Cultural Consensus Theory: Aggregating Complete Signed Graphs Under a Balance Constraint – Part 2

Kalin Agrawal William Batchelder

Institute for Mathematical Behavioral Sciences,
University of California Irvine

Support is gratefully acknowledged from grants to Batchelder from the Air Force Office of Scientific Research, the Army Research Office, and the Intelligence Advanced Research Projects Activity

International Network for Social Network Analysis
Sunbelt XXXII Conference, Redondo Beach, CA, March 2012

The Model

In case you missed it...

- We ask each of M experts to provide a complete signed response graph on N nodes. Accumulated data is a 3D array $\mathbf{X}$.
- Each $X_{i,jk}$ is expert i's judgment on whether they believe the tie between j and k is positive or negative.
- There is a consensus binary partition $\mathbf{W}$ on which experts base their responses.
- Experts each have ability $\alpha_i \ \epsilon \Re$ and response bias $G_i \ \epsilon (0, 1)$.
- Ties each have difficulty $\beta_{jk} \ \epsilon \Re$.

Parameter Estimation I

Overview of estimation process

- Bayesian inference
- Use the Metropolis-Hastings algorithm
- Sample from the joint posterior parameter space, given the data

Parameter Estimation II

The posterior distribution of the parameters is proportional to the likelihood times the prior,

$$\Pr(\mathbf{W}, \mathbf{G}, \alpha, \beta | \mathbf{X}) \propto L(\mathbf{X} | \mathbf{W}, \mathbf{G}, \alpha, \beta) \pi(\mathbf{W}, \mathbf{G}, \alpha, \beta) \ . \tag{1}$$

$L(\mathbf{X} | \mathbf{W}, \mathbf{G}, \alpha, \beta)$ is the response model as a likelihood function.

$\pi(\mathbf{W}, \mathbf{G}, \alpha, \beta)$ is the joint prior on the partition, expert biases and abilities, and tie difficulties.

Parameter Estimation III

The response model was given in the previous talk, so omitted here.

Independent priors,

$$\pi(\mathbf{W}, \mathbf{G}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{i=j}^{N} \pi(W_j) \prod_{i=1}^{M} \pi(G_i) \prod_{i=1}^{M} \pi(\alpha_i) \prod_{j=1}^{N-1} \prod_{k>j}^{N} \pi(\beta_{jk}) .$$

- Node cell assignment: $W_j \sim \mathrm{Bernoulli}(1/2)$ ("flat")
- Expert guessing bias: $G_i \sim \mathrm{Uniform}(0, 1)$ ("flat")
- Expert ability: $\alpha_i \sim \mathrm{Normal}(0, 1)$
- Tie difficulty: $\beta_{jk} \sim \mathrm{Normal}(0, 2)$

Parameter Estimation IV

- Componentwise univariate updates for continuous parameters (response bias, ability, tie difficulty).
  - Use standard candidate generators: Beta, Normal, Normal.
- Blockwise updates for the discrete partition use a "defection" idiom for candidate generation.
  1. Randomly choose one the two cells.
  2. Randomly select nodes from the chosen cell to move to the the other.

Recovery of Parameters Used for Generating Simulated Data

Using multiple generated datasets with various parameter values, we were able to recover parametercs using our estimation procedure.

Note: The ability to recover parmeters depends on the detectability of the ties.

- Recall from previous talk: the probability of a single detection is $D_{i,jk}$, with average tie detectability across experts given by $\overline{D}_{jk}$, and overall average $\overline{D}$.

Recovery of Consensus Partition I

An example:

Use a generating average α_i of 0 and β_{jk} of 1.5, with average generating detectabiliity of $\overline{D} = 0.18$.

$M = 12$ experts and $N = 6$ nodes (15 ties). Cell labels "a" and "b".

Estimated posterior distribution of the partition:

Partition	Freq	Modal	Generating
aaabbb	9613	*	$<$
aaaaaa	283		
aaaabb	70		
abbaaa	20		
abbbbb	14		

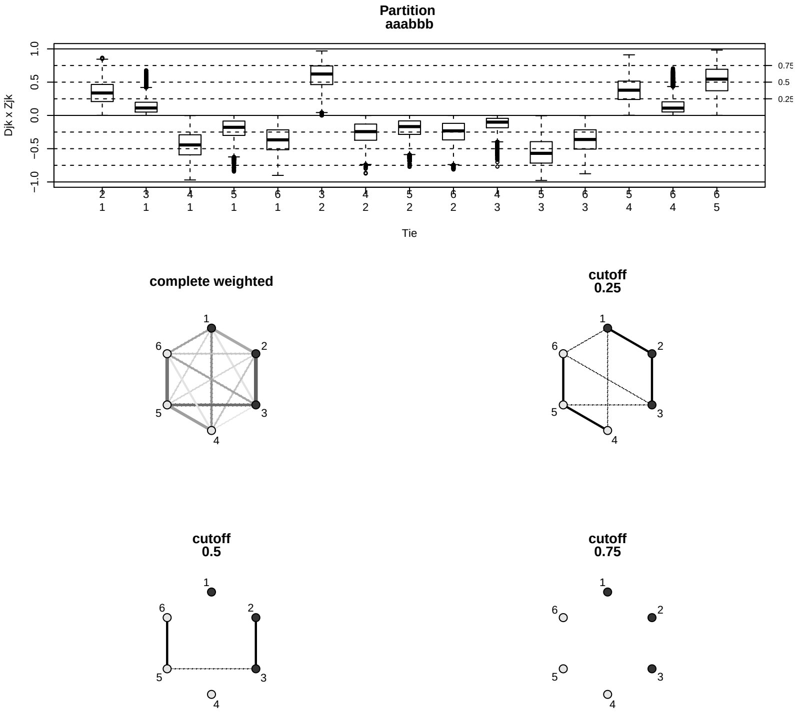
Recovery of Consensus Partition II

The modal partition, aaabbb, implies a complete signed graph.

