Overview

Location plays an essential role in our lives, bridging our online and offline worlds. We explore the interplay between people's location, interactions, and their social ties within a large real-world dataset. Our system, Flap, solves two intimately related tasks: link and location prediction in online social networks. We evaluate Flap on a sample of 11 thousand highly active users from New York City and Los Angeles, and show that it (1) reconstructs the entire friendship graph (60M possible edges) with high accuracy even when no edges are given; and (2) infers people's fine-grained location, even when they keep their data private and we can only access the location of their friends.

Our models significantly outperform current comparable approaches to either task.

Friendship Prediction

Flap infers social ties by leveraging patterns in friendship formation, the content of people's messages, and user co-location. Friendships on Twitter are predicted by the process shown on the left. The belief propagation algorithm is summarized below.

For supervised learning, the optimal set of parameters $\theta$ given observed $(x)$ and hidden $(y)$ random variables can be estimated directly from training data:

$$\theta^* = \arg \max_{\theta} \log \left( \Pr(x_{1:t}, y_{1:t}|\theta) \right)$$

In unsupervised learning, we apply expectation-maximization method:

$$\theta^* = \arg \max_{\theta} \log \sum_{y_{1:t}} \Pr(x_{1:t}, y_{1:t}|\theta)$$

While each observation type is a weak predictor of friendship when considered in isolation, combining them results in a strong model, accurately identifying the majority of friendships.

The plot on the left shows Flap's performance in terms of ROC curves, and compares it to alternative models. We vary the proportion of friendships that are given to Flap from 0% to 50%.

Conclusions

We show that much information can be inferred about individuals from their interactions in online social media, without active user participation. Namely, we predict Twitter friendships with high accuracy even when no ties are given. Additionally, we infer fine-grained user location, even when they keep their data private and we can only access the location of their friends.

References