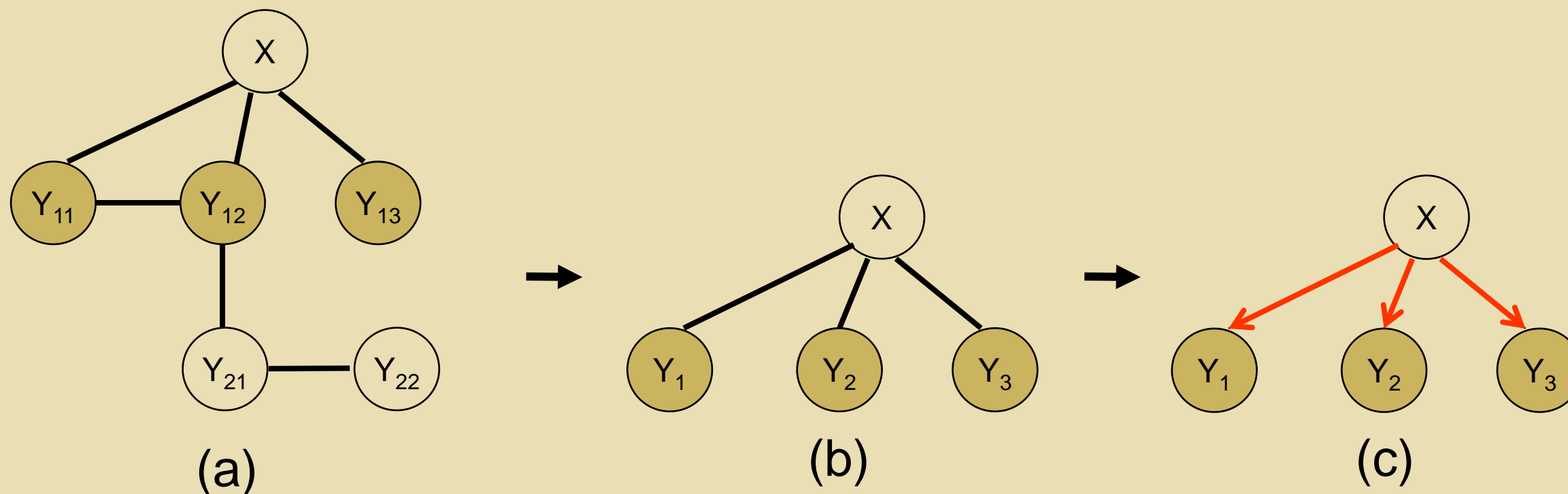
- Personal private information can be inferred from friends
- Privacy confidentiality becomes more challenging for online social networks
- Goal
 - Study private attribute inference in social networks
 - Propose schemes for privacy protection



Mapping Social Networks into Bayesian Networks

Bayesian network Construction

- Localization Assumption:** Given the attribute value of X's friend at i (i ≥ 1) hops away, Y, the attribute value of X is conditionally independent of the descendants of Y
- Naive Bayesian Assumption:** Given the attribute value of X, the attribute values of Y (direct friends of X) are conditionally independent of each other



Reduction of a social network (a) into a Bayesian network to infer X from friends Y via Localization Assumption (b) and Naive Bayesian Assumption (c). The shaded nodes represent friends with known attributes.

Bayesian Inference
$$\hat{x} = \arg \max_x P(X = x | Y_1, Y_2, \dots, Y_i), x \in \{t, f\}$$

