CLASSIC: Software Platform for Solving Real-World Problems by Understanding Virtual Communities

Supported by AFOSR grants FA9550-07-1-0253 (CCPV) and FA9550-09-1-0261; ONR grant N00014-09-1-0805. Algorithms under U.S. Patent 7499965.
Role in CCPV Project

COHERENCE-BASED MODELING OF CULTURAL CHANGE AND POLITICAL VIOLENCE (CCPV) - Tasks and interdependencies -

Cross-national statistical analysis of impact of culture

Experiments on the role of culture in decision-making

Case studies on ethnic conflict

Theory-building & model specification

Provides action, cultural change and structural models

Programming of the simulation

CROSS-COUNTRY DATASET

Compiles

Collection of empirical data on ethnicity

Provides input data

SIMULATION

WEB CRAWLER

Adapting social science theories to virtual communities

Provides theories

Programming of the web crawler

Software program

Retrospective testing of predictions

Software program
Past and Present Members of CCPV Team

Sun-Ki Chai (Dept. of Sociology)
David Chin (Dept. of Information & Computer Sciences)
Scott Robertson (Dept. of Information & Computer Sciences)
Gurdal Arslan (Dept. of Electrical Engineering)
Min-Sun Kim (Dept. of Speech Communication)
Jang-Hyun Kim (Dept. of Speech Communication)
Mooweon Rhee (Dept. of Management & Industrial Relations)
Ruben Juarez (Dept. of Economics)
Keisuke Nakao (Dept. of Economics, Hilo)
Ilan Noy (Dept. of Economics)
Katherina Sherstyuk (Dept. of Economics)
Abdul Karim Khan (Dept. of History, LCC)
Leonard Andaya (Dept. of History)

Research Assistants: Aaron Herres, Dong-Wan Kang, Evan Yazawa, Suren Dorj, Ming Liu, Debbie Gundaya, Kiran Sagoo, Cecilia Noble, Quincy Edwards, Turro Wonkaren, Kar-Hai Chu, Leo DeCandia, Robert Puckett
Why the Need for General Web Social Analysis Application?

- The internet growing importance of as medium for communication, social interaction, commerce, and political activity.
- Amount of data available is unprecedented.
- Most data publicly available, accessible from anywhere in the world, and can be collected with minimal intrusion on human subjects.
- Business, government, academia, are rapidly turning to internet as a resource to analyze social phenomena.
- Despite the quantity and accessibility of data, relatively few large-n studies of web social dynamics.
- Absence of convenient technology for collecting and organizing this data is the major roadblock.
- Expansion in practical uses of search technology in last several years disappointing.
General Kinds of Things You Might Want to Find Out

• Finding which combination of content, link, “annotation”, and offline (geographical, demographic) characteristics are most significant in determining a site or member’s popularity, power, influence, etc. Subset by target “audience”.

• Determining attitudes by site/member along multiple user-specified dimensions towards a particular entity or issue. Associating with lifestyle, ideology, inferred SES, demographic characteristics.

• Measuring changes in attitudes over time and the generation and diffusion of ideas across networks.

• Using virtual community analysis to identify the key concepts and areas of conflict that define a particular virtual community. What are their ancillary interests?

• Using attitudinal and relational data, along with available structural information, to predict behavior and/or explain the processes behind behavior.
“Reality Mining” of Web: 
A Platform for Answering Social Questions

• Improved methods for identifying boundaries of virtual communities of interest.
  – group network measures, discourse theme coherence, homogeneity characteristics

• More versatile methods for measuring the attitudinal, ideological, and cultural of defined populations
  – plug-in dictionaries, limited syntactic parsing, mini-dictionaries for general cultural typologies, inferred homogeneity

• Multi-method validation of such methods through comparison with established social science methodologies.
  – human coder comparison, survey-based indices, computer-mediated experiments, event history analysis

• General framework for generating short and long-term behavioral predictions
  – rational choice model with coherence-based endogenous preference change
“Intermediate” User Interface

Seed sites and key terms:

Rank results by:
- Broker Power
- Authority
- Similarity
- Prestige
- Influence
- Popularity

Community members
This is an Example web page title
www.exampleURL.com
Put description here. Limit: 1 line.
<more information> <reseed search> <analyze this forum>

Visualizations
Data Content Options
Data content types
Emotions
Content type entries
Chosen content
good
excellent
could be

Graph Types
Pie Graph
Line Graph
Bar Graph
Network Graph
CLASSIC: Specification of Seed Sites and Freeform Crawl Criteria

It is not safe to modify the sequence being iterated over in the loop (this can only happen for mutable sequence types, such as lists). If you need to modify the list you are iterating over (for example, to duplicate selected items) you must iterate over a copy. The slice notation makes this particularly convenient:
Using Online Data to Predict Behavior: Example of Grid-Group Framework

**Grid**: extent to which social rules prescribe and restrict action

**Group**: extent to which identity is directed towards others

Closely related to classical sociological concepts of regulation and integration

- Operationalization methods straightforward and well-tested
- Works well as front-end to “thin” rational choice models of decision-making
- Fits with abstract dimensions of social organization found in social theories, e.g. regulation and integration
Expected Regret (single-period, individual form):

\[ d = \int_s (u(s,a^*(s)) - u(s,a))) p(s) \, ds \]

where

\[ a^*(s) = \arg\max_{a \in A} u(s,a) \]

\[ a = \arg\max_{a \in A, s \in S} \int_s u(s,a) \, p(s) \, ds \]

s states of the environment, a actions, u utility function, and p subjective probabilities

adjustment of g, h to minimize d within “non-yogic” utility and information constraints
### A Priori Validation: Grid-Group and Behavior in Voluntary Contributions Mechanism

<table>
<thead>
<tr>
<th></th>
<th>Shuffled</th>
<th>Partner</th>
<th>Pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Contribution Level, No Pun</td>
<td>Contribution, with Punish</td>
<td>Punishment Expenditure</td>
</tr>
<tr>
<td>Grid Pearson correlation</td>
<td>-0.100</td>
<td>0.038</td>
<td>0.263**</td>
</tr>
<tr>
<td>Grid Sig. (2-tailed)</td>
<td>0.392</td>
<td>0.748</td>
<td>0.022</td>
</tr>
<tr>
<td>Group Pearson correlation</td>
<td>0.350***</td>
<td>0.096</td>
<td>-0.116</td>
</tr>
<tr>
<td>Group Sig. (2-tailed)</td>
<td>0.002</td>
<td>0.411</td>
<td>0.318</td>
</tr>
<tr>
<td>N</td>
<td>76</td>
<td>76</td>
<td>76</td>
</tr>
</tbody>
</table>

***Correlation is significant at the 0.01 level (2-tailed).
** Correlation is significant at the 0.05 level (2-tailed).
* Correlation is significant at the 0.10 level (2-tailed).
Field Study: Risk Preferences Among Moro Ethnic Groups

Graphs by eth2
How It Works

Community

Candidate sites

Single page

SEED SITE

SEED SITE

LINKS=10
CONTENT=0.65

LINKS=8
CONTENT=0.78

LINKS=10
CONTENT=0.89

Community sites
Attitudinal Content Tracking
Member Network Analysis

Includes:

- Connections through thread sharing
- Time prioritization
- Rank of reply prioritization
- All previous filtering types on threads
Contructionist Behavioral Simulation: Identity Groups Predicted, Not Assumed