Solving Mysteries with the Wisdom of Crowds: a Modularized Pipeline and Context Slices

Tianyi Li
Virginia Tech
Blacksburg, VA
tianyili@vt.edu

ABSTRACT

The increasing volume of text data is challenging the cognitive capabilities of experts. Machine learning and crowdsourcing present opportunities for large-scale distributed sensemaking, but we must overcome the challenge of modeling the holistic process so that many distributed agents can contribute to suitable components asynchronously and meaningfully. My dissertation work is devoted to addressing this challenge from a crowdsourcing perspective. Specifically, I study 1) how novice crowds can build theories from raw datasets without expert intervention; 2) what bottlenecks exist for crowds in the theory-building process; and 3) how previous crowd analyses can be refined to enable iterative crowdsourced sensemaking.

CCS CONCEPTS

• **Human-centered computing → Collaborative and social computing systems and tools**; *Computer supported cooperative work*; *Empirical studies in collaborative and social computing*.

CSCW '19, November 09–13, 2019, Austin, TX

ACM ISBN 978-1-4503-9999-9/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

KEYWORDS

Sensemaking; Text Analytics; Intelligence Analysis; Mysteries; Crowdsourcing; Empirical investigations

ACM Reference Format:

Tianyi Li. 2019. Solving Mysteries with the Wisdom of Crowds: a Modularized Pipeline and Context Slices. In *CSCW 2019 Doctoral Consortium.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/1122445.1122456

INTRODUCTION

Making sense of large datasets, in which analysts must sort through many snippets of textual information to identify a latent plot, challenges the scalability of traditional expert processes. My dissertation is devoted to exploring how this process can be modularized to support distributed, asynchronous collaboration among novice crowds. Specifically, I focus on the case of solving a mystery, such as identifying the suspect in a murder case or the target in a terrorist attack.

Sensemaking is a highly integrated cognitive process. Analysts iteratively forage, schematize, and synthesize information in an ad hoc manner [4] that is difficult to formalize into one single workflow of microtasks for novice crowd workers [5]. In addition, sensemaking requires a holistic view of the data. While aggregating local views can produce a synthesized view of the dataset [1], it introduces biases and mistakes when used uncover hidden relationships and make decisions [6]. The two problems intersect and cannot be solved independently in the crowdsourcing setting. For each modularized process, the information should also be modularized for individual crowd workers to contribute.

My dissertation work aims to confront this double modularization challenge and further alleviate the burden on experts. I divide my dissertation into three main phases: 1) *constructing* a crowdsourced sensemaking pipeline [2], 2) *analyzing* errors and bottlenecks in crowdsourced sensemaking [3], and 3) *formulating* the refining process based on the pipeline (in progress).

COMPLETED WORK

Phase 1: Building the CrowdIA system

My phase 1 research focused on designing a pipeline of modularized steps connected by clearly defined inputs and outputs [2]. The pipeline builds on the expert sensemaking process [4] and partitions information into "context slices" for individual workers in each step (Figure 1). I implemented CrowdIA, a software platform to enable unsupervised crowd sensemaking using our pipeline. CrowdIA successfully guided the crowd to solve two mysteries, and were one step away from solving the third. The crowd's intermediate results also revealed their reasoning process and explained their conclusions.

Research Questions for Phase 1

The first phase focuses on designing and experimenting with a pipeline that guides the crowd to develop a theory from the source documents without expert intervention.

We explored the following research questions in [2]:

- RQ1-1: To support crowds, how can we formally modularize the sensemaking process into a series of steps that each defines the information needs (*Step Input*) and intermediate analysis results (*Step Output*)?
- RQ1-2: Within each step, how do we slice the *Step Input* into contextualized microtasks for individual crowd workers, and aggregate the local analysis results into *Step Output*?
- RQ1-3: How well do crowds perform in solving mysteries with the modularized sensemaking process, and specifically, how do crowds perform in each step?

Research Questions for Phase 2

The second phase evaluates the crowdsourced sensemaking process to understand the errors and bottlenecks. This further assesses the proposed pipeline and educates the refining process design.

We explored the following research questions in [3]:

RQ2-1: What are the errors (type and frequency) workers make in a crowdsourced sensemaking pipeline, both within each step and across steps?

RQ2-2: How does the amount of local context affect the errors within and across steps in a crowdsourced sensemaking pipeline?

