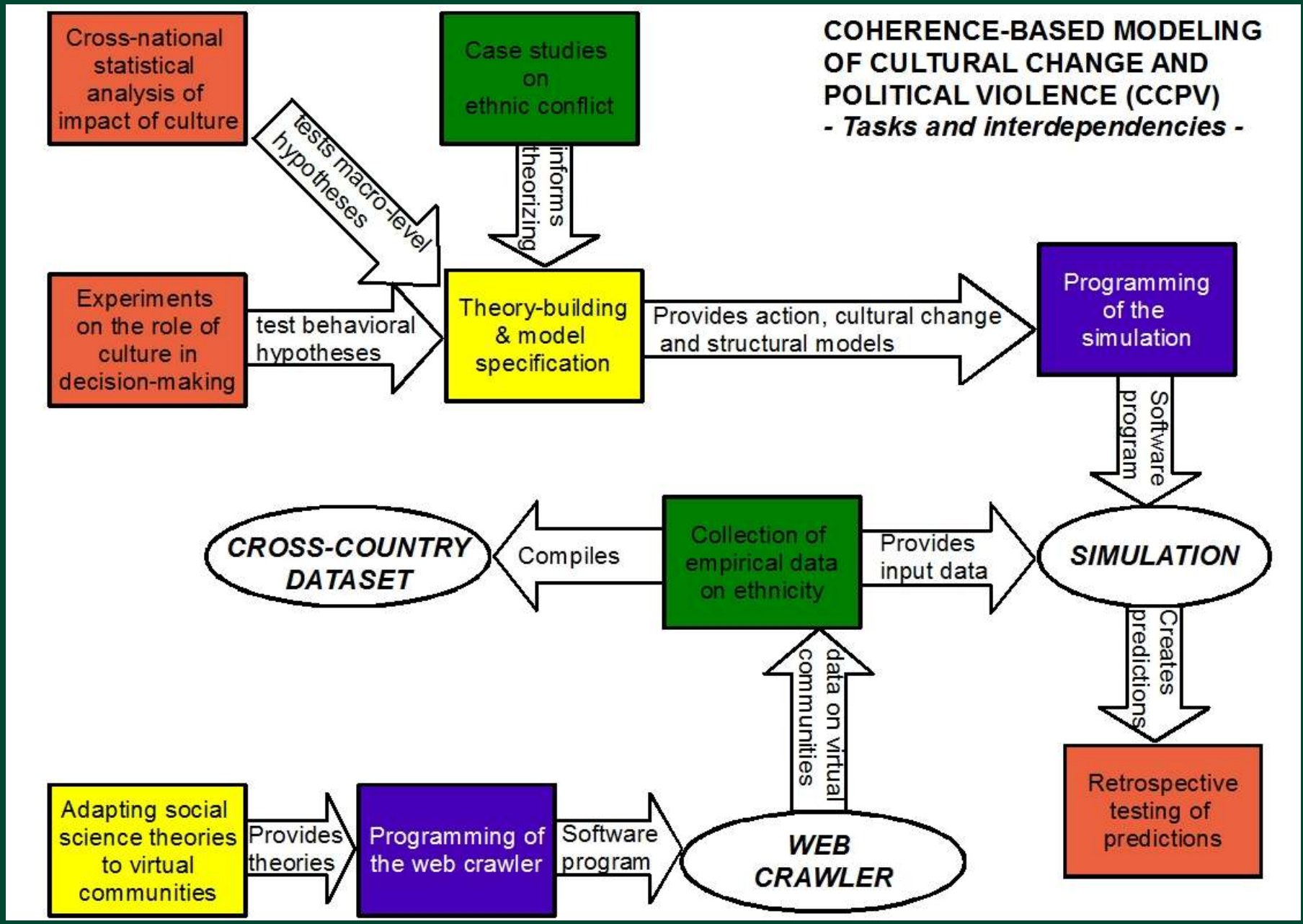
CLASSIC: Software Platform for Solving Real-World Problems by Understanding Virtual Communities

Supported by AFOSR grants FA9550-07-1-0253 (CCPV) and FA9550-09-1-0261; ONR grant N00014-09-1-0805.

Role in CCPV Project



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.

