# Quantitative Analysis of Interaction Patterns in Chat

**Rationale:** The purpose of this branch of the VMT project is to conduct a micro-level analysis of collaborative math problem solving
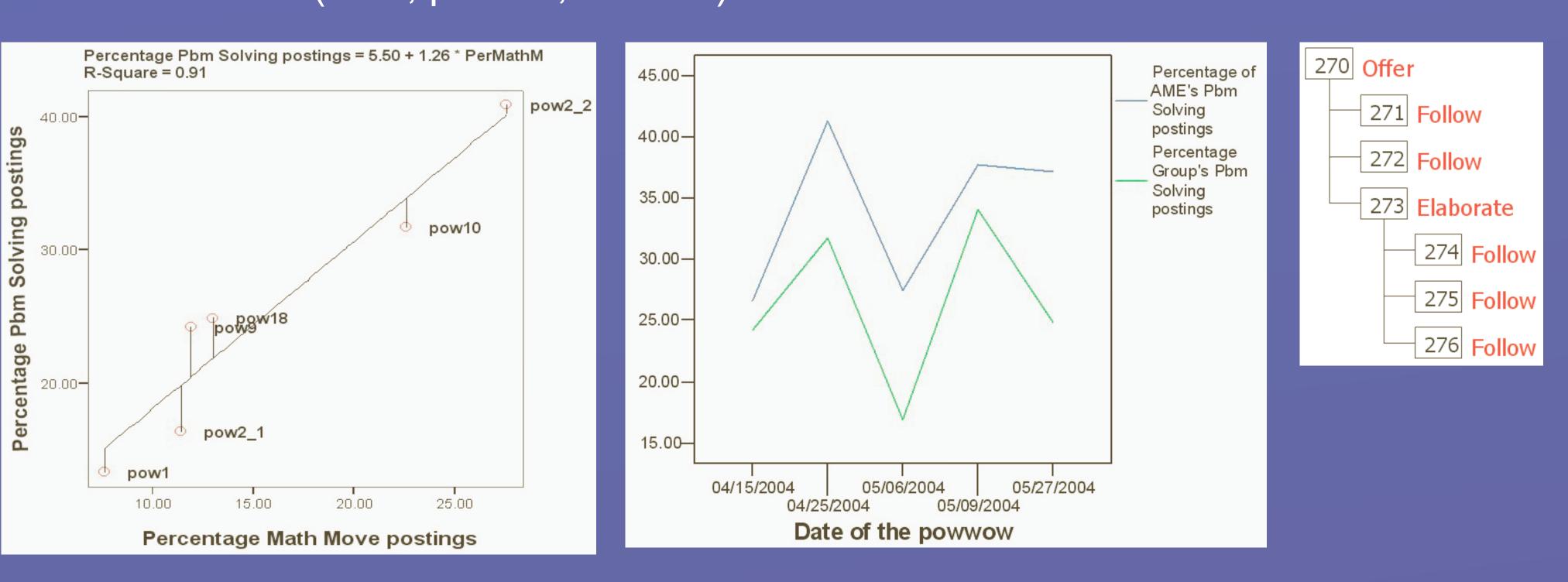
## The Virtual Math Teams Coding Scheme

* For quantitative analysis of collaborative interactions, chat transcripts are segmented and coded in six dimensions, noting thread links between postings.

* Multi-dimensional coding scheme design:
  - Segmentation: a chat line is selected as the unit of analysis
  - Coding: categorization along multiple dimensions
    1. Conversation (based on Beers et al., 2004)
    2. Social (based on Renninger & Shumar, 2002)
    3. Problem-solving (based on Polya, 1957)
    4. Math Moves (mathematical procedures/calculations)
    5. Math Progress (evaluation of the group's problem solving steps)
    6. Support (system support and facilitation)
  - Threading: postings are linked based on:
    1. Conversational content (e.g. adjacency pairs, utterances extended to multiple chat lines)
    2. Problem solving characteristics (e.g. carrying out a previously proposed strategy)

* 10 transcripts were coded to aid explorative analysis

* Initial inter-coder reliability results: 50%-92%, depending on dimension



## Explorative Statistical Analysis

* Uncovering the structure of transcripts across dimensions
  - We have found that the distributions of problem solving and math move codes exhibit a good approximation of the normal distribution

* Exploring similarities and differences across transcripts
  - The clustering results are interpreted in terms of the way participants organize their interaction (based on expository versus exploratory modes of interaction)
  - For each interaction mode we are devising alternative representations in terms of codes to cluster the groups based on their interaction styles.