Research Questions for Phase 3

The third phase investigates how to refine the previous analysis with the pipeline.

I plan to explore the following research questions:

RQ3-1: How well can experts locate errors in crowd analysis provenance?

RQ3-2: How well can the crowd refine previous analysis with expert guidance?

Phase 2: Understanding CrowdIA bottlenecks

Interconnecting the inputs and outputs in different crowdsourcing processes introduces another level of complexity and is subject to errors and bottlenecks. The mixed-quality crowd performance in phase 1 also points to this challenge. My phase 2 research probes the errors and bottlenecks by analyzing crowd performance with different levels of quality (gold-standard input versus previous crowd-generated input) and quantity (number of items in each microtask) of input data.

The results indicate that while errors happen in each step and propagate to later steps, the crowd had varying performance in different steps of the pipeline. Surprisingly, some errors were mitigated in later steps without expert intervention. This can help prioritize the refining process. My results also shed light on the trade-offs between increased local context and analysis quality, which can inform the design of refining microtasks.

WORK IN PROGRESS

Drawing on these lessons learned, the last part of my dissertation (phase 3) explores how the pipeline could support refining existing crowd analyses and complete the iterative sensemaking process.

Phase 3: Enabling expert-crowd collaboration

The main challenge in designing a refining path of the pipeline is that without a gold-standard solution, we can only judge the logic reasoning but not the correctness of crowd analysis. Thus, I scope the refining path research with the goal to surface and structure the relevant evidence that helps solve a mystery. Specifically, the refining path needs to support 1) filling the information holes about the case (looking for relevant information missed in previous analysis and fitting in the known knowledge); and 2) removing the irrelevant information from the previous analysis to eliminate distraction.

The design space of a refining path can be described with a 2×2 matrix of *who* to refine (expert vs. crowd) and *how* to refine (provide feedback vs. execute feedback). Leaving the entire refining task to experts could lead to another complicated project about expert-crowd systems and is not as relevant to my dissertation topic. A fully crowd-powered approach is challenged by the lack of a global view of the analysis provenance and proved difficult in my pilot studies. Such being the case, I plan to focus on a refining process where experts provide feedback and guide crowds to refine the previous analysis.

Study 1: Helping experts locate errors in crowd analysis provenance. I plan to first conduct a study to assess expert feedback on crowd errors. This involves designing and implementing an expert interface to represent the crowd analysis provenance in the pipeline. By analyzing how many of the crowd errors experts find and how well experts can locate the origins of errors, I can iterate on the system design to facilitate efficient navigation of previous crowd analysis.

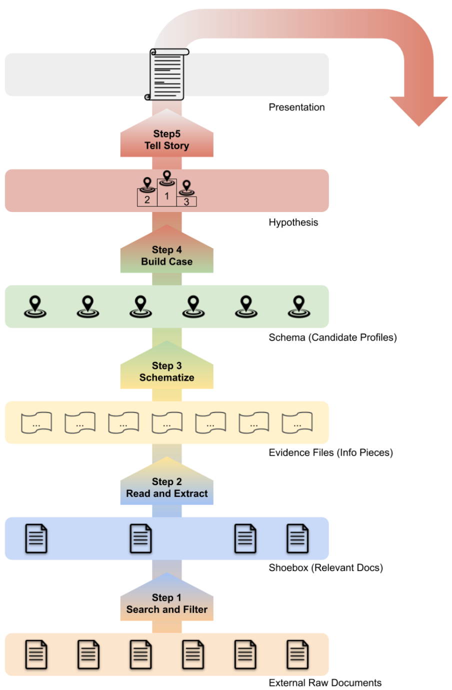


Figure 1: The CrowdIA pipeline transforms raw documents to a final presentation.
Step 1: Crowds rate document relevance.
Step 2: Crowds extract important information pieces from relevant documents.
Step 3: Crowds tag the information pieces with potential answers and evidence types. Tagged info pieces are organized into profiles of candidate answers.
Step 4: Crowds rank candidate profiles by likelihood of being the correct answer.
Step 5: Crowds write a narrative with the information in the top-ranked profile.

Study 2: Evaluating crowd refinement with expert guidance. I plan a second study to evaluate crowd performance in addressing expert feedback. Given the ongoing, iterative nature of intelligence analysis, I will scope my work to one batch of refinement to address the question of "when to stop." I structure expert feedback as context, critique, todo to support crowds in refining analysis. Crowd refinement can be evaluated by comparing to the gold-standard solutions and expert assessment.

