Supporting High-Stakes Investigations with Expert-Led Crowdsourcing

Sukrit Venkatagiri
sukrit@vt.edu
Virginia Tech
Arlington, VA, USA

ABSTRACT
My dissertation introduces the concept of expert-led crowdsourcing (ELC) to support expert investigators who increasingly face limits on their time and attention. ELC combines experts’ domain knowledge and experience with the speed and scale of crowds. I study ELC in two investigative domains: journalism and law enforcement. Through four studies, I show: 1) how novice crowds can effectively augment expert investigators’ work practice; 2) the ethical tensions in conducting an ELC investigation for real-world, sensitive investigations; 3) how capture-the-flag competitions increase inter-team collaboration in ELC investigations; and 4) how different teamwork structures affect intra-team collaboration in ELC investigations.

CCS CONCEPTS
• Human-centered computing → Collaborative and social computing systems and tools.

KEYWORDS
crowdsourcing, investigations, journalism, misinformation, experts

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

GROUP '23, January 8–11, 2023, Hilton Head, SC, USA
© 2023 Copyright held by the owner/author(s).
https://doi.org/10.1145/3565967.3571764

1 INTRODUCTION
Expert investigators in fields such as journalism and law enforcement are tasked with conducting investigations of increasing scope but with dwindling resources. Experts must trace and verify the provenance of every piece of evidence and corroborate facts using digital tools. These labor-intensive tasks can take hours to days, and may not always prove fruitful. They also do not scale easily, meaning that successful investigations are limited by individual experts’ time and attention.

An alternative approach for conducting investigations leverages the powerful and adaptive capabilities of online crowds. This approach has seen success in locating missing persons and supporting crisis response efforts. Crowdsourced investigations, however, are perhaps better known for their high-profile failures, such as when Reddit users misidentified a perpetrator in the 2013 Boston Marathon bombing. Despite this criticism, crowdsourced investigations continue to flourish and evolve.

In my dissertation, I explore how to support experts’ investigative practice with expert-led crowdsourcing (ELC), an approach that combines experts’ domain knowledge and experience with the speed and scale of crowds. I focus on the high-stakes domains of journalism, law enforcement, and misinformation studies, where professionals avoid inaccurate statements, unethical or vigilante conduct, and slow completion or response rates. Designing a successful ELC investigation in high-stakes domains not only requires effectiveness, and efficiency, but also ethical conduct. I explore these dimensions through three research questions across four projects:

• RQ1: How can novice crowds effectively augment expert investigators?
• RQ2: What are the ethical tensions in expert-led crowdsourced investigations?
• RQ3: How do competition, collaboration, and teamwork structures affect efficiency in an ELC investigation?

2 COMPLETED RESEARCH

2.1 Augmenting Expert Work Practice with Crowdsourcing

The first part of my dissertation addresses RQ1. I focused on a key step in many experts’ investigative process, called image geolocation, a sensemaking process that involves verifying the exact location where an image was taken. This manual task may take hours or days and does not scale easily. Computer vision attempts at automating this process are also insufficiently accurate.

I developed GroundTruth [5], a web-based system that helps expert investigators geolocate images with a novice crowd. I proposed an approach, called crowd-augmented expert work, that extends Heer’s idea of shared representations between humans and intelligent agents [1]. This approach allows visual context to be communicated between experts and crowds, augmenting but not replacing the expert.

I conducted a mixed-methods evaluation of GroundTruth involving a think-aloud protocol, log analysis, and semi-structured interviews with 11 experts working with 567 crowd workers from Amazon Mechanical Turk. I found that GroundTruth merges the benefits of both expertise and crowdsourcing. Experts worked with crowds in real-time to narrow the search area substantially, and frequently succeeded in geolocating the image. Experts were also excited by the idea of incorporating GroundTruth into their toolset since it provided them with added functionality.
vestigations involving crowds also pose ethical challenges, such as the potential for vigilantism and doxxing. In the second part of my dissertation, I answer RQ2 through an ethnographic study of CrowdSolve, a real-world, crowdsourced investigation led by law enforcement [4].

CrowdSolve blended top-down guidance by law enforcement experts with bottom-up participation by 250 amateur sleuths. Over four days, experts and the novice crowd collaborated in a co-located setting to investigate two decades-old unsolved murder cases. I described how CrowdSolve represents a third unique model of crowdsourced investigations — called expert-led crowdsourced investigations (ELCI) — by blending the two traditional top-down and bottom-up models.

