# Measuring Perceived Causal Relationships Between Narrative Events with a Crowdsourcing Application on Mturk

Dian Hu, David A. Broniatowski

Department of Engineering Management & Systems Engineering, The George Washington University, Washington, DC, 20052, USA
{hudian,broniatowski} @email.gwu.edu

**Abstract.** The computational study of narrative is important to multiple academic disciplines. However, prior research has been limited by the inability to quantify each subject's comprehension of the causal structure. With the aid of big data technology and crowdsourcing tools, we aim to design a new approach to analyze the content of narratives in a data-driven manner, while also making these analyses scientifically replicable. The goal of this research is therefore to develop a method that can be used to measure people's understanding of the causal relationships within a piece of text.

**Keywords:** Crowdsourcing · Narrative · Network · Causal Relationships

## 1 Introduction

The study of narratives is uniquely important to various groups including, but not limited to, public health officials, public relations professionals, market analysts and policy makers. In recent years, the emergence of various social media provides new platforms for the dissemination of narratives. With these new platforms, narratives reach more people simultaneously and the effects of narratives last longer [13, 22]. However, the new platforms have also introduced new problems. One study shows that the spread of information online has accelerated the spread of misinformation with negative implications for public understanding of science and public health [7]. Furthermore, in the field of understanding vaccine hesitation, several studies have found that misconceptions are often transmitted in the form of narratives on social media [3,8]. These misconceptions have hindered the public's acceptance of factually-accurate public health messages. Thus, to increase the public's familiarity with, and acceptance of, factually-accurate message among the public, we must first understand how the public comprehends these messages, what forms of messages are more acceptable for them and how the public perceives the internal causal relationships within an article.

Many studies have done extensive work to explain how information is spread [2,6,16]. However, most prior work has focused on describing the underlying mechanisms, and the external variables affecting the spread. Less work has analyzed how the content of a narrative affects its spread. Merely analyzing the external variables and the spread pattern is not enough to understand why certain misinformation can be more popular than others. More importantly, understanding the dissemination mechanism alone cannot help to improve the quality of messages: the factors that lead people to understand the gist of the message. Effective communications can be better achieved if we can understand both factors: how publics comprehend messages and how publics transmit messages. In this project, we focus on one key internal feature: A narrative's causal structure, since prior work has shown that causal structure is a key component of narrative comprehension [19,20]. Each reader may comprehend the same piece of text differently. Therefore, the causal structure generated by individual readers is a direct lens through which we can see their perception of the internal logic within a text. Furthermore, the extracted causal structure could provide a new variable that can be used in other studies of information diffusion. The goal of this current project is to develop a new method that can conveniently collect readers' perceptions of a narrative's internal causal structure. We have also proposed a systematic approach to measuring causal structures from a group using graph theory.

Although essential for future work, the tool presented in this paper is still preliminary. In the future, using data collected from this tool, we would first explore what factors lead people to disagree on the existence of causal links. These factors would include both variables from the individual subjects (e.g., individual's education level, familiarity with background information and common sense) and variables from the text messages (e.g., grammar, writing style, conjunction and reasoning effectiveness). Such analyses cannot be achieved without this work: a convenient web-based application that can collect causal graphs from multiple subjects.

## 2 Background: Representing Text as Causal Networks

The study of narrative reading comprehension has been an active field since the 1970s. Since that time, the field had identified three generations of research and investigations [21]. In the first generation, researchers focused on what readers remember after they read the text [12,20,21]. The second generation of research focused on the process by which readers generate products during readings [9,15,17,18]. Nowadays, the third generation of research is still actively constructing theories to explain the findings from the first and the second generations [21]. Among these studies, causal network theory emerged as a replicable way to describe readers' mental presentations of text [19,20]. When describing a narrative using causal network theory, each event in the narrative is represented by a node, and each causal relationship between events is represented as a link [20]. The rules to establish the causal networks is derived from psycholinguistic literature and rigorous grammars. Traditionally, the graph is constructed by trained judges in this discipline [18,20]. A simplified rule to determine whether a causal link exists from Event A to Event B is to determine "if A did not occur, then B would not have occurred" [20]. Table 1 shows an example of labeled events in a simple narrative while Fig. 1 shows the associated causal network.

