Two is Better Than One: A Mixed Methods Approach to Human-Centered Data Science

Jim Maddock
Northwestern University
Technology and Social Behavior
maddock@u.northwestern.edu

Darren Gergle
Northwestern University
Technology and Social Behavior
dgergle@northwestern.edu

Kate Starbird
University Of Washginton
HCDE, DUB
kstarbi@uw.edu

Abstract
Neither deep qualitative nor broad quantitative methods alone fully describe complex human behavior across massive datasets. Mixed methods approaches offer a solution by strengthening findings through mutually supportive analyses. Here we present two examples that illustrate our iterative, exploratory technique for examining human-centered phenomena and social mechanisms within large, complex datasets.

Author Keywords
Social computing; mixed methods; data science

ACM Classification Keywords
H.5.3 [Information Interfaces & Presentation]: Group and Organization Interfaces—Computer-supported cooperative work, Evaluation/methodology

Introduction
Social computing research currently straddles a methodological divide. Both qualitative and quantitative approaches have led to new insights into social phenomena, yet in practice each lacks the strengths of the other. While “big data” techniques offer unprecedented breadth, they often cannot provide the descriptive depth required to understand individual actors. On the flip-side, many qualitative approaches reveal detailed insights, but they lack the analytical scope to

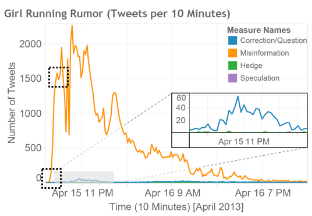

Figure 1: Volumes of misinformation and correction over time. Dashed boxes indicate points of interest.

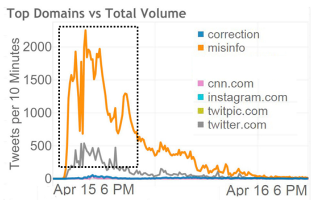

Figure 2: Volumes of specific URLs over time compared to misinformation volume.

describe and model overarching social mechanisms. Due to the complexity of human-centered data science, no single methodological approach shows a complete picture of human behavior.

In the case of social computing however, data-science is not a zero sum game. While quantitative and qualitative methods will always have trade-offs, mixed methods analysis not only mitigates inherent disadvantages, but also strengthens descriptive power. Used in tandem, quantitative and qualitative techniques complement one another, surfacing insights not evident through a single methodological approach.

Social computing and human computer interaction studies that combine these techniques certainly exist, but researchers have yet to standardize and formalize mixed methods approaches. As a result, separation between qualitative and quantitative research is evident across a diverse subset of topics, such as social network, collective action, peer-production, and crowd sourcing research.

Here we argue the advantages of a mixed methods approach to exploratory, human-centered data science through two studies, each addressing a separate sub-domain of social computing research. The first example summarizes a well established, ongoing project in which the authors employ an iterative, exploratory process in order to analyze social media data [11]. The second describes a developing project that will use a similar methodological approach in order to model collaboration dynamics within peer-production communities. Through these two examples we hope to formalize a process that leverages mixed methods analysis in order to explore complex social phenomena across multiple topics.

Project 1: Rumoring Behavior During Crisis

The first project explores rumoring on social media during crises and disasters, with the dual goals of better describing rumoring behavior and building towards automated detection algorithms. Over 2.5 years each iteration of the larger project spawned several sub-projects, though due to space constraints we focus on a smaller subset.

Prior crisis informatics research that analyzes social media or blog data tends to downplay or skim over methodological complexity. Though some studies successfully straddle the divide between qualitative and quantitative [6, 1], research that includes both qualitative coding and quantitative statistics frequently compresses discussions of methodological contributions in favor of other findings. We therefore hope to surface these methodological contributions from our projects in order to begin to formalize mixed methods techniques.

Temporal, Domain, and Lexical Signatures

The initial phase of our project focused on a 10.5 million tweet dataset collected during the 2013 Boston Marathon Bombings. Through quantitative graph-based exploration of co-occurring hashtags we identified five salient rumors, constructed a qualitative coding scheme, and categorized each tweet as “misinformation”, “correction”, or one of three additional codes that expressed varying degrees of uncertainty.

Quantitative temporal analysis and comparison of each code was a natural—if uninformed—first step towards visualizing our dataset (Figure 1). These temporal graphs revealed patterns of peaks and valleys, which we hypothesized could indicate consistent temporal “signatures” across similar types of rumors [8].