But what if we want a more nuanced graph?

Use posterior distribution of $\overline{D}_{jk}$, conditioned on the partition aaabbb to make a *weighted signed graph*.

Adjusting a threshold (“turning a knob”) makes this into a discrete graph.


Partition
aaabbb
Djk x Zjk
1.0
0.5
0.0
−0.5
−1.0
0.75
0.5
0.25
2 3 4 5 6 3 4 5 6 4 5 6 5 6 6
1 1 1 1 1 2 2 2 2 3 3 3 4 4 5
Tie
complete weighted
cutoff
0.25
cutoff
0.5
cutoff
0.75
1
2
3
4
5
6

Applied to real data I

We issued a survey to 19 students, presenting all 45 pairs of names of 10 famous sports players.

5 were basketball “MVP”s and 5 were baseball “MVP”s.

For each tie we asked, “Do they play the same sport?” and elicited their confidence.

Using the design mentioned earlier to make it hard for experts to logically infer correct answers: Q1-9 completed no cycles, Q10 completed one long cycle.

Applied to real data II

Elicited confidences:

Don't know	516
Unsure	179
Certain	148
N/A	12

Correctly recovered the true partition.

Mean marginal **W**, Q1-10: (1, 1, 1, 1, 1, 2, 2, 2, 2, 2)

Mean marginal **W**, Q1-45: (1, 1, 1, 1, 1, 2, 2, 2, 2, 2)

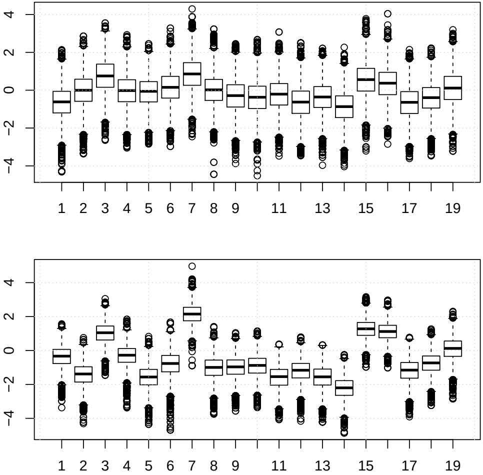
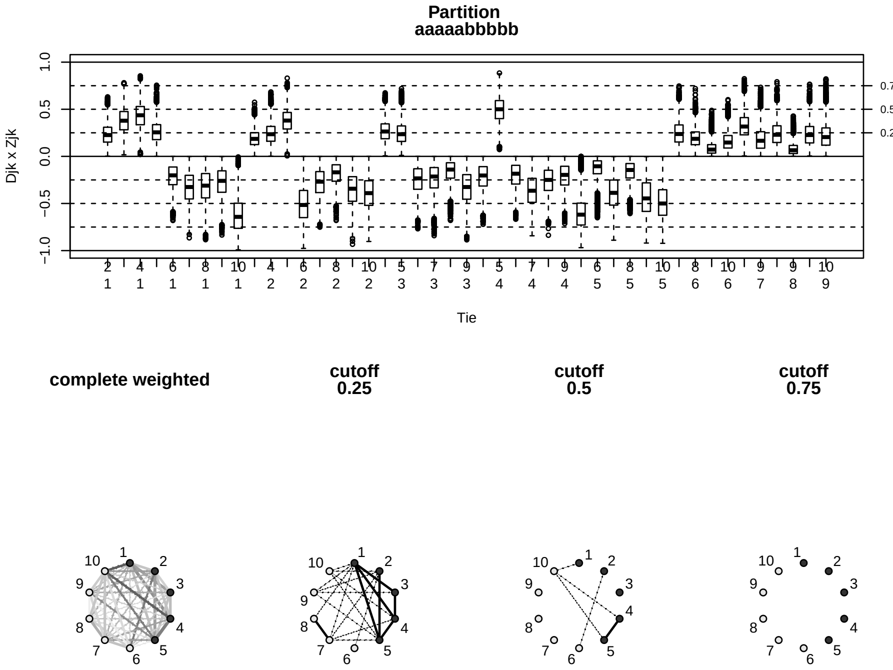


Figure: Marginal posterior α, Q1-10 (top), Q1-45 (bottom).


Partition
aaaaabbbbb
Djk x Zjk
Tie
complete weighted
cutoff
0.25
cutoff
0.5
cutoff
0.75

Compare the observed number of unbalanced triads against the expected number of unbalanced triads assuming independent responses.

Expected value is Chi-square distribution with 19 df, and $p = 0.95$ critical value of 30.14.

Observed value was 71.03 (signficant).

Discussion I

Caveat: The data was based on the question of whether two nodes belong to the same class or not.

Might expect an expert who knows the cell for a node will apply that knowledge to judging all ties with that node.

We have begun collecting complete-graph tie-based data, but would appreciate your suggestions.

Working on a “nodal” response model, incorporating a more complicated sampling mechanism. (Ask about this afterwards if you are interested.)

Discussion II

Added a plausible hard constraint (balance) on CCT for social networks.

The approach is open to more complicated response models as well as network models (e.g. model homophily, cultural mixtures)

We are exploring other constraints on pairwise relations.

Thanks!

Feel free to email us: kagrawal@uci.edu, whbatche@uci.edu.