**Table 1. Labeled events in a simple narrative**

| Event Index | Text |
|---|---|
| 1 | Daniel arrived in his aunt's house. |
| 2 | He knocked at the door and |
| 3 | rang the doorbell |
| 4 | Daniel's cousin was waiting at home. |
| 5 | He opened the door for Daniel then |
| 6 | two dogs ran out to greet Daniel. |

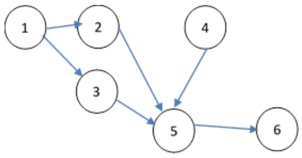


**Fig. 1. Causal network associated with the narrative presented in Table 1.**

Furthermore, prior studies have shown that readers will be more likely to recall events that have a high number of connections when they perform a recall task after reading the narrative [19,20]. However, several concerns still exist surrounding these findings: Can average readers generate repeatable causal networks that are as precise as those generated by trained scholars? Do readers compare each pair of potential causal relationships following the complicated grammar rules? If each reader has a causal graph in mind, would it be possible that each of them is indeed recalling the events that are highly connected in the unique graph in his/her own mind? To address these questions, we propose another approach:

Building upon this foundation of causal network theory, we aim to assess if a large group of readers can identify narrative events and construct these causal networks with a certain level of agreement, using our web-based crowdsourcing application.

## 3 Methods

Due to space limitations, we can only present the outline of the method in this paper. More detailed information on performance and sample outputs of the method section can be found in our prior working paper [11].

Identifying events can be a complicated and subjective process, especially when the task is assigned to a group of people. Thus, we built an event parser based on natural language processing using NLTK [1] to consolidate narrative events. Therefore, in this study, the events are fixed when they are presented to the subjects.

Historically, in off-line settings, creating these causal networks has been both time- and labor-intensive. However, several recent studies have demonstrated that online crowdsourcing is an effective way to perform participatory human research [4]. For example, Amazon's Mechanical Turk (Mturk) has been used in various studies to collect responses from a relatively large group of participants with a wide range of educational and cultural backgrounds [4,5]. Therefore, we aim to gather causal networks from crowdsourcing workers on Mturk. The current version of this web application is based on HTML-5, CSS and JavaScript technology, a combination that is especially suited to developing cross-platform application [10]. The tool can directly save the network information into standard JSON data, therefore granting swift responses and data integrity. To ensure that subjects understanding the task, we formulated specific instructions containing the following information: what the nodes represent, what the links represent and how to use the tool to draw and modify causal networks. A short demographic survey is also included at the end of the task.

## 4 Results from the usability study

On March 6, 2017, we launched our first study on Mturk using this tool. Turkers were eligible to complete the task if they had a HIT approval rate of 95% or higher and were in the United States. 20 subjects were recruited on Mturk without intervention or selection beyond these criteria. The first study used one narrative, borrowed from a prior study in narrative comprehension [19], with 22 events. The Mturk workers could refer to the instructions while finishing the task. In our first case study using this tool, we wanted to assess three objectives: (1) Can Mturk Workers understand our instructions? (2) How long does it take for Mturk Workers to accomplish the task? (3) To what level do the Mturk user agree with each other on each potential causal link? Are there links that exist (or don't exist) with statistical significance?

Among the 20 workers, the average time to finish the task (both drawing and survey) was 23 minutes, with a minimum of 9 minutes, and a maximum of 50 minutes, indicating that our tool can effectively reduce the drawing time for creating each graph. Even with a demographic survey, this is a 50% percent reduction compared to our pilot study on fellow researchers without the drawing tool. To analyze our causal network data, we first aggregated each subject's responses into a 22*22 matrix with one entry for each pair of events in this narrative. That is, we have a matrix where each cell represents the frequency of Mturk Workers who agree that there is a link between these two events. Table 2 shows the partial results of the first five nodes. For example, the number 11 means 11 out of 20 workers had agreed that there should be a causal relationship from Event 4 to Event 5. As we can observe from the partial results, there are many cells with value 0 inside the matrix. Theoretically, it means that 20 subjects all agree that there should be no link between these two events.