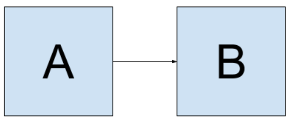

Figure 3: Editor A makes an edit at T1. Editor B makes a later edit at T2.

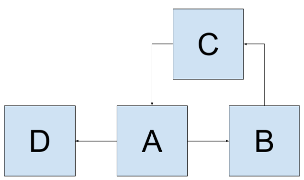

Figure 4: Following Figure 3, Editor C makes an edit at T3. Editor A makes a second edit at T4, and Editor D makes an edit at T5. Revisiting an article creates a loop in the graph structure (Editor A edits at both T1 and T4).

Qualitative investigation of tweets sampled from key moments in each rumor's lifespan—e.g. a peak or a valley—showed that embedded URLs played an important role in rumor propagation, often appearing at the beginning of sudden increases in misinformation volume. In the next iteration of analysis we therefore graphed specific URLs over time, and found patterns that looked similar to our original graphs (Figure 2).

As we qualitatively coded each tweet, we observed that misinformation tweets were characterized by low lexical diversity while corrections leveraged a more diverse vocabulary. We tested this observation by applying a quantitative corpus wide approach to calculate lexical diversity and found consistently lower measures in misinformation tweets [5].

Expressed Uncertainty

Throughout subsequent phases of this project we retooled our coding scheme to improve coder agreement across our additional three codes—"hedge", "question", and "speculation". The transformation resulted in a general "uncertainty" code, which could apply to both "affirmations" and "denials" of the central rumor narrative.

As we coded new rumors from new events, we applied analytical techniques developed earlier in the project and graphed our codes over time. These visualizations showed less consistency due to the diversity of event types, but they indicated the importance of uncertainty as an early predictor and critical component of rumoring [9].

Notably, our qualitative observations showed consistent lexical patterns across tweets coded "uncertainty", regardless of the event or type of rumor. Further investigation and engagement with existing literature on "hedging" revealed distinct but generalizable structures to tweets that contained expressed uncertainty [10].

Mutually Supportive Techniques

While none of these techniques were individually unique, we hope to underscore their descriptive power when appropriately combined. Each analytical iteration suggested a new measure to observe. In our final publications we not only noted quantitative similarities across our temporal, lexical, and geographic measures, but also surfaced the underlying story through qualitative descriptions. Ultimately we derived multidimensional signatures from several relatively simple analyses in order to describe complex human phenomena.

These rapid cycles of exploration, analysis, and discussion were possible due to the data availability and processing power that characterize the "big data" paradigm, yet they do not forfeit the descriptive depth of qualitative techniques. Instead, quantitative and qualitative methods support one another to encompass both analytical breadth and depth, which we argue better describe the complexity of underlying social mechanisms.

Project 2: Hyperlingual Collaboration Dynamics

We hope to use similar methodologies in a subsequent project that will compare collaboration dynamics across different but parallel language communities. Our initial study will analyze Wikipedia talk and article pages from the top 25 largest language editions. Again, while prior work extensively documents collaboration within Wikipedia editor communities, it tends to privilege either quantitative statistical approaches [7, 3, 4], or highly targeted qualitative description [2].

Chain Revision Graphs

Our project attempts to use targeted qualitative description to support network analysis. We will employ chain revision

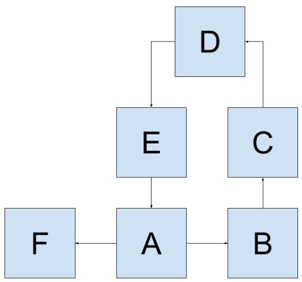

Figure 5: The more edits between when editor A first edits an article and revisits the same article, the "looser" the loop.

graphs, [3], which consist of editor nodes and directed edit edges that model the entirety or a subsection of a talk page conversation. Each node represents a unique editor, who has contributed one or more edits to the page. Directed edges then show the edit history, connecting editors in chronological order (Figure 3). These graphs model editor persistence within the page, as repeat edits from a single editor create a "loop" structure and therefore form tighter graphs (Figure 4). By comparing distributions of the degree of tightness (Figure 5) across different language communities, we can model and compare editor use of collaboration resources over time. Healthy collaborations, for instance, might have many topics of similar length with many contributors, while a "flamewar" might feature a single high activity exchange between two contributors.

