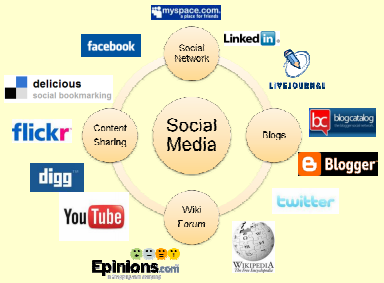
Relational Learning via Latent Social Dimensions

Lei Tang and Huan Liu
L.Tang@asu.edu, Huan.Liu@asu.edu
Computer Science & Engineering, Arizona State University, USA

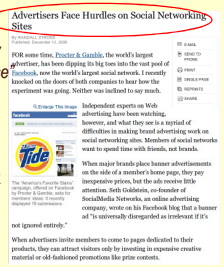
ASU ARIZONA STATE UNIVERSITY

Boom of Social Media



A Motivating Example

"In 2008, 57% of all users of social networks clicked on an ad and only 11% of those clicks lead to a purchase"



Reality:
Limited user profile information;
Readily available Social Network

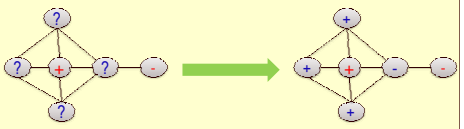
Core Problem:
How to utilize Social Network information to help predict user preference or behavior?

Problem Formulation

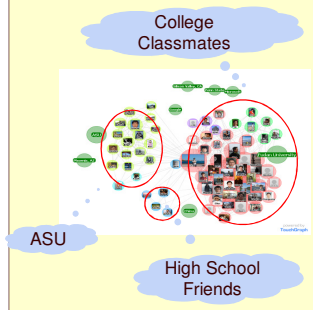
Preference or Behavior can be represented by labels (+/-)
- Whether or not clicking on an ad
- Whether or not interested in certain topics
- Subscribed to certain political views
- Like/Dislike a product

Given:
- A social network (i.e., connectivity information)
- Some actors with identified labels

Output:
- Labels of other actors within the same network



Heterogeneous Relations in Social Media



Connections in a social network are heterogeneous
- Colleagues
- Family/Relatives
- Classmates

Each type of relation is likely to be associated with one affiliation.

Each actor involves in multiple affiliations (e.g., ASU, college, high school)

This affiliation information in social media is not always available

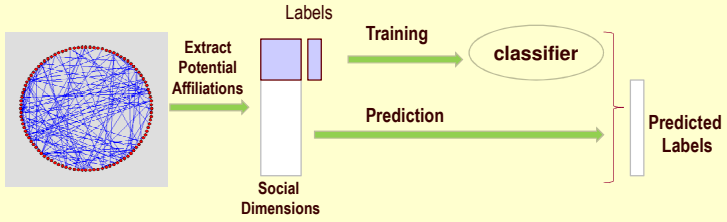
Simply taking a homogeneous view of these connections is often insufficient

Extracting Actor Affiliations

Key observation:
- Actors of the same affiliation are likely to interact with each other more frequently.
  - Each affiliation corresponds to one community.
- Each actor can involve in multiple affiliations
  - Use soft community detection methods to extract potential affiliations

Actor	Affiliation 1	Affiliation 2
1	1	1
2	1	0
3	0	1
...	.....	.....

Relational Learning Framework based on Social Dimensions (SocDim)



- Training:
  - Extract social dimensions to represent potential affiliations of actors
    - Applicable modules: modularity maximization, graph Laplacian, etc.
  - Build a classifier to select those discriminative dimensions
    - Applicable modules: SVM, logistic regression, etc.
- Prediction:
  - Predict labels based on one actor's latent social dimensions

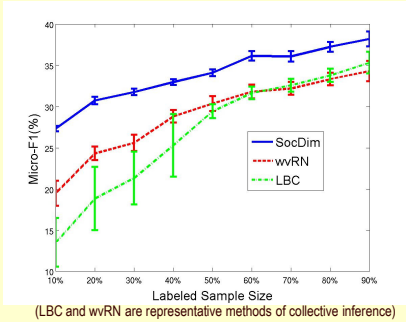
Baseline Method: Collective Inference/Label Propagation

Markov Assumption:
- Label of one node depends on that of its neighbors

Training:
- Build a relational model based on labels of neighbors

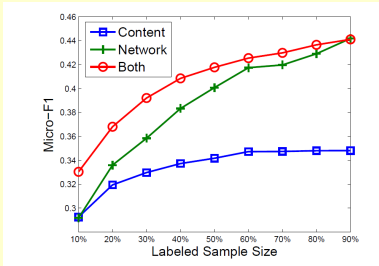
Prediction --- **Collective inference**
- Predict the labels of one node while fixing labels of neighbors
- Iterate until convergence
- Treat connections in a network homogeneously

Empirical Results



(LBC and wvRN are representative methods of collective inference)

SocDim converts network into features; If node features are available, social dimensions can be combined directly for discriminative learning.



Content: use node content feature alone; *Network*: use network alone; *Both*: combine social dimensions and content info together

Conclusions

1. Networks in social media are *noisy* and *heterogeneous*
2. SocDim, capturing potential affiliations of actors, outperforms other representative methods based on collective inference.
3. SocDim converts network-format data into feature format; can be combined with other node features

Acknowledgement: AFOSR Grant FA95500810132

If you have any questions, please contact Lei Tang.
http://www.public.asu.edu/~ltang9/