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.
Role in CCPV Project

- 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
- Collection of empirical data on ethnicity
- Adapting social science theories to virtual communities

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

Programming of the simulation

CROSS-COUNTRY DATASET

Collection of empirical data on ethnicity

Provides input data

SIMPULATION

Retrospective testing of predictions

WEB CRAWLER

Programs the web crawler

Software program

Provides theories

Data on virtual communities
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)
Mooween Rhee (Dept. of Management & Industrial Relations)
Ruben Juarez (Dept. of Economics)
Keisuke Nakao (Dept. of Economics, Hilo)
Ilan Noy (Dept. of Economics)
Katherina Sherstuyk (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
CLASSIC: Specification of Seed Sites and Freeform Crawl Criteria

Table Of Contents
4. More Control Flow Tools — Python v2.6.4 documentation — Mozilla Firefox

4.1. ur Site
4.2. for Staters
4.3. The range Function
4.4. break and Statements, a Causes on Lo
4.5. pass State
4.6. Defining
4.7. More on Functions
  4.7.1. Defa Argument
  4.7.2. Key Arguments
  4.7.3. Arbit Argument
  4.7.4. Uimp Argument
  4.7.5. Lambda
  4.7.6. Document Strings
  4.8. Internets Coding Style

Previous topic
3. An Informal In to Python

Next topic
5. Data Structures

This Page
Show Source
Quick search
Go
Enter search terms or a module, class or function name.

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:

```python
>>> for x in a[:]: # make a slice copy of the entire list
...     if len(x) > 6: x.insert(0, x)
```
“Intermediate” User
Initiation/Runtime/Result

---CLASSIC---

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>

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

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>

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
  - Positive
  - wonderful
  - great
  - Negative
  - Uncertain
  - maybe
  - perhaps
  - perchance

Graph Types

- Pie Graph
- Line Graph
- Bar Graph
- Network Graph

Advanced Search Options
**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</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson</td>
<td>-0.100</td>
<td>0.038</td>
<td>0.263**</td>
</tr>
<tr>
<td>correlation</td>
<td>0.392</td>
<td>0.748</td>
<td>0.022</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>0.052</td>
<td>0.582</td>
<td>0.683</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.075</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.108</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.132</td>
<td>0.131</td>
</tr>
<tr>
<td>Group</td>
<td>Pearson correlation</td>
<td>0.350***</td>
<td>-0.116</td>
</tr>
<tr>
<td></td>
<td>0.096</td>
<td>0.251*</td>
<td>-0.218</td>
</tr>
<tr>
<td></td>
<td>0.002</td>
<td>0.171</td>
<td>0.319***</td>
</tr>
<tr>
<td></td>
<td>0.411</td>
<td>0.106</td>
<td>0.138</td>
</tr>
<tr>
<td></td>
<td>0.318</td>
<td>0.000</td>
<td>0.155*</td>
</tr>
<tr>
<td></td>
<td>0.062</td>
<td>0.114</td>
<td>0.077</td>
</tr>
<tr>
<td></td>
<td>0.208</td>
<td>0.132</td>
<td>0.132</td>
</tr>
<tr>
<td></td>
<td>0.106</td>
<td>132</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>132</td>
<td>132</td>
</tr>
<tr>
<td>N</td>
<td>76</td>
<td>132</td>
<td>132</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 sites

Candidate sites

Single page
Forum Analyzer

Raw Text Message:

else, am i right?

Joe, Not sure on the switch aspect of the mirror. But I think you can get this place to re-silver the mirror. You need to send just the mirror glass.

Hey, looking for a (cheap) place to stay in Albuquerque for one night going to be passing through on the way to Vegas from Dallas, pull in out here (dallas) early tuesday morning, if anyone knows of a clean but cheap place to stay or something like that, let me know.

Super 8 Motel on I-40, exit 196. Only US $ 42.85

the all OM car show is at Woodley Park next Sunday, June 28th, who's going? Basketball Sam says he is coming up, I should be there with a mmmmm black auburn racecar?

Will try to make it. Just moved to Palm Springs from L.A. No Buick yet, but sounds like fun. Would like to meet B & Sam too.

To anyone that's never been there: it has always been a great event. I would like to attend but might not be able to go. Some stupid little Toyota will be showing up here next weekend to get its bad struts replace...
Attitudinal Content Tracking

Analyze in real time high-level social constructs drawn from one of multiple established content analysis dictionaries that can be plugged into system.

Constructs can be incorporated into crawling algorithm inclusion criteria. Raw n-grams also collected and can be incorporated as well.
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