Qualitative Support for Quantitative Hypotheses

Despite the power of graph-based approaches, they lack depth without qualitative support, and numerical results from network analysis can easily be misinterpreted without looking at individual data points. In our work we could speculate about the significance of loose or tight loops within chain revision graphs, yet without qualitative samples these hypotheses amount to little more than conjecture.

Qualitative analysis can also validate sampling and component separation within large graph structures. Isolating components within complex networks is difficult because grouping and separation methods easily reshape results. For example, one common technique removes nodes that have less than a specified number of edges in order to isolate dense clusters. Increasing or decreasing this number, however, can dramatically change cluster composition.

As illustrated by the previous study, our qualitative findings can inform sampling methods and coding schemes, which both validate quantitative findings and inform future rounds of analysis. After we build chain revision graphs, we will sample conversations based on graph shape—e.g. loose loops vs tight loops, number of loops, number of authors, or number of interactions. We can then code these conversations based on categories that reflect conflict or coordination interactions. Qualitative coding will hopefully suggest new measures, such as lexical patterns or topic modeling, though these iterations are difficult to predict and will develop as the project progresses.

Conclusion: From Experimentation to Exploration

Iterative, exploratory analysis is possible due to modern data availability. Before the existence of large-scale, public trace-data, researchers generally constructed datasets from observation, surveys and interviews, or experiments. These methods—both qualitative and quantitative—result in highly specific datasets, which force researchers to preemptively estimate important measures and statistics in order to argue a particular, distinct point.

As the previous two studies illustrate however, persistent availability of rich trace data allows researchers to quickly develop and explore new methodologies based on previous results without re-running experiments or taking multiple series of field observations. This rapid, iterative process is human-centered itself, leveraging human capacity to make meaning out of data by repeatedly choosing from a range of analyses. We argue that our human-centered mixed methods approach to data-science can help unwind complex human interaction and behavior, and we suggest that formalization or standardization of mixed method techniques would benefit the larger HCI community.

REFERENCES

1. Prashant Bordia and Nicholas DiFonzo. 2004. Problem solving in social interactions on the Internet: Rumor as

social cognition. *Social Psychology Quarterly* 67, 1 (2004), 33–49.

2. Susan L. Bryant, Andrea Forte, and Amy Bruckman. 2005. Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. In *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*. ACM, 1–10.
3. Brian Keegan, Darren Gergle, and Noshir Contractor. 2012. Staying in the loop: structure and dynamics of Wikipedia's breaking news collaborations. In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration*. ACM.
4. Aniket Kittur and Robert E. Kraut. 2008. Harnessing the wisdom of crowds in wikipedia: quality through coordination. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*. ACM, 37–46.
5. Jim Maddock, Kate Starbird, Haneen J. Al-Hassani, Daniel E. Sandoval, Mania Orand, and Robert M. Mason. 2015. Characterizing online rumoring behavior using multi-dimensional signatures. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 228–241.
6. Onook Oh, Manish Agrawal, and H. Raghav Rao. 2013. Community intelligence and social media services: A rumor theoretic analysis of tweets during social crises. *MIS Quarterly* 37, 2 (2013), 407–426.
7. Aaron Shaw and Benjamin M. Hill. 2014. Laboratories of oligarchy? How the iron law extends to peer production. *Journal of Communication* 64, 2 (2014), 215–238.
8. Kate Starbird, Jim Maddock, Mania Orand, Peg Achterman, and Robert M. Mason. 2014. Rumors, false flags, and digital vigilantes: Misinformation on twitter after the 2013 boston marathon bombing. *iConference 2014 Proceedings* (2014).
9. Kate Starbird, Emma Spiro, Ahmer Arif, F. Chou, Sindhuja Narisimhan, Jim Maddock, Kelly Shanahan, and John Robinson. 2015. Expressed Uncertainty and Denials as Signals of Online Rumoring. *Proc. of Collective Intelligence* (2015).
10. Kate Starbird, Emma Spiro, Isabelle Edwards, Kaitlyn Zhou, Jim Maddock, and Sindhuja Narashimhan. 2016a. Could this be true? I think so! Expressed Uncertainty in Online Rumoring. To appear in *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM.
11. Kate Starbird, Emma Spiro, Jim Maddock, Ahmer Arif, and John Robinson. 2016b. Examining Online Rumoring during Crisis Events: Articulating a Method for Analyzing "Big" Social Data [Special Issue]. *Social Science Research* (2016). Manuscript submitted for publication.