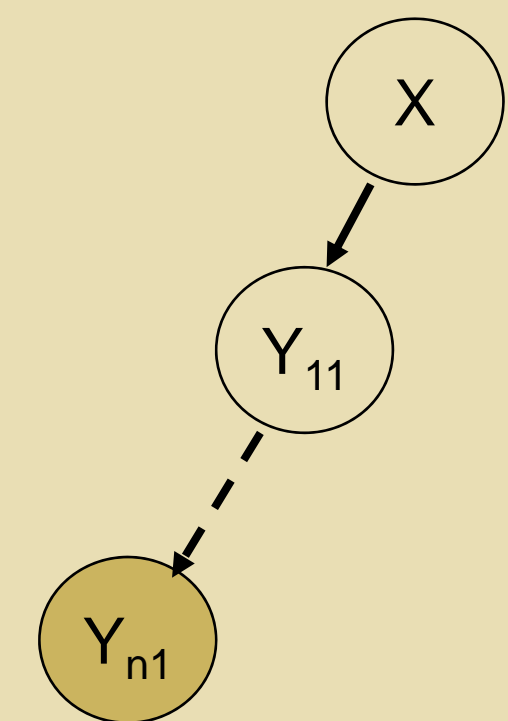
Protection

Casual Effects From Friends Attribute Values

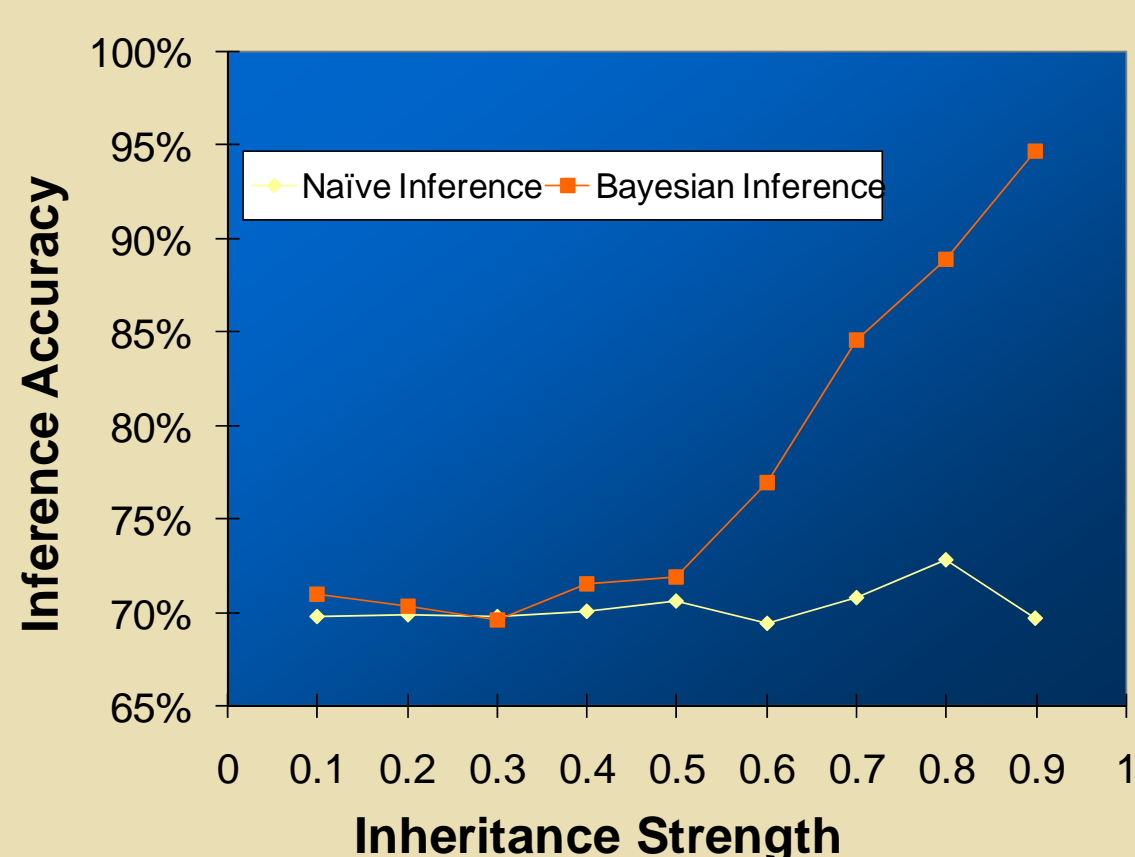
- For a descendant Y_{n1} n hops away from the ancestor X in a chain, if the attribute values of all the intermediate nodes Y_{i1} (i < n) are unknown, then $P(X=t | Y_{n1}=t) > P(X=t)$ iff $(I-M)^n > 0$, and $P(X=t | Y_{n1}=f) > P(X=t)$ iff $(I-M)^n < 0$, where I is Inheritance Strength, the tendency that a child inherits its parent's attribute, and M is Mutation Strength, i.e., the tendency that a child develops its attribute by mutation

Protection Schemes

- Attributes
 - Randomly Hiding Attributes (RHA), Selectively Hiding Attributes (SHA), Selectively Falsifying Attributes (SFA)
- Relations
 - Selectively Hiding Relations (SHR), Selectively Adding Relations (SAR)



