

Social Identity Group Discovery: A Data Mining Research Challenge

Gary W. Strong
Human Language Technology
Center of Excellence
Johns Hopkins University

Integrating Data Mining with Social Science

- Panel Title: Future Research Challenges and Needed Resources for the Web, Semantics, and Data Mining
- Social science perspective
- Drawing from a MITRE Internal R&D effort now in its 3rd year

“Leaderless Resistance” article by Simpson Garfinkel*

- Social Identity Groups can exist without members at times
- Role of the Internet in group persistence

* Garfinkel, S. [Leaderless Resistance Today](#), *First Monday*, 8:3, March 3rd, 2003.

MITRE internal research project on “Phase Change Behavior”

- Influence of social groups (conference and publication venues) on papers
- “Team Assembly” paper by Guimera et al*
- Created a modeling framework in which the attributes of social identity groups play a key role in modeling group dynamics

* Guimera, R., B. Uzzi, J. Spiro, and L. A. N. Amaral. "Team Assembly Mechanisms Determine Collaboration Network Structure and Team Performance." *Science* 308, no. 5722 (2005): 697-702

Evolution of Terrorism from Organized Efforts to “a bunch of guys”*

- 9/11 to Madrid to London to recent cells in Germany
- Publication of technical how-to on sites like “YouTube”**

* Marc Sageman, personal communication

** YouTube Bomb Video Brings Scrutiny

Terrorism Suspect Allegedly Gave Tips on Using Detonator

<http://www.washingtonpost.com/wp-dyn/content/article/2007/09/30/AR2007093001648.html>

Internet offers canned solutions associated with social identity groups

- People rarely decide complex issues on bare facts and reasoning
- It is simpler to adopt an identity group's "solutions"
- Propensity for this is embedded in cultural evolution



When confronted with a problem to solve, an individual may choose from a variety of identity groups that offer solutions.

Relevance to Data Mining?

- Paradigm of social networking is not able to deal with social identity groups
- Identity group affiliations play a significant role in individual decisions
- Discovering such groups is a data mining problem in need of better formulation
 - Similar to “social radar” approaches to early detection of infectious disease