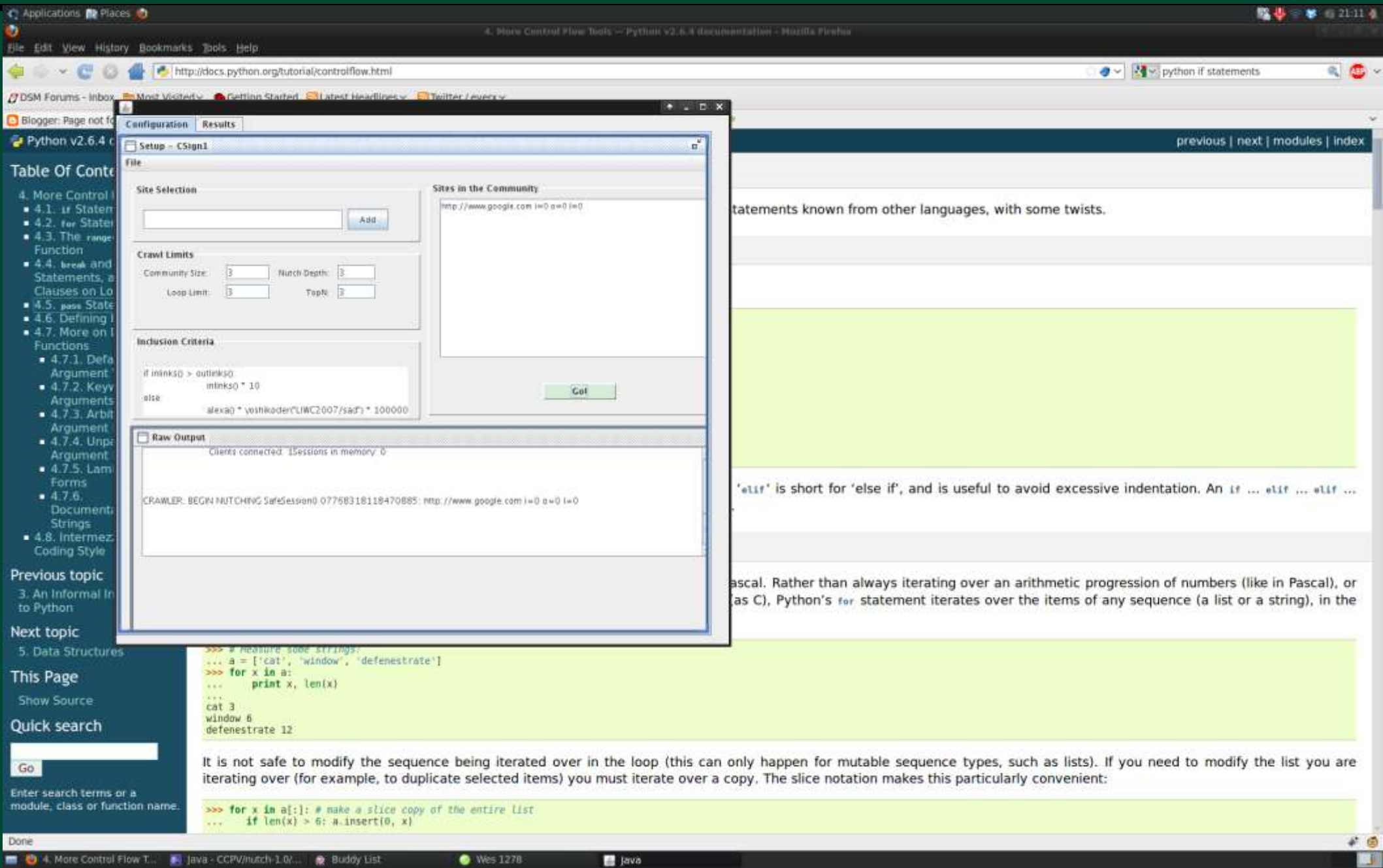
General Kinds of Things You Might Want to Find Out

- 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

CLASSIC: Specification of Seed Sites and Freeform Crawl Criteria



“Intermediate” User Initiation/Runtime/Result

File Options Analysis Tools Help

-=CLA$SIC=-

Seed sites and key terms:

www.example.com, www.seed.jp, www.aSeedSite.org, sports, hiking

Rank results by:

☐ Broker Power ☑ Authority ☐ Similarity
☑ Prestige ☑ Influence ☐ Popularity

Search | Advanced Search Options

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>

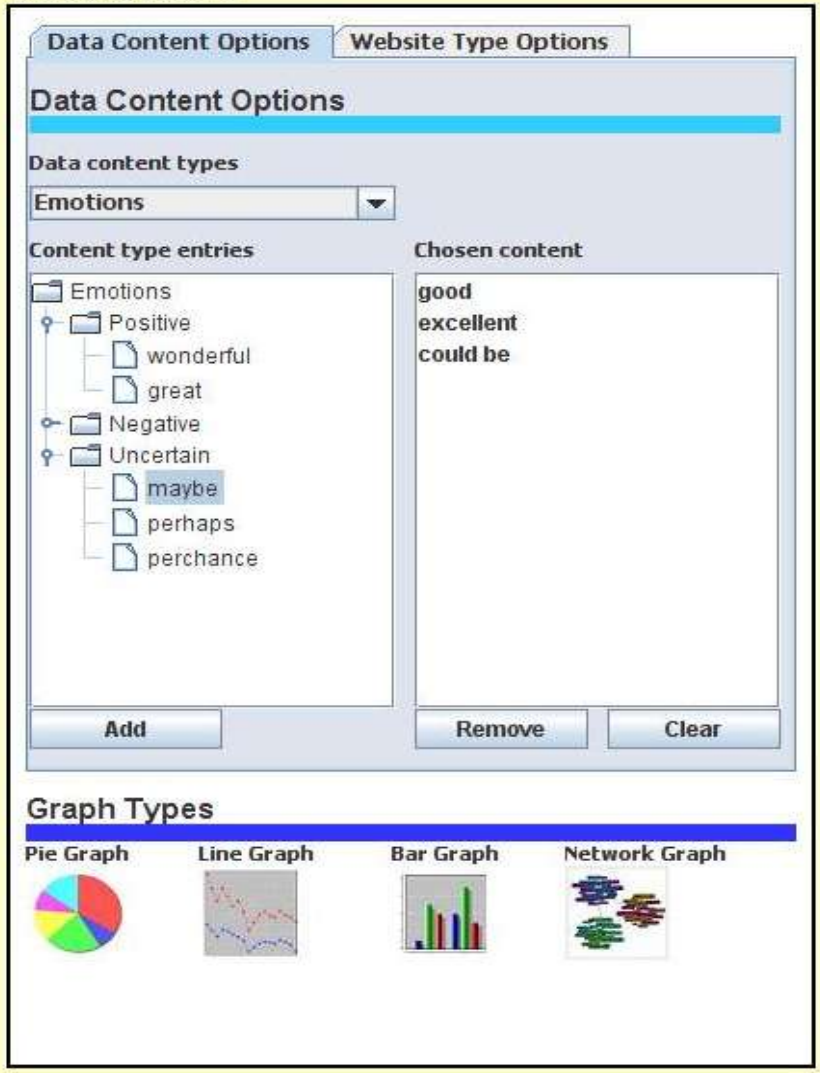
This is an Example web page title
www.exampleURL.com
Put description here. Limit: 1 line.
<more information> <reseed search> <analyze this forum>

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>

Select all | Hide details

Visualizations



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

Using Coherence Model to Predict Preference and Belief Change

Expected Regret (single-period, individual form):