Using Lee et al.’s lens of human infrastructure [2], I examined how CrowdSolve successfully navigated the ethical tensions between: (1) openness and security and privacy; (2) entertainment and reality; and (3) experts vs. crowds. I also found that CrowdSolve leveraged the complementary strengths of experts and novice crowds. Experts helped the crowd learn valuable skills. In turn, the crowd applied these skills to generate new and useful leads for experts.

3 RESEARCH IN PROGRESS

My prior work showed that shared representations can effectively enable crowds to augment expert work practice. I also showed that ELC can be used to conduct high-stakes, real-world investigations while balancing the ethical tensions. However, at CrowdSolve I found that there was duplication of effort across teams and that there was minimal shared context between teams, which may have reduced the efficiency of the event.

In the final two parts of my dissertation, I answer RQ3 by exploring how to improve inter- and intra-team collaboration to increase the overall efficiency of an ELCI. I focus on debunking misinformation using open-source intelligence (OSINT) techniques. OSINT relies on publicly available information online, thus enabling crowds to conduct real-world investigations without requiring access to confidential law enforcement files. Focusing on debunking misinformation (proving a claim as false) also enables success to be measured.

3.1 Improving Inter-Team Collaboration through Capture-the-Flag Competitions

Using an iterative design-based research approach with a class of 40 students over several sessions, I developed QuriOSINTy a web-based system that enables crowd workers to submit and view other teams’ work output. This reduces redundancies within an investigation and allows for greater inter-team collaboration. The system leverages a “capture the flag”-style (CTF) points-based setup to motivate teams to compete against each other to score the most points. Teams not only get points for the quality of their work, but also for building upon other teams’ work, thus increasing collaboration.

I conducted semi-structured interviews with six teams, consisting of 16 participants in total. I am now analyzing transcripts from the interviews as well as inspecting system log data. My preliminary findings show that teams performed better when they: (1) consisted of members who enjoy competition; (2) tried multiple different tactics instead of just one; and (3) took up tasks that suited their strengths. I also found that many teams were motivated more by altruism (i.e., combating misinformation in the real-world) than competing to win the CTF.

3.2 Increasing Intra-Team Efficiency through Liberating Structures

Prior work has shown that the structure of a team can influence its performance. However, traditional team structures (top-down and bottom-up) may not be optimized for ELCIs. ELCIs are different from traditional, homogeneous collaborations because experts simultaneously lead and participate in the collaboration. Further, experts not only possess greater domain expertise than crowds, but also perform the same (and a superset) of the crowd’s tasks. In this project, I will evaluate how different team structures affect the efficiency of an ELCI. More specifically, I will study four different liberating structures [3] — microstructures that enhance relational coordination and trust — over a semester-long class with
30 students working with expert investigators. To understand participants’ experiences and performance, I will conduct a mixed-methods evaluation consisting of surveys, document analysis, and semi-structured interviews.

4 ANTICIPATED CONTRIBUTIONS
My dissertation informs the design of more effective, ethical, and efficient expert-led crowdsourced investigations. More specifically, I seek to contribute three mixed-methods evaluations of two web-based systems and four team structures, as well as an ethnographic study of a 250-person crowdsourcing event. Throughout, I highlight the challenges and opportunities faced for in conducting expert-led crowdsourced investigations. My findings also inform the design of future systems that support high-stakes investigations.

I also make two conceptual contributions. The first contribution is the idea of shared representations for crowd-augmented expert work that enables novice crowds to effectively contribute to experts’ investigative practice. The second is the concept of expert-led crowdsourcing (ELC). ELC supports experts’ investigative practice by blending top-down expert guidance with the bottom-up participation of a crowd.

5 GOALS FOR DOCTORAL CONSORTIUM
One of my goals for attending the GROUP Doctoral Consortium is to receive feedback on the framing of my dissertation and suggestions for my final study. I also seek mentorship from established HCI and CSCW scholars about their experiences in academia and industry and how they built out a broader research narrative post-graduation. Finally, I hope to provide feedback and build relationships with other Ph.D. students who are at a similar stage.

REFERENCES