Expected contributions

My dissertation aims to contribute: 1) a modularized sensemaking pipeline to guide crowdsourced sensemaking, implemented as the CrowdIA system; 2) analyses of the bottlenecks and the error propagation in holistic crowdsourced sensemaking; 3) a refining strategy and workflow to improve and utilize crowd analyses. I hope that the CrowdIA pipeline can open up the sensemaking process to enable design and evaluation of novel systems at a finer granularity. I expect my findings to further our understanding of the opportunities and limitations of incorporating crowdsourcing efforts into complex problem-solving and help integrate existing research efforts to systematically augment intelligence using different agents in a more decentralized and scalable manner.

GOALS FOR CSCW DOCTORAL CONSORTIUM

My first goal is to gather feedback on the framing and direction of my dissertation proposal. In the final year of my program, I need fresh perspectives on both the coherence and gaps in my proposed work. Specifically, I’m interested in suggestions regarding my upcoming refining path research. My second goal is to connect to senior researchers and seek for career mentoring. I’m applying for both academia and industry research positions and I hope to learn from successful experiences in both worlds to broaden my horizon with job search and career planning.

REFERENCES

[1] Nathan Hahn, Joseph Chang, Ji Eun Kim, and Aniket Kittur. 2016. The Knowledge Accelerator: Big picture thinking in small pieces. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2258–2270.
[2] Tianyi Li, Kurt Luther, and Chris North. 2018. CrowdIA: Solving Mysteries with Crowdsourced Sensemaking. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 105 (Nov. 2018), 29 pages. https://doi.org/10.1145/3274374
[3] Tianyi Li, Chandler J. Manns, Chris North, and Kurt Luther. 2019. Dropping the baton? Understanding errors and bottlenecks in a crowdsourced sensemaking pipeline. *Proceedings of the ACM on Human-Computer interaction* CSCW (2019).
[4] Peter Pirolli and Stuart Card. 2005. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proc. international conference on intelligence analysis*, Vol. 5. McLean, VA, USA, 2–4.
[5] Daniela Retelny, Michael S Bernstein, and Melissa A Valentine. 2017. No workflow can ever be enough: How crowdsourcing workflows constrain complex work. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 89.
[6] Yla Tausczik and Mark Boons. 2018. Distributed Knowledge in Crowds: Crowd Performance on Hidden Profile Tasks. In *Twelfth International AAAI Conference on Web and Social Media*.

Solving Mysteries with the Wisdom of Crowds

Tianyi Li

College of Engineering Computer Science Virginia Tech

Problem:

The information we need to make sense of today challenges the scalability of traditional expert processes.

1. Integrated cognitive process vs. large-scale collaboration
2. Holistic view vs. voluminous data
3. Errors occur in all stages and propagate to later analysis
4. Audit and refine existing analysis

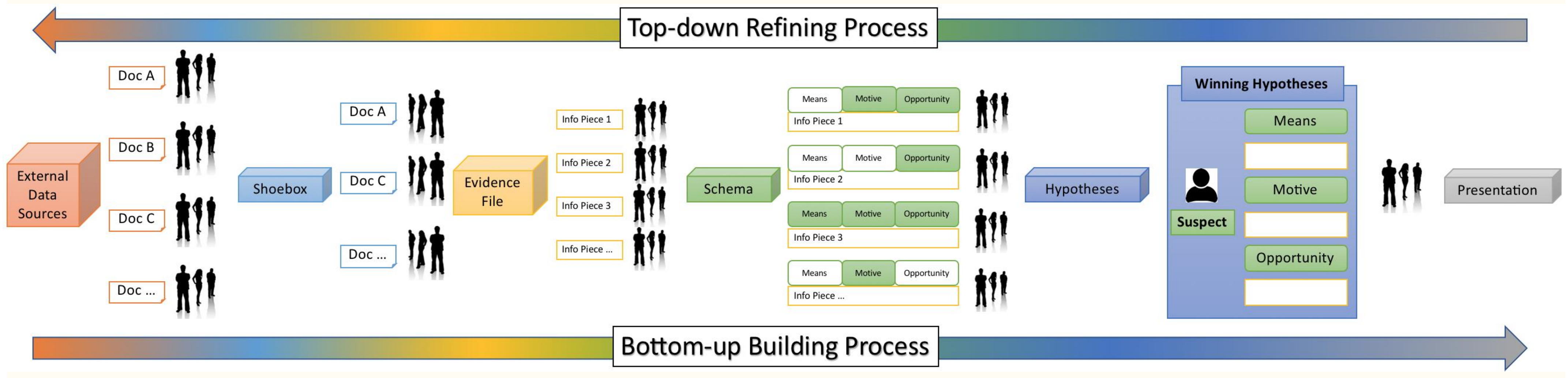
Solution:

Enable many distributed agents to contribute to suitable components asynchronously and meaningfully.