$$d = \int_S (u(s,a^*(s)) - u(s,a))) \, p(s) \, ds$$

where

$$a^*(s) = \operatorname{argmax}_{a \in A} u(s,a)$$

$$a = \operatorname{argmax}_{a \in A,\, s \in S} \int_S u(s,a) \, p(s) \, ds$$

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

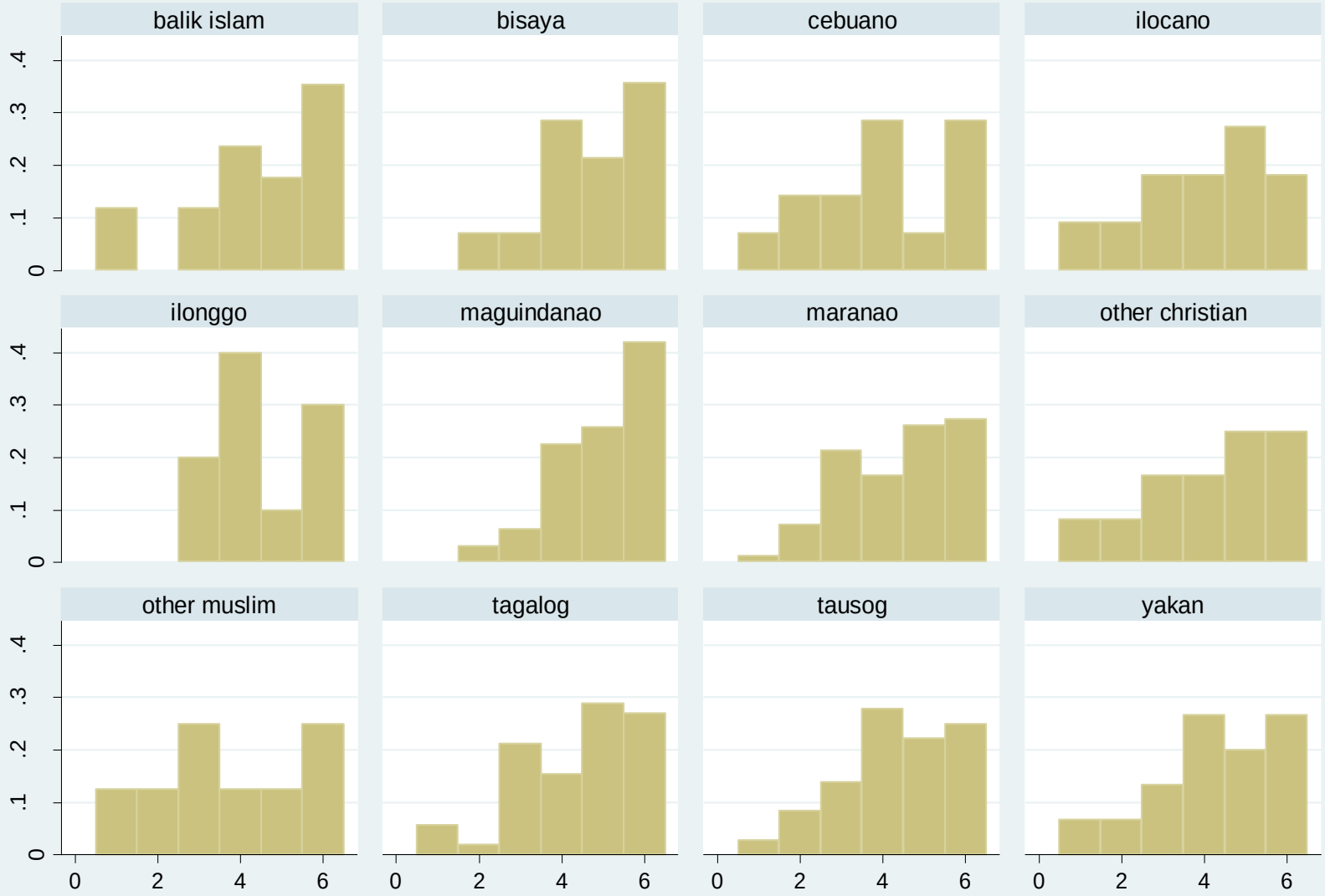
adjustment of ***g, h*** **to minimize** ***d*** **within "non-yogic" utility and information constraints**

A Priori Validation: Grid-Group and Behavior in Voluntary Contributions Mechanism

