Mining and Mine Field:
From Understanding to Forecasting

Dipak K. Gupta

Principle Investigator: Dr. Ming-Hsiang Tsou mtsou@mail.sdsu.edu, (Geography), **Co-Pis:** Dr. Dipak K Gupta (Political Science), Dr. Jean Marc Gawron (Linguistic), Dr. Brian Spitzberg (Communication), Dr. Li An (Geography). *San Diego State University, USA*.

Root Causes of Mass Movements: The Necessary Conditions

- A collective movement takes places only when there is a shared set of grievances: economic, political, ethnic, religious, or environmental

- However, these factors of frustration and anger are but the necessary conditions for actual formation of collective movements. For a movement to take place we need to have the presence of the sufficient conditions

The Sufficient Conditions of Creating a Movement

- For sufficient condition, we need conduits which can channelize the feeling of shared frustration in an organized form

- This is done by "Political Entrepreneurs," the leaders of political parties, labor and student unions, leaders of NGOs, heads of religious orders

What do "Political Entrepreneurs" do?

They provide a psychological "frame" through which events can be understood by their followers. This framing is done by:

- Defining the boundaries of the aggrieved community ("Us" factor or the "In-group")

- Defining the offending group ("Them" or the "Out-group)

- Articulating a clear threat from "them" to "Us"

Understanding the spread of extremist ideology: The case of White Supremacists

In the CDI project, we created a system for automatically ranking documents by degree of militancy, designed as a tool both for finding militant websites and prioritizing the data found.
Our ranking system employed a small hand-selected vocabulary based on group membership markers used by insiders to identify members and member properties (us) and outsiders and threats (them). We used the same vocabulary to build a classifier. Evaluating several
ranking systems by their correlations with human judgments, we showed that the best ranker uses the small us-them vocabulary, outperforming one system with a much larger vocabulary, and another with a small vocabulary chosen by Mutual Information.

From Understanding to Forecasting of Civil Strife

The next step from "understanding" is forecasting. In my other project, EMBERS, we attempted to forecast civil strife in Latin American countries with the help of my model of mass movement.

For an accurate forecasting based on "Big Data", we use all kinds of social media and other Internet postings to identify the necessary and sufficient conditions, along with a third factor: the "trigger events."

These trigger events can take many forms, from elections to incidents of police brutality, from a Papal visit to the decision to construct a dam.

The Puzzle of Going Viral

Mass movements from the earliest times have been caused by rumors. The paths of those rumors must have traversed similar paths (without the technology) as they do in today's world.

1. The best reason for a message going viral is the existence of repeatable visual or audio tape, which can be reproduced many times.
2. The message must be shocking (e.g., the picture of an actual beating or rape as opposed to the general knowledge that such things happen).
3. Ready identification with the individual victims (facial/voice recognition)
4. The "political entrepreneurs" repeating and rebroadcasting the messages in the social media
5. The messages jumping from social media to the traditional media
6. The messages entering into regional/national/international political discourse through the framing of "Us", "Them," and "Threat."

The Classification of Forecasting

Causality	Timing	
	Timing known	**Unknown Timing**
Causality Known	*Known Knowns* (almost certain events: preplanned demonstrations, protest occurring of specific, politically significant days)	*Known unknowns* * (Widespread grievance but unknown trigger events – cutting down trees in a park near Taksim Square in Istanbul)
Causality unknown	*Unknown knowns* (Regular events in social strife: perhaps the commemoration of emotionally charged events on specific days of religious/political significance)	*Unknown unknowns* (Black swan events: perhaps precipitated by natural disasters, political assassinations, or sudden death of a leader)

Mining and The Minefield

"Data mining" was considered a "dirty word" in social science research. Yet, now this is proving us with the greatest opportunity to understand our collective behavior.

As computing technology exponentially increases in capabilities, we are in a position to analyze incredible volumes of data by fusing it with human judgment. For the first time, since the development of earliest recorded scholarship, we are breaking the barriers of strict disciplinary boundaries.

At the same time, careful examination of our effort to understand human communication on a grand scale exposes the limits of our cognitive capabilities in making sense of the complexities of human communications and the grave policy mistakes that we may make while trying to do so.