* Tracking the collaboration activity of a student in different chat sessions

* Studying the correlation and regression between different dimensions.
  - There is a strong negative correlation between social reference and math specific dimensions (Pearson correlation: -.942 ($p<0.05$) & .-970 ($p<0.01$))
  - There is a strong positive correlation between problem solving and math move dimensions (.967, $p<0.01$, 2-tailed)
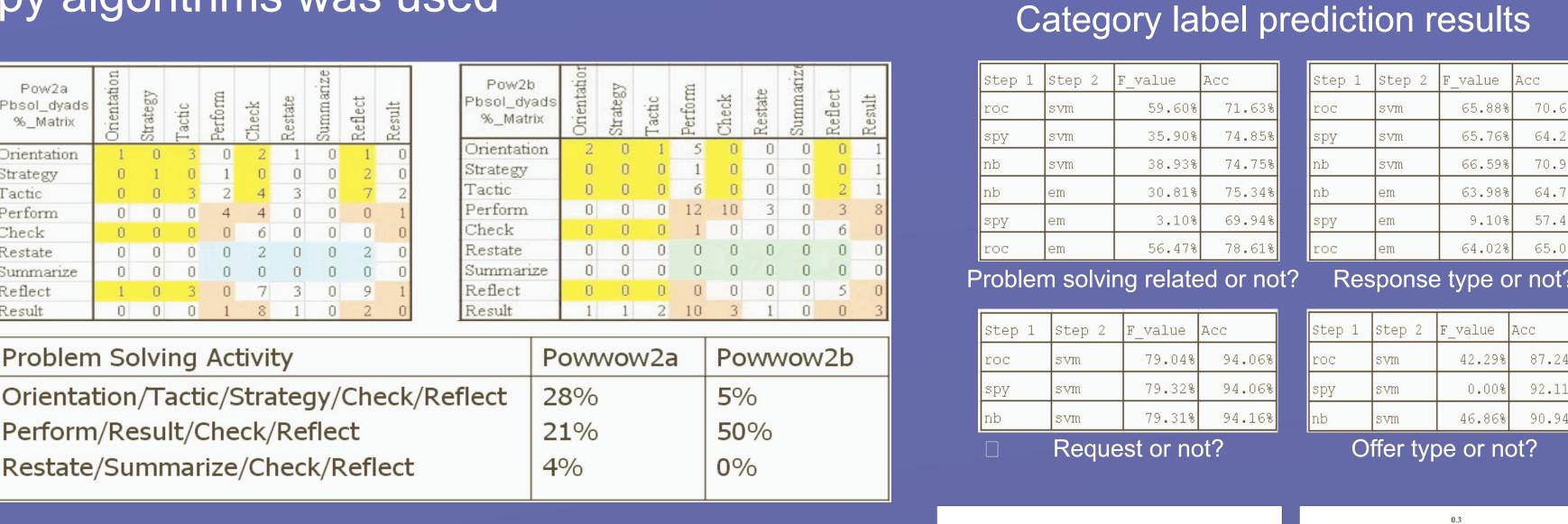


Note: Explorative analysis is performed on a selection of 6 transcripts. In 3 cases the problem was revealed at the beginning of the session, whereas in the latter cases the participants had a chance to work on the problem in advance
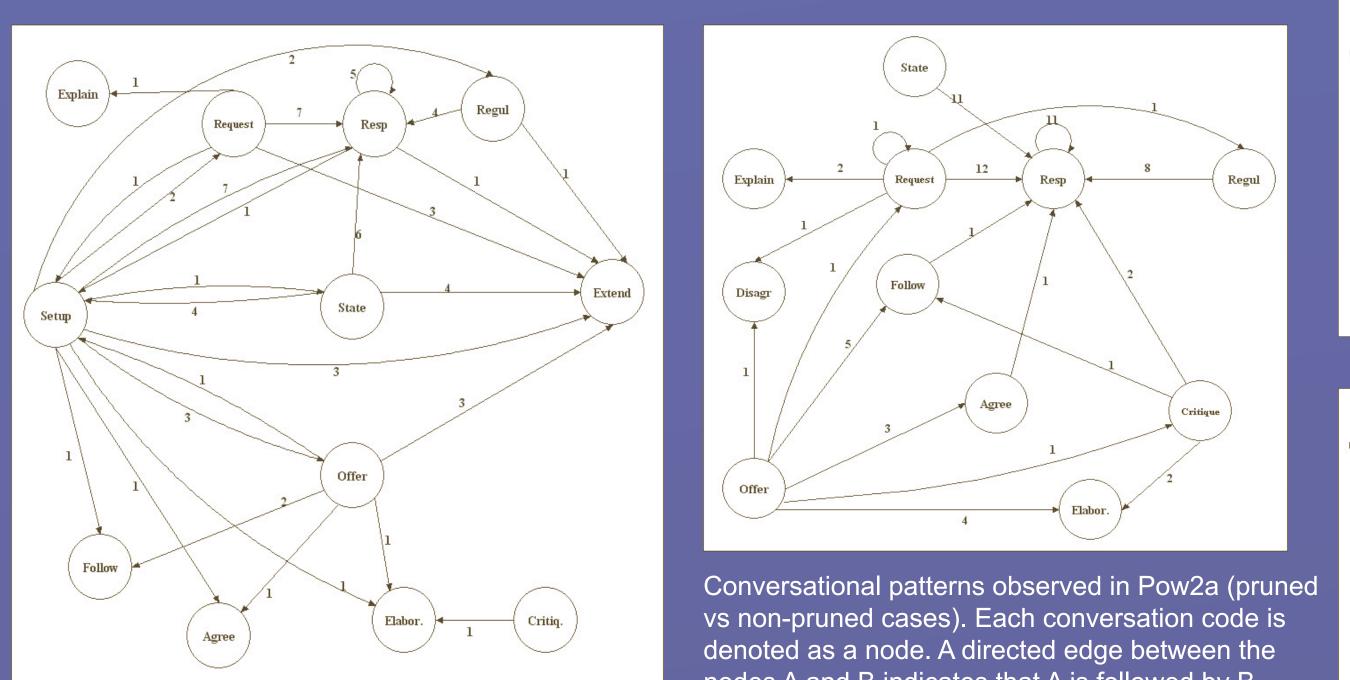
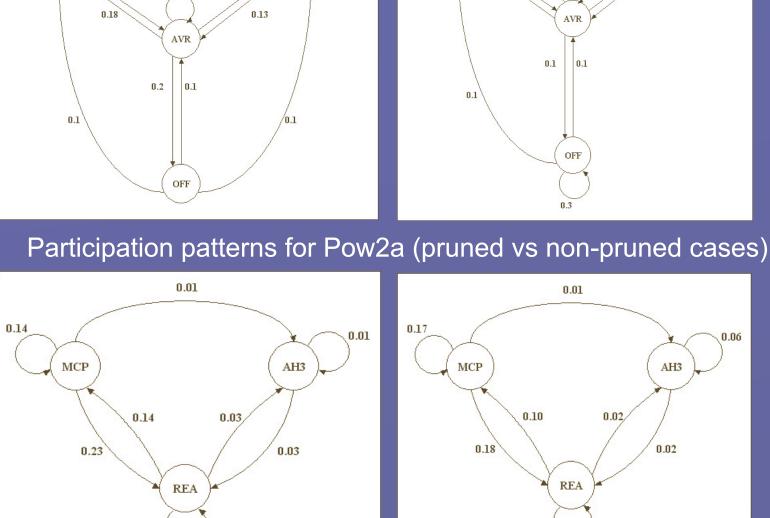## A Thread-based Scheme for Mining Interaction Patterns