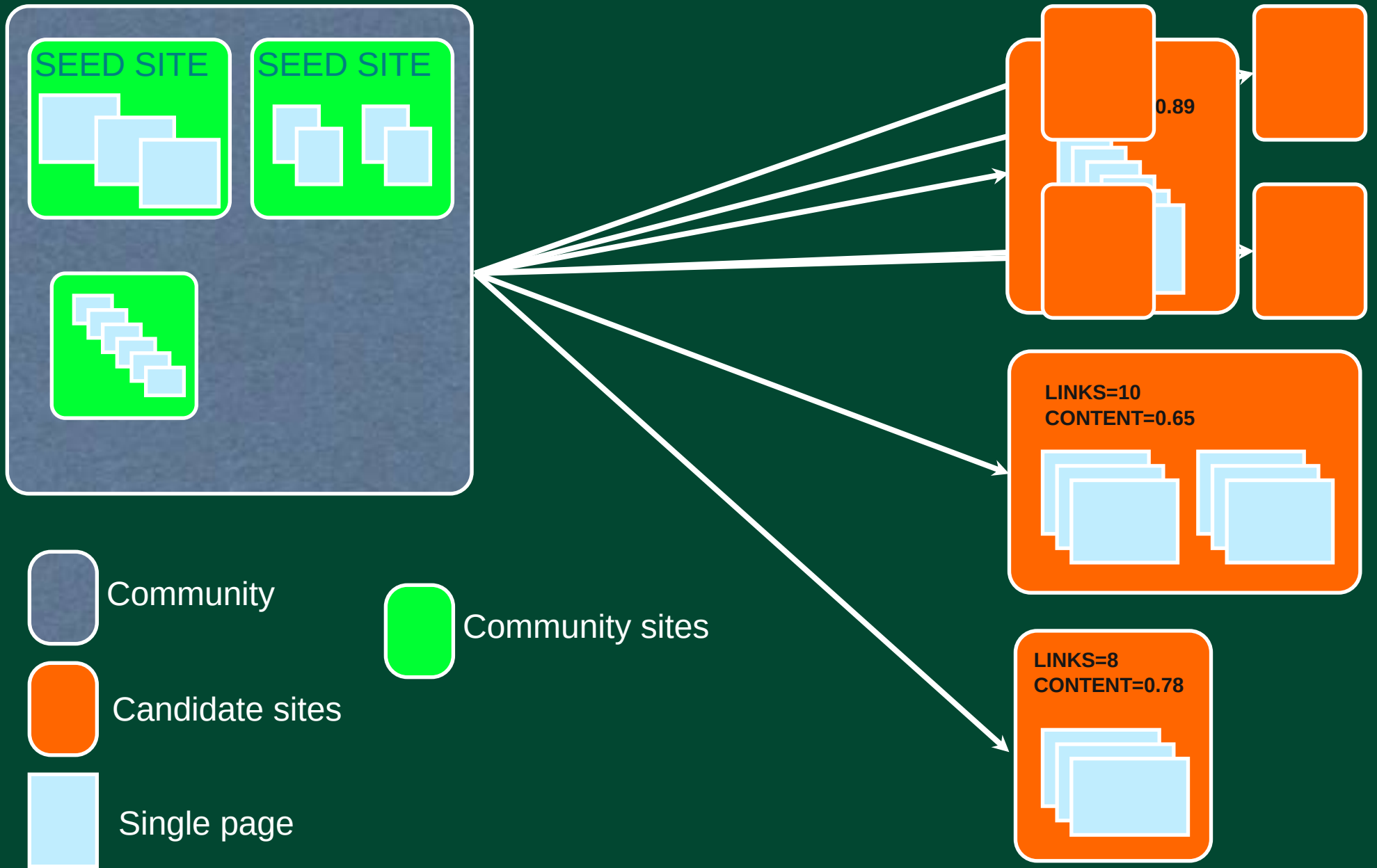
		Shuffled			Partner			Pooled		
		Contribution Level, No Pun	Contribution, with Punish	Punishment Experndit ure	Contribution Level, No Pun	Contribution, with Punish	Punishment Experndit ure	Contribution Level, No Pun	Contribu tion, with Punish	Punishme nt Experndit ure
Grid	Pearson correlation	-0.100	0.038	0.263**	0.261*	-0.075	0.056	0.108	-0.002	0.132
	Sig. (2-tailed)	0.392	0.748	0.022	0.052	0.582	0.683	0.217	0.983	0.131
Group	Pearson correlation	0.350***	0.096	-0.116	0.251*	0.171	-0.218	.319***	0.138	-0.155*
	Sig. (2-tailed)	0.002	0.411	0.318	0.062	0.208	0.106	0.000	0.114	0.077
	N	76	76	76	56	56	56	132	132	132