**Table 2. Number Of Workers who agree that there should be a link between two events**

| | Event 1 (From) | Event 2 | Event 3 | Event 4 | Event 5 |
|---|---|---|---|---|---|
| Event 1 (To) | 0 | 0 | 1 | 0 | 0 |
| Event 2 | 14 | 0 | 0 | 0 | 0 |
| Event 3 | 8 | 3 | 0 | 0 | 0 |
| Event 4 | 1 | 1 | 14 | 0 | 1 |
| Event 5 | 0 | 0 | 8 | 11 | 0 |

Fig.2 is a histogram of the frequency observed in Table.2. The data indicate that, when dealing with most pairs, participants are more likely to think there is no link between two events.

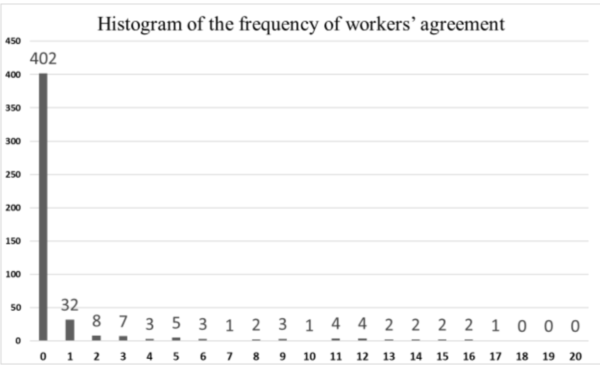

**Fig. 2. Histogram of the frequency of workers' agreement**

## 5 Discussion

To make more sense of the data, we performed a power analysis to assess whether the graph as a whole is statistically significant ($p<0.05$). Between two ordered events, there can only be two scenarios: 1) link exists, or 2) link does not exist. Therefore, we can treat a causal network graph as a series of binary decisions. Thus, the binomial distribution would be appropriate to assess the graphs' significance. To reach P (graph)<0.05 as a graph, given that we have 22 narrative events, and excluding self-links:

$$\mathrm{P(link)} < \frac{\mathrm{P\ (graph)}}{\textbf{number of links}},\ where\ \frac{\mathrm{P\ (graph)}}{\textbf{number of links}} = \frac{0.05}{\mathbf{22^2-22}} = 0.000108$$

Therefore, in this study, to ensure statistical significance, each link must ensure that

$$\binom{N}{n} p^n (1-p)^{N-n} < 0.000108$$

We assume that, absent information regarding causal relationships, people will randomly draw a connection between two events with a probability of 0.5 *a priori*. Assuming a binominal distribution with $p_1=0.5$, and $N=20$ (number subjects), we need $n=19$ or 20 to say that a link exists with statistical significance based on the result. Similarly, we need $n=0$ or 1 to say that a link does not exist with statistical significance. However, in our current data, there is no pair where 19 or 20 out of 20 subjects had agreed that there is a link. Besides of the fact that we only have a small pool of subjects, another theoretical explanation is that the $p_1$ might not be 0.5 in our experimental settings. Since we are launching the task on Mturk, our experiment is limited by participants' effort in critical thinking. When Mturk workers are working on this project, they have the motivation to go through the task as quickly as possible to gain maximum profit per hour. Larger sample sizes will likely yield more significant results.