* A computational model for performing sequential pattern mining
  - Frequent conversational patterns (source-sink pairs)
  - Participation patterns - e.g. who interacted with whom, who initiated more threads of discussion, who showed more attention to other members' postings
  - Patterns in problem solving activity of a group (tied to expository vs exploratory modes)

* Experiments with text mining tools to automatically assign code labels to postings
  - Preprocessing: setup-extension pruning, stemming, stop word removal
  - The LPU package containing implementations of SVM, Naive Bayesian, Rocchio and Spy nd Spy algorithms was used



| Problem Solving Activity | Powwow2a | Powwow2b |
|---|---|---|
| Orientation/Tactic/Strategy/Check/Reflect | 28% | 5% |
| Perform/Result/Check/Reflect | 21% | 50% |
| Restate/Summarize/Check/Reflect | 4% | 0% |



Conversational patterns observed in Pow2b (pruned vs non-pruned cases). Each conversation code is denoted as a node. A directed edge between the nodes A and B indicates that A is followed by B. Weights correspond to percentage values.

Participation patterns for Pow2a (pruned vs non-pruned cases)

Participation patterns for Pow2b (pruned vs non-pruned cases)

## Findings - Significance:

* We propose a thread-based analytical scheme to conduct a sequential analysis of interaction in chat environments (Cakir et al.). We employ this scheme in order to remedy the confounding effects of artificial turn orderings introduced by text-based chat systems.

* The initial results we obtained from clustering and conversation analysis of transcripts led us to the distinction between expository versus exploratory modes of group interaction.

* Existence of complex sequential relationships among postings, fragmented turn organizations, chat specific acronyms, and misspelled words make it extremely difficult to design automated schemes for processing chat data. Our initial results with popular text mining algorithms demonstrate the potential of better schemes that can go beyond keyword (i.e. content) based algorithms by considering the sequential relationships between postings.

## Next Steps:

* We are in the process of building a statistical model to study interaction patterns we have found so far from a statistical perspective.

* We will further improve our quantitative analysis schemes based on the results we obtain from conversation analytic and ethnographic studies conducted at the VMT project.

### Reference:

Cakir, M., Xhafa, F., Zhou, N., and Stahl, G. (2005). Thread-based analysis of patterns of collaborative interaction in chat. To appear in the proceedings of AI in Education 2005 conference, Amsterdam, Netherlands.

**Principal Investigators:** Gerry Stahl, Stephen Weimar, Wesley Shumar
**Researchers:** Jan-Willem Strijbos, Stefan Trausan-Matu. Fatos Xhafa, Alan Zemel
**Doctoral Students:** Murat Cakir, Johann Sarmiento, Ramon Toledo, Nan Zhou

The Math Forum • DREXEL UNIVERSITY

Drexel UNIVERSITY

College of Information Science & Technology

NSF — National Science Foundation