***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).

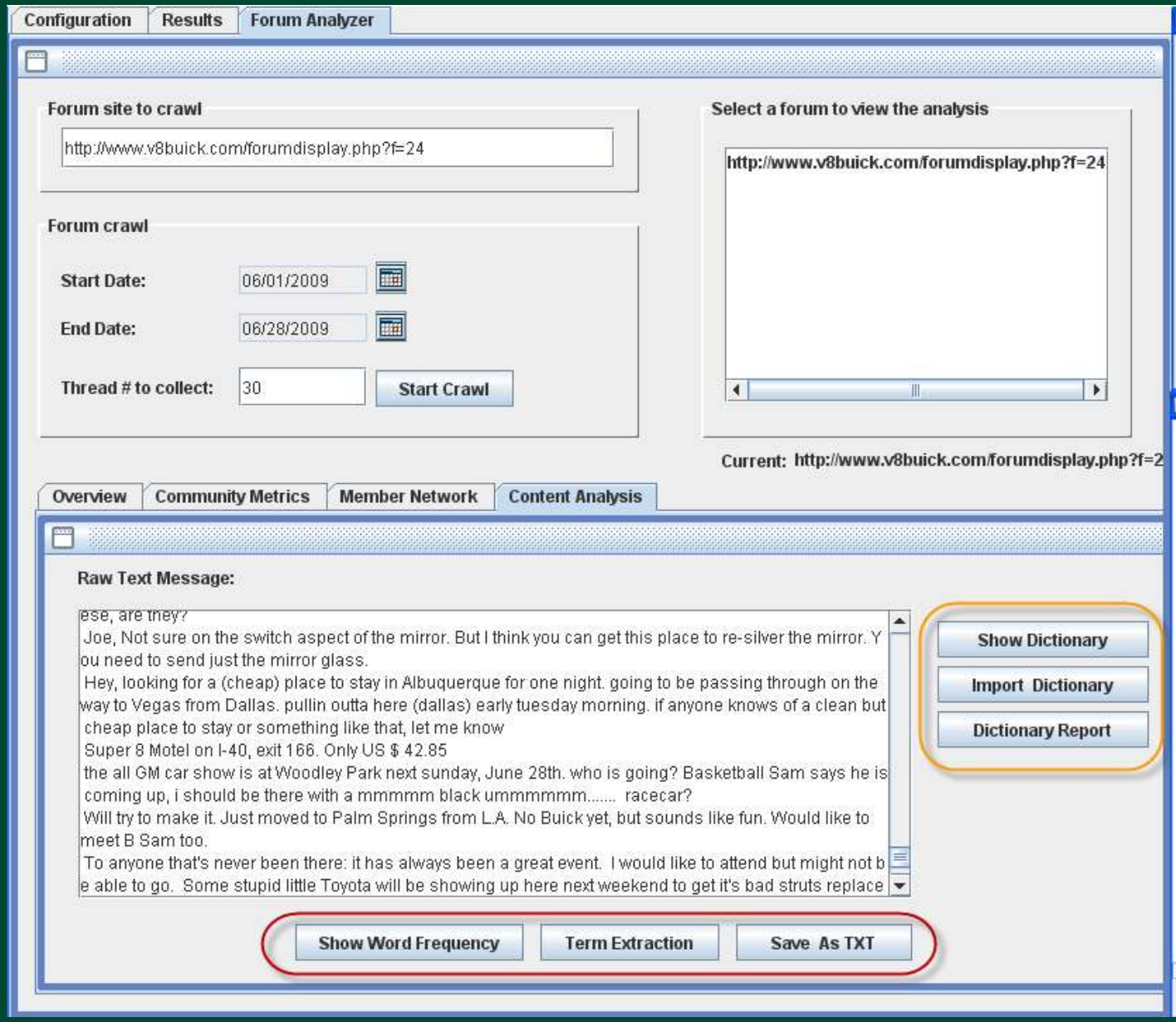
Field Study: Risk Preferences Among Moro Ethnic Groups