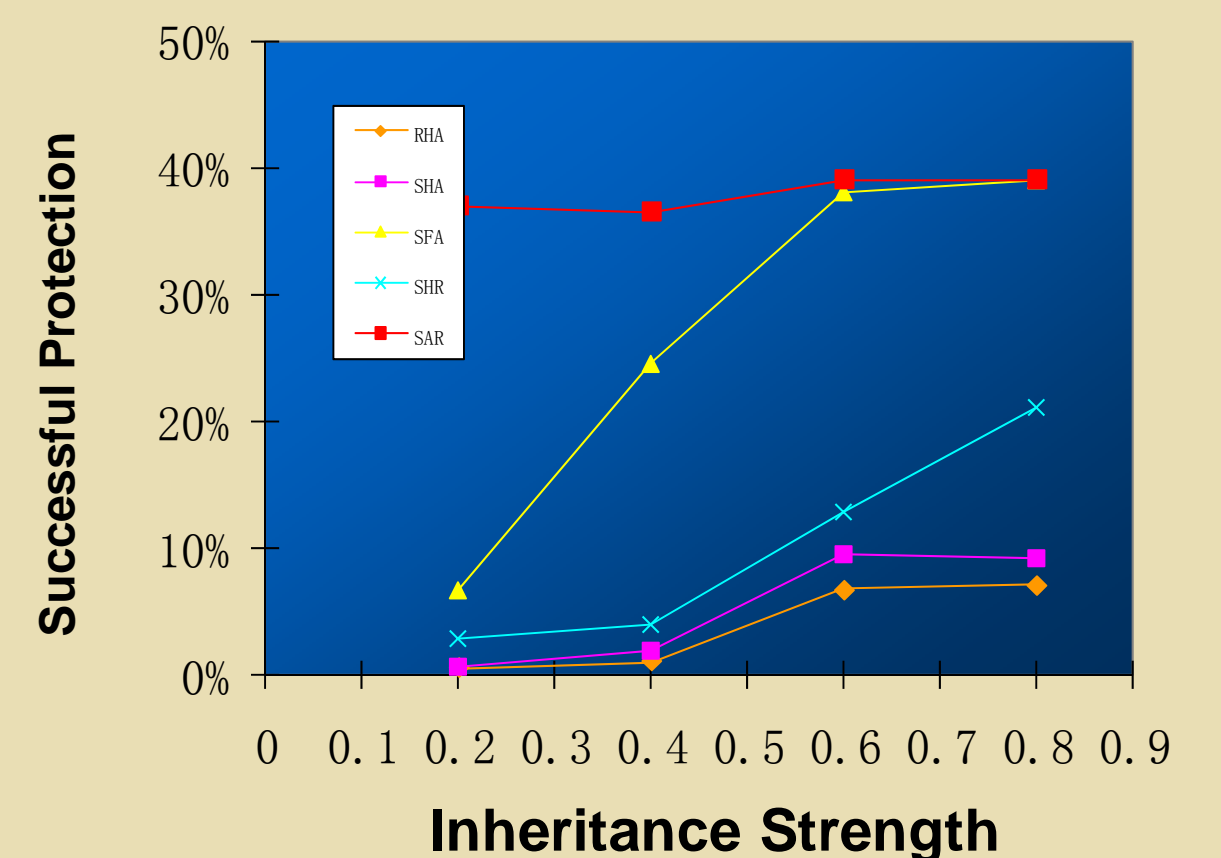
Experimental Results



(a) Inference Accuracy of Naive Inference and Bayesian Inference (prior probability = 0.3), Simulation based on Livejournal Dataset Structures

| Target Interest | Naive Inference | | | Bayesian Inference | | |
|--------------------------|-----------------|-----------|--------|--------------------|-----------|--------|
| | Accuracy | Precision | Recall | Accuracy | Precision | Recall |
| Health | 0.539 | 0% | 0% | 0.638 | 61.4% | 58.1% |
| Online Stores & Services | 0.522 | 0% | 0% | 0.606 | 58.4% | 85.4% |
| Restaurants & Gourmet | 0.568 | 0% | 0% | 0.642 | 63.4% | 40.5% |
| Electronics | 0.766 | 0% | 0% | 0.765 | 76.6% | 99.7% |

(b) Inference Accuracy of Naive Inference and Bayesian Inference, Experiment based on Epinions Dataset



(c) Performance of Privacy Protection Schemes, Simulation based on Livejournal Dataset Structures

Conclusion

- Privacy can be inferred from social relations especially when people are closely connected
- Privacy protection needs to consider both social relations and network structure
- Selectively Falsifying Attributes is the most effective scheme for privacy protection