Cross-Domain Entity Resolution in Social Media

Problem
- Associate entities (typically people and organization) across multiple social media sites
- Provide integration of information from multiple sites to create a complete picture of user and community activities, characteristics and trends

Our Approach
- Profile-based methods
  - Approximate string matching
- Content-based methods
  - Author identification
- Graph-based methods
  - Cross-domain community detection
  - Neighborhood-based features

Profile-based and Content-Based

Profile-Based Methods
- Content-based methods
  - Associate entities (typically people and organization)
- Provide integration of information from multiple sites to create a complete picture of user and community activities, characteristics and trends

Content-based entity resolution is performed with an SVM authorID system

Conclusions
- Examined novel methods for extracting features
- Showed profile and graph features worked the best in combination
- Achieved excellent performance for non-trivial entity resolution
- Software released on https://github.com/mitll
- Future Work:
  - Create dynamic features
  - Extend to non-English social media platforms
  - Improve the scalability using hashing
  - Perform collective entity resolution—joint community-detection, entity resolution iteration

MIT LL — Lin Li, William Campbell, Charlie Dagli, Joel Acevedo-Aviles, Kelly Geyer, Joseph Campbell; Johns Hopkins University – C. Priebe