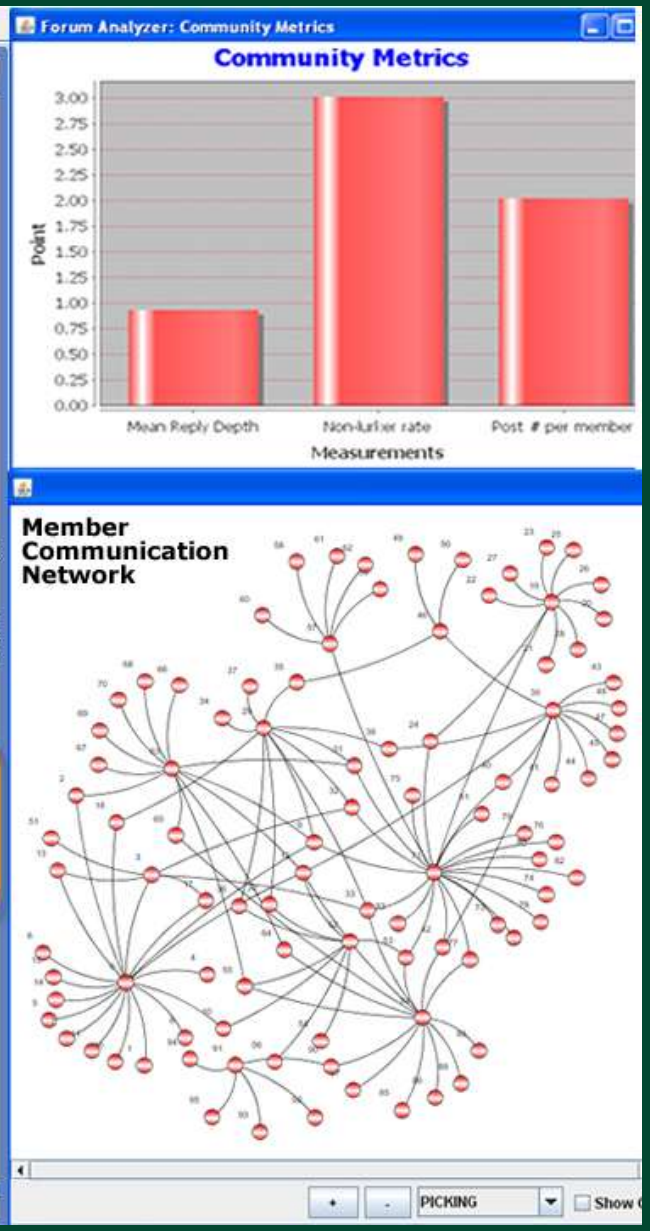


How It Works


SEED SITE
SEED SITE
0.89
LINKS=10
CONTENT=0.65
LINKS=8
CONTENT=0.78
Community
Community sites
Candidate sites
Single page

