Social Network Extraction from Texts: Thesis Proposal

Apoorv Agarwal
Department of Computer Science, Columbia University

**Overall Goal**
Extract a social network from text where nodes are people and links are social events.

**Impact**
- Social network analysis community
- Study of literary and journalistic texts
- Linguistics (add frames to FrameNet)

**Social Events (Agarwal et al. 2010)**
- **Social Event**: An event between two people or group of people where at least one party is aware of the other party and aware of the event. Types:
  - Interaction event (INR): both parties mutually aware
  - Observation event (OBS): only one party aware of the other

**Future Work 1: Domain adaptation**
- **Current training data**: English part of Automatic Content Extraction (ACE) 2005 multilingual news corpus
- **Need domain adaptation**: to port the technique to richer sources of social networks like literary and historical texts.
- **Notable work** by Hal Daumé III, 2007: present a straightforward kernelized version of domain adaptation
  
  \[ K'(T_1, T_2) = \begin{cases} 
  2K(T_1, T_2) & \text{if same domain} \\
  K(T_1, T_2) & \text{if different domain} 
  \end{cases} \]

**Future Work 2: Scaling convolution kernels**
- **Kernels**
  - **PET**: used for all other structures
  - **PT**: used for PET

**Future Work 3: Interpreting convolution kernels**
- **Linearization of SST kernel** (Pighin and Moschitti, 2009):
  \[ f(x) = \sum_{i=1}^{N_s} \alpha_i y_i K(s_i, x) + b \] (Burges, 1998) (1)
  \[ K(T_1, T_2) = \sum_{s} h_s(T_1) h_s(T_2) \] (Collins and Duffy, 2002) (2)

**Future Work 4: Applications**
- **Predict hierarchy of Enron email corpus**: capture “who talks to whom about whom” type of links
- **Use phrase-level polarity analysis** (Agarwal et al. 2009, 2011) to add polarity on edges of the network