## 6 Discussion and Future Work

In this paper, we described the development of a new method that can be used to crowdsource the analysis of narratives. Moreover, we have examined the performance of the tool based on a usability study. We proposed a systematic and replicable approach to directly measure people's understanding of the causal relationships within online narratives. In the future, we plan to modify the tool as follows: we would present each pair of narrative events to the workers, asking them whether there is a causal relationship. After the workers have finished assessing all pairs, we would present the drawing panel to the workers again, giving them another chance to review their own decisions. Furthermore, network information saved by this tool can be integrated with other popular network analysis software. SNAP [14], for example, can import JSON based data and extract network properties for more in-depth analysis. We, therefore, aim to continue improving the use cases of this tool to make it more convenient and flexible for scholars and other professionals.

## 7 Acknowledgment and Disclaimer

This work was supported in part by the National Institute of General Medical Sciences under grant number 5R01GM114771. The content is solely the responsibility of the

authors and does not necessarily represent the official views of the National Institutes of Health.

# 8 References


1. Bird, S. et al.: Natural language processing with Python. O'Reilly Media, Inc. (2009).
2. Boyd, S. et al.: Randomized gossip algorithms. IEEE/ACM Transactions on Networking (TON). 14, SI, 2508–2530 (2006).
3. Broniatowski, D.A. et al.: Effective vaccine communication during the Disneyland measles outbreak. Vaccine. 34, 28, 3225–3228 (2016).
4. Buhrmester, M. et al.: Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data? Perspectives on psychological science. 6, 1, 3–5 (2011).
5. Callison-Burch, C., Dredze, M.: Creating speech and language data with Amazon's Mechanical Turk. In: Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk. pp. 1–12 Association for Computational Linguistics (2010).
6. Cowan, R., Jonard, N.: Network structure and the diffusion of knowledge. Journal of Economic Dynamics and Control. 28, 8, 1557–1575 (2004).
7. Del Vicario, M. et al.: The spreading of misinformation online. Proceedings of the National Academy of Sciences. 113, 3, 554–559 (2016).
8. Dredze, M. et al.: Zika vaccine misconceptions: A social media analysis. Vaccine. 34, 30, 3441 (2016).
9. Gernsbacher, M.A. et al.: Investigating differences in general comprehension skill. Journal of Experimental Psychology: Learning, Memory, and Cognition. 16, 3, 430 (1990).
10. Heitkötter, H. et al.: Evaluating cross-platform development approaches for mobile applications. In: Web information systems and technologies. pp. 120–138 Springer (2012).
11. Hu, D., Broniatowski, D.A.: Designing a Crowdsourcing Tool to Measure Perceived Causal Relationships Between Narrative Events. Retrieved at: http://sbp-brims.org/2016/proceedings/LB_129.pdf
12. Kintsch, W., Van Dijk, T.A.: Toward a model of text comprehension and production. Psychological review. 85, 5, 363 (1978).
13. Kozinets, R.V. et al.: Networked narratives: Understanding word-of-mouth marketing in online communities. Journal of marketing. 74, 2, 71–89 (2010).
14. Leskovec, J., Sosič, R.: SNAP: A general purpose network analysis and graph mining library in C++ http://snap.stanford.edu/snap. June (2014).
15. McKoon, G., Ratcliff, R.: Inference during reading. Psychological review. 99, 3, 440 (1992).
16. Shutters, S.T., Cutts, B.B. : A simulation model of cultural consensus and persistent conflict. In: Proceedings of the second international conference on computational cultural dynamics. pp. 71–78 (2008).
17. Singer, M. et al.: Minimal or global inference during reading. Journal of Memory and Language. 33, 4, 421 (1994).
18. Trabasso, T. et al.: Logical necessity and transitivity of causal relations in stories. Discourse processes. 12, 1, 1–25 (1989).
19. Trabasso, T., others: Causal cohesion and story coherence. (1982).
20. Trabasso, T., Van Den Broek, P.: Causal thinking and the representation of narrative events. Journal of memory and language. 24, 5, 612–630 (1985).
21. Van den Broek, P. et al.: The landscape model of reading: Inferences and the online construction of a memory representation. The construction of mental representations during reading. 71–98 (1999).
22. Winterbottom, A. et al.: Does narrative information bias individual's decision making? A systematic review. Social science & medicine. 67, 12, 2079–2088 (2008).