Forum Analyzer

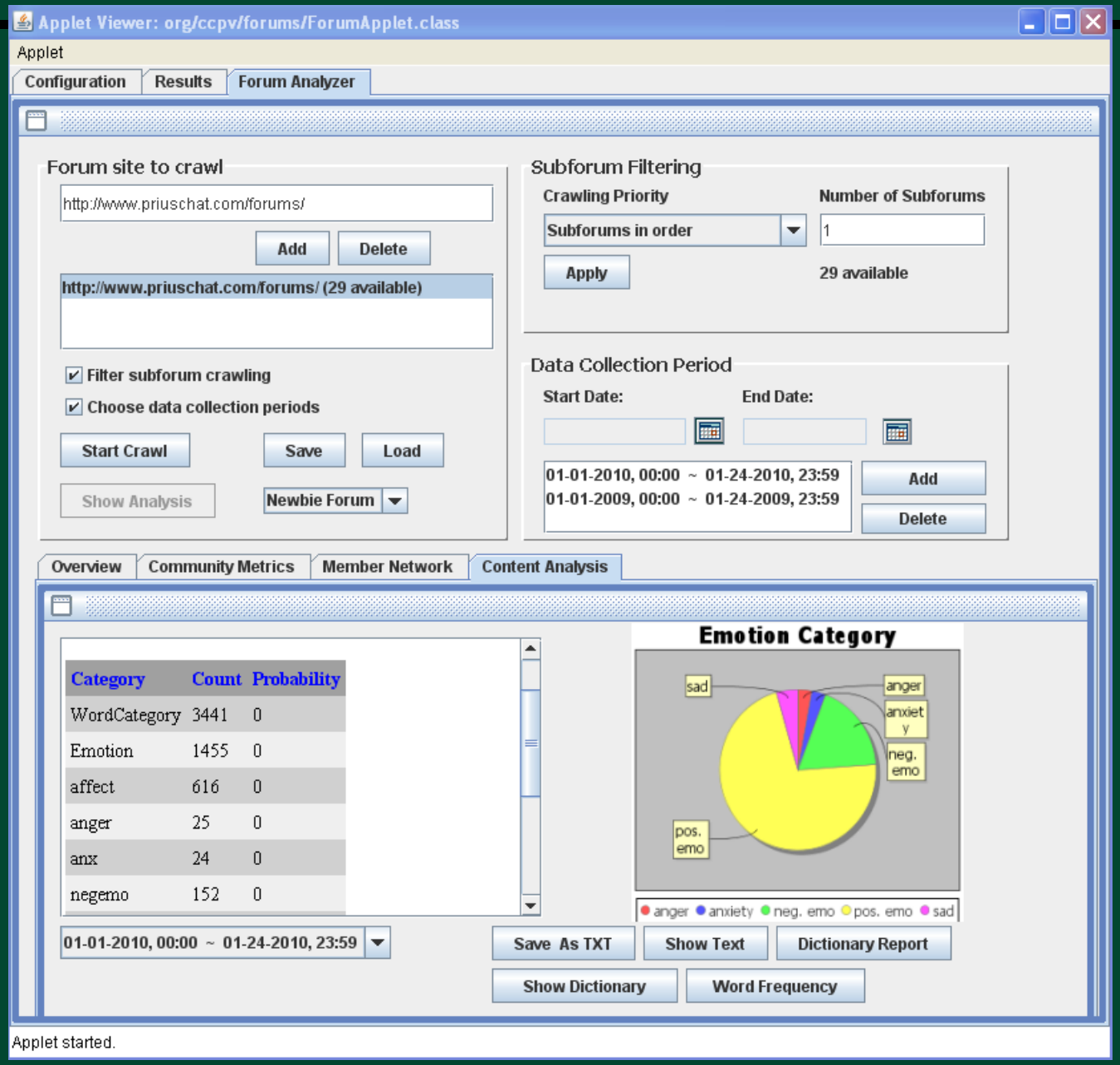




Attitudinal Content Tracking

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

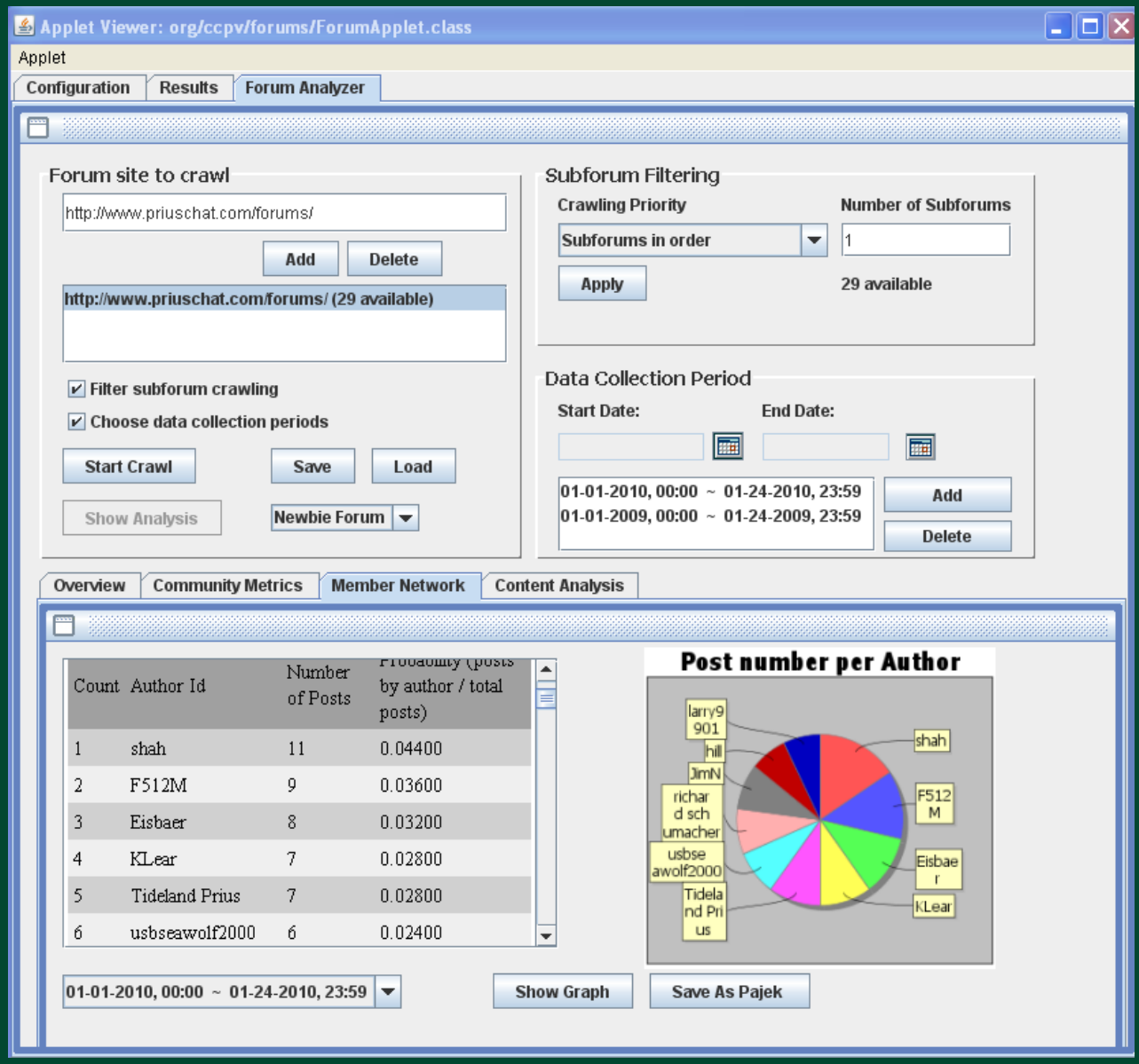
Constructs can be incorporated into crawling algorithm inclusion criteria Raw n-grams also collected and can be incorporated as well.



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