PennState
College of Education

Discovering Informal Learning Cultures of Blind Individuals Pursuing STEM Disciplines

JooYoung Seo, M.Ed.

The Pennsylvania State University, Learning, Design, and Technology

Background

- **Blindspot in STEM (Science, Technology, Engineering, and Mathematics) Education**
  - Significant difficulties of students with disabilities.
  - Extra challenges of Blind students:
    - Material accessibility issue
    - Instructional inclusivity issue
    - Networking issue
    - Low social and self-expectation issue
- **Existing STEM accessibility studies:**
  - Top-down approach:
    - Usability filed test
    - Special curriculum design
  - Absence of either systematic quantitative or in-depth qualitative analysis

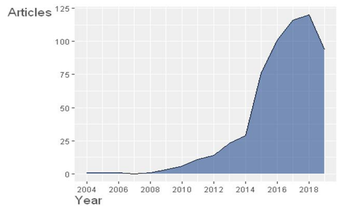

Figure 1. Growing interest in STEM

Purpose

- To discover collective knowledge sharing patterns and informal learning cultures of blind individuals pursuing STEM disciplines as captured through computer-mediated mailing listservs

Research Questions

- **Overarching Qualitative Research Questions**
  1. What are the common STEM *issues* of blind learners that *provoke* discussion?
  2. What are the *patterns* of interaction between blind mentors and mentees in the STEM-related mailing lists?
  3. What *strategies* are proposed or utilized by blind individuals pursuing STEM disciplines in the mailing lists?

- **Quantitative Research Questions**

  1. Descriptive Questions

  1.1. What is the frequency and variation patterns of collective knowledge participations of members in the target mailing listservs?

  1.2. What are the *top-10 most participated topics* among members found in the target online listservs?

  1.3. Who are the *most represented* population of the mailing listserv?

  1.4. Who are the *least represented* population of the mailing listserv?

  2. Exploratory Network Question

  2.1. What does the *directionality* of the relationship between discussion starters and participants among the mailing lists members look like?

  3. Data Clustering Questions

  3.1. What are estimated *latent topics* across all of the four target mailing lists calculated by Structural Topic Models?

  3.2. In what ways are the estimated structural topics *correlated* with each other?

  3.3. How does the *rate of topics* change *over time*?

  3.4. How do these topical distributions vary by the *four types* of the NFB mailing lists?

  3.5. How do these topical distributions vary by the *number of discussants*?

  3.6. How do the detected topical distributions vary by the *number of discussants and the type of mailing lists over time*?

Methodology

✓ Research Design

- **Quantitative Ethnography** (Shaffer, 2017)
- **Mixed-method** interaction between data-driven computation and context-driven interpretation
- "Computer-assisted data science exploration" + "Humanistic contextual deep interpretation"
- Syntactical pattern discovery; semantical interpretation
- A large-scale corpus in the least intrusive fashion

✓ Data Collection

- Target community: the National Federation of the Blind (NFB); the four STEM-oriented listservs <Table 1>.
- Data: NFB members' email message archives between December 2008 and December 2018
- Sample size: **a total of 23,540 messages**
- Data crawling and tidying methods:
  - Crawling: Unix shell wget command.
  - Tidying: R package mboxr (Seo and Choi 2019) to convert an mbox mail format into a tibble data frame <Table 2>

Categories	Mailing Lists	Topics	Number of Subscribers	Number of Messages (Dec 2008-Dec 2018)
Science & Engineering	NFB Science and Engineering Division	To promote education solutions for the blind, share professional successes, and encourage new solutions and techniques to succeed in science and engineering as a blind person.	259	1920
Technology	NFB in Computer Science	The Discussion of the business and operation of the NFB in Computer Science. To share information about the worlds of computer science and technology.	302	12428
Arts	Artists-Making-Art	explore art with all senses.	83	378
Math	BlindMath	Sources for accessible texts, information about tactile and auditory graphing programs, suggestions for insuring that math lectures are accessible to blind students, and strategies used by blind math instructors.	655	8814

Table 1. A summary of the target NFB mailing lists.