1. Pipeline: modularize the process
2. Context Slice: modularize the data
3. Systematic error analysis
4. Refining path and expert UI

1 2 CrowdIA: a crowdsourced sensemaking pipeline

Li, Luther, North. 2018. Proc. ACM Hum.-Comput. Interact. 2, CSCW



- A pipeline of modularized steps connected by clearly defined inputs and outputs
- Each step partitions information into "context slices" for individual workers
- Implemented the pipeline as a web-based software to enable unsupervised crowdsourced sensemaking

Mysteries	No. Docs	No. Workers	Outcome
Serina ruined Mr. Potter's flowerbed	3	5	✓ **Serina**
Scarlett killed Mr. Boddy in his kitchen with a knife	9	76	✓ **Scarlett**
NYSE is the target of terrorist attack	13	134	✗ **carpet shop**

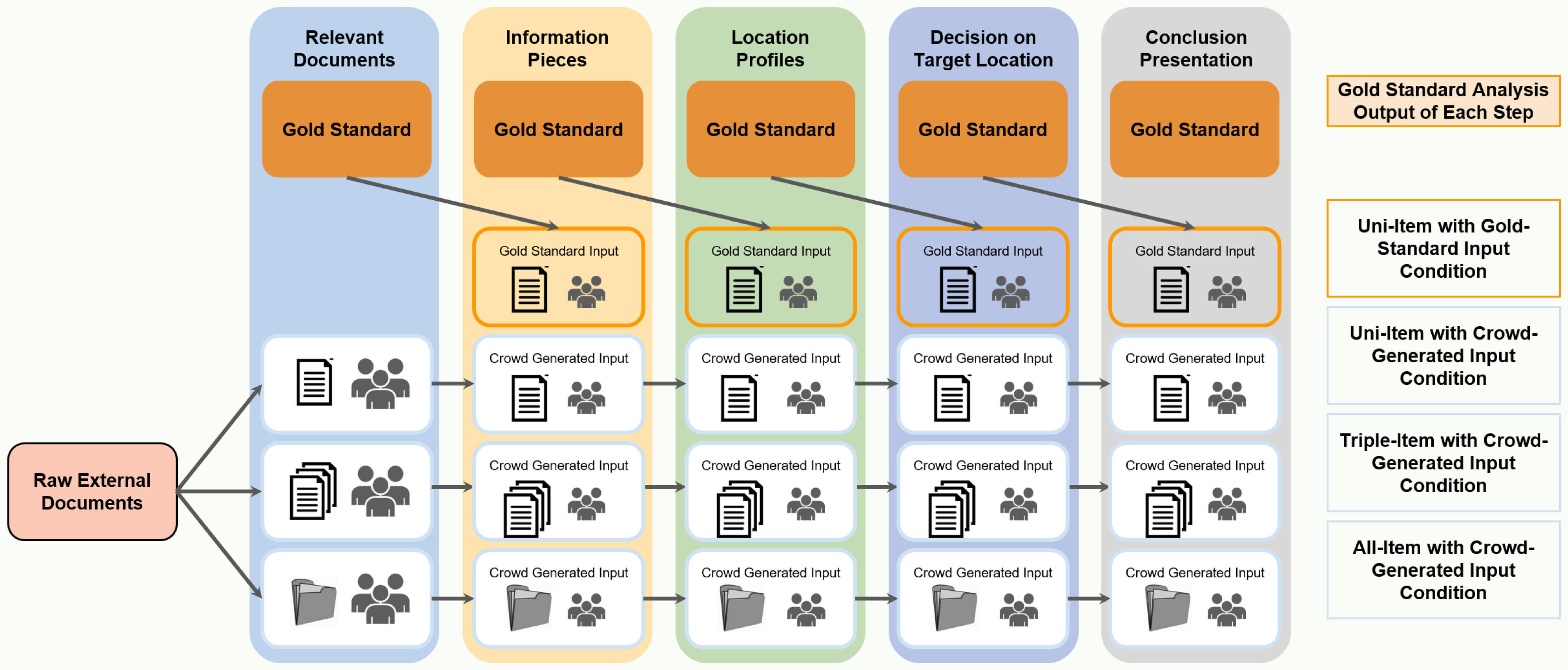
Contributions and limitations:

- Stepwise debugging and optimization
- A much bigger pool of contributors
- Meaningful and scalable division of work
- Crowd mistakes propagate and affect analysis outcome

3 Errors and bottlenecks of crowd sensemaking

Li, Manns, North, and Luther. 2019. Proc. ACM Hum.-Comput. Interact. CSCW

- ➢ What are the errors (type and frequency) workers make in a crowdsourced sensemaking pipeline?
- ➢ How does the amount of local context affect the errors within and across steps?



Types of local errors:

- Insufficient context
- Misinterpretation
- Inattention to background knowledge
- Failing the task goals
- Low effort

4 Completing the loop: enabling expert-crowd collaboration *(work in progress)*

Challenge:

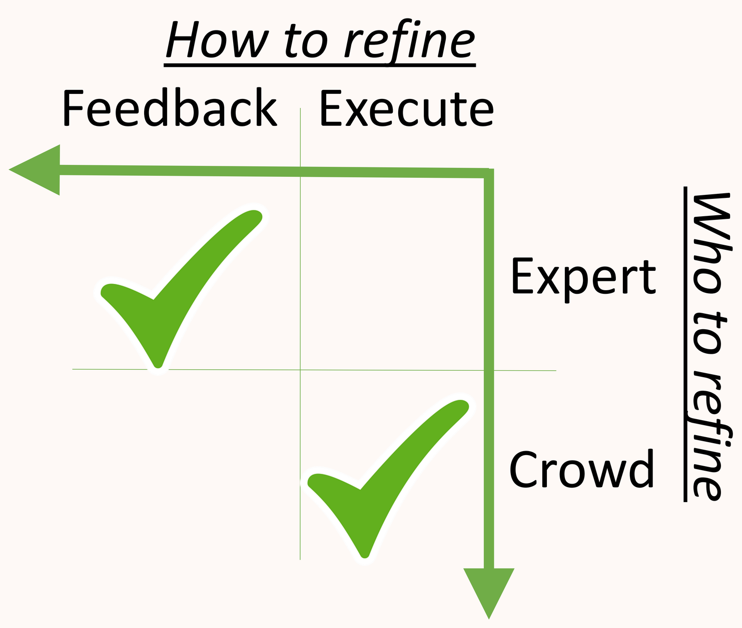
Without a gold-standard solution, we can only judge the logical reasoning but not the correctness of the crowd's analyses.

Scope:

Surface and structure the relevant evidence that helps experts solve the mystery.

Specifically, the refining path needs to support:

1) filling the information holes (looking for relevant information missed in the previous analysis and fitting in the known knowledge
2) removing the irrelevant information from the previous analysis to eliminate distractions.



Planned work

Study 1: Helping experts locate errors in crowd analysis provenance.

- Develop an expert interface to represent the crowd analysis provenance
- Assess expert feedback on crowd errors
  - How many crowd errors can experts find?
  - How well can experts locate the origins of errors?

Study 2: Evaluating crowd refinement with expert guidance.

- "When to stop"? – one batch
- Structure expert feedback as {context, critique, todo}.
- Evaluation: compare to the gold-standard solutions and expert assessment

Acknowledgements

I would like to express my deep gratitude to my advisors **Dr. Chris North** and **Dr. Kurt Luther** for their guidance and encouragement in the past four years. I also want to thank the 2019 DC chairs **Dr. Shaowen Bardzell** and **Dr. Bryan Semaan**, as well as the 2019 DC mentors, **Dr. Steve Jackson, Dr. Neha Kumar, Dr. Michael Muller, Dr. Andrea Parker, Dr. Bryan Semaan, Dr. Volker Wulf** for the attention and feedback. My research is supported in part by **NSF** grants IIS-1527453, IIS-1651969, and IIS-1447416.

Contact:
Tianyi Li
tianyili@vt.edu
http://people.cs.vt.edu/tianyili/