Learning to Discover Social Circles in Ego Networks

Julian McAuley, Jure Leskovec

Stanford University

Abstract

Our personal social networks are big and cluttered, and currently there is no good way to organize them. Social networking sites allow users to manually categorize their friends into **social circles** (e.g. 'circles' on Google+, and 'lists' on Facebook and Twitter), however they are laborious to construct and must be updated whenever a user's network grows. We define a novel machine learning task of identifying users' social circles. We pose the problem as a node clustering problem on a user's ego-network, a network of connections between her friends.

"Knows your circles better than you do!"

– Wired

Properties of circles

Our goal is to automatically detect **circles** using **profile** and **network** information. We develop a model of circles with the following properties:

- Circles form around nodes with common properties.
- Different circles are formed by different properties, e.g. one circle might be formed by family members, and another by students who attended the same university.
- Circles can overlap, and 'stronger' circles form within 'weaker' ones, e.g. a circle of friends from the same degree program may form within a circle from the same university.
- We leverage both profile information and network structure in order to identify circles.

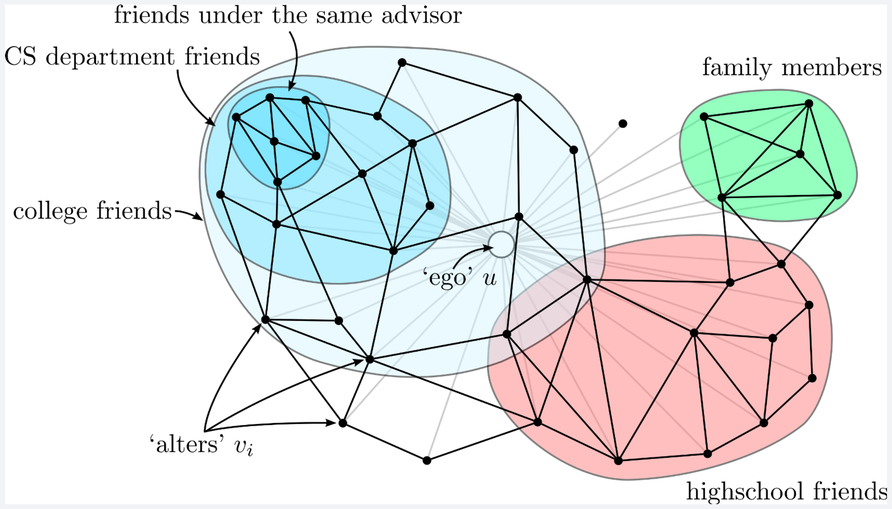
Model

Our model predicts **hard** memberships to multiple, **overlapping circles**, using both **profile** and **network** information.

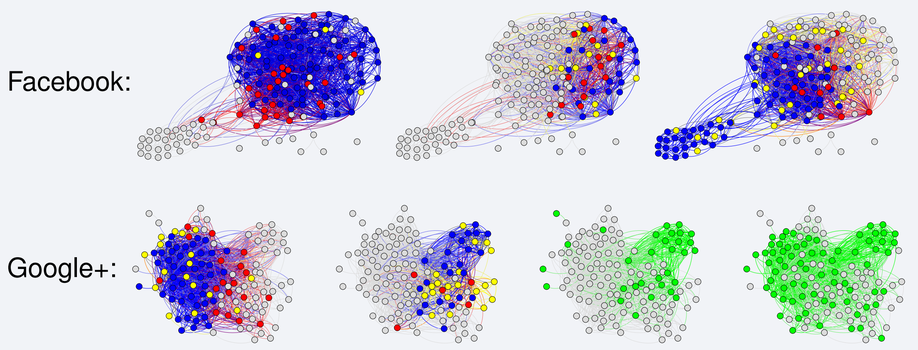
$$p((x,y) \in E) \propto \exp\left\{ \underbrace{\sum_{C_k \supseteq \{x,y\}} \langle \phi(x,y), \theta_k \rangle}_{\text{circles containing both nodes}} - \underbrace{\sum_{C_k \not\supseteq \{x,y\}} \alpha_k \langle \phi(x,y), \theta_k \rangle}_{\text{all other circles}} \right\}$$

Training is done by maximum likelihood, using QPBO and L-BFGS.

An ego-network

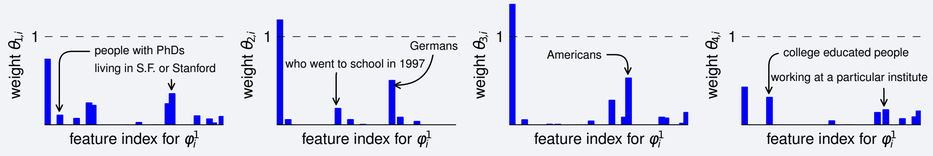


Some detected circles



Blue = true positive; gray = true negative; red = false positive; yellow = false negative; green = detected circles for which we have no groundtruth.

Examples of model parameters for four circles



Data

We collect data from **Facebook, Google+, and Twitter**

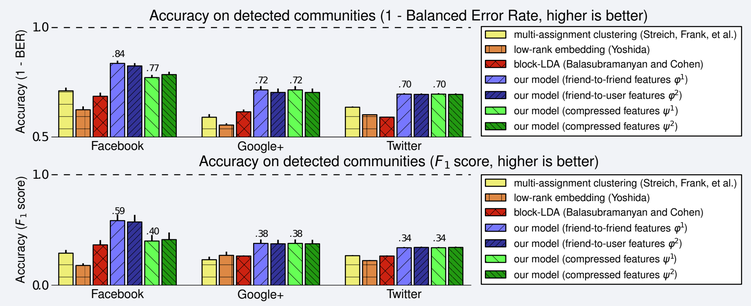
	ego-networks	circles	nodes	edges
Facebook	10	193	4,039	88,234
Google+	133	479	107,614	13,673,453
Twitter	1,000	4,869	81,306	1,768,149

All data are **available** on `snap.stanford.edu/data/`

"Results are decent"

– MIT Technology Review

Some results



Bibliography

McAuley, J., and Leskovec, J. 2012. Discovering Social Circles in Ego Networks. In *arXiv:1210.8182*

Streich, A., Frank, M., Basin, D., and Buhmann, J. 2009. Multi-assignment clustering for boolean data. In *International Conference on Machine Learning*.

Yoshida, T. 2010. Toward finding hidden communities based on user profiles. In *ICDM Workshops*.

Balasubramanyan, R. and Cohen, W. 2011. Block-LDA: Jointly modeling entity-annotated text and entity-entity links. In *SIAM International Conference on Data Mining*.

Rother, C., Kolmogorov, V., Lempitsky, V., and Szummer, M. 2007. Optimizing binary MRFs via extended roof duality. In *Computer Vision and Pattern Recognition*.

jmcauley@cs.stanford.edu, jure@cs.stanford.edu