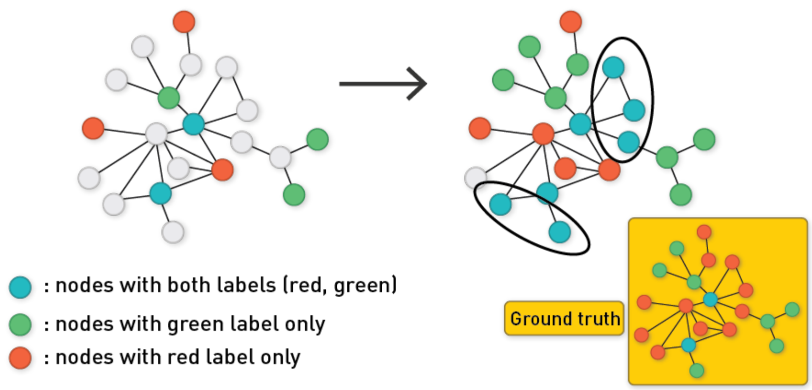
Apply RN in Multirelational Network



Profile Inference

Profile:

Hometown:	Palo Alto
High school:	Gunn
College:	Stanford
Employer:	Facebook
Current city:	Sunnyvale

Hobbies, politics, music, …

A complete profile is a boon:

People are easily searchable

Tailored news recommendations

Group recommendations

Ad targeting (especially local)

How can we fill in missing profile fields?

Relational Neighbor

- The relational neighbor (RN) classifier estimates class probabilities solely based on entities of the same type whose class labels are known.
- The classifier works by making two strong, yet often reasonable, assumptions:
  - Some entities' class labels are known within the same linked structure.
  - The entities exhibit homophily—entities related to each other are similar and likely belong to the same class along one or more dimensions.
  - The classifier may not perform well if entities are isolated or if no labels are known.

Motivation

- The conventional relational classification model focuses on the single-label classification problem.
- Real-world relational data sets contain instances associated with multiple labels.
- Connections between instances in multilabel networks are driven by various casual reasons.

Previous Work

- Random Walks: Talukdar and Crammer (2009), Baluja et al. (2008)
- Statistical Relational Learning: Lu and Getoor (2003), Macskassy and Provost (2007)
- Relational Dependency Networks: Neville and Jensen (2007)
- Latent Models: Palla, Knowles, and Ghahramani (2012)
- Either:
  - Too generic; require too much labeled data
  - Do not handle multiple label types
  - Are outperformed by label propagation (Macskassy & Provost 2007)

Contribution

- A new multilabel iterative relational neighbor classifier (SCRN).
- Extract social context features using edge clustering to represent a node's potential group membership.
- Use of social features boosts classification performance over benchmarks on several real-world collaborative networked data sets.

The EdgeExplain Model

- Instead of taking friendships as given, â�¨explain friendships using labels.
- A friendship u~v is explained if â�¨u and v share the same:
  - Hometown *or*â�¨
  - Current city *or*
  - High school *or*
  - College *or*
  - Employer

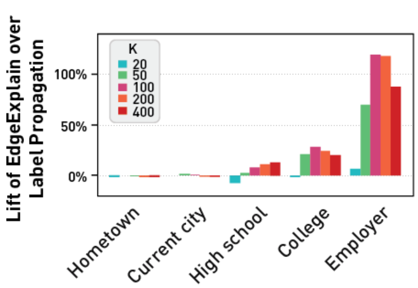
The EdgeExplain Model

- Instead of taking friendships as given, â�¨explain friendships using labels.
- A friendship u~v is explained if â�¨u and v share the same:
  - Hometown *or*â�¨
  - Current city *or*
  - High school *or*
  - College *or*
  - Employer
- Use k-Nearest Neighbors
- Probabilistic model

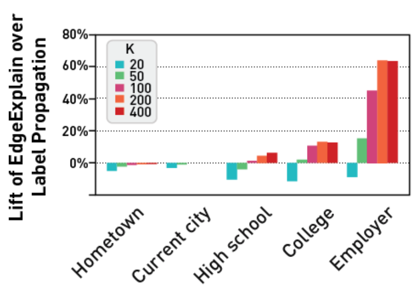
Experiments

- 1.1B users of the Facebook social network
- O(10M) labels
- Fivefold cross-validation
- Measure recall
    - Did we get the correct label in our top prediction? Top three?
- Inference:
    - Proximal gradient descent
    - Implemented via message passing in Apache Giraph (Ching 2013)
- Sparsify graph by considering k closest friends by age

Results (vs. Label Propagation)



Recall@1



Recall@3

- Joint modeling helps most for employers
- Significant gains for high school and college as well

EdgeExplain Algorithm Conclusions

- Assumption: Each friendship has one reason.
- Model: Explain friendships via user attributes.
- Results: Up to 120% lift for recall at 1 and 60% for recall at 3.

References

- Mackskassy, S. A., & Provost, F. (2003). A simple relational classifier. *Proceedings of the Second Workshop on Multi-Relational Data Mining (MRDM) at KDD*, 64–76.
- Tang, L., & Liu, H. (2009). Relational learning via latent social dimensions. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 817–826.
- Tang, L., & Liu, H. (2009). Scalable learning of collective behavior based on sparse social dimensions. *Proceedings of International Conference on Information and Knowledge Management (CIKM)*, 1107–1116.
- Neville, J., Gallagher, B., Eliassi-Rad, T., & Wang, T. (2011). Correcting evaluation bias of relational classifiers with network cross validation. *Knowledge and Information Systems (KAIS)*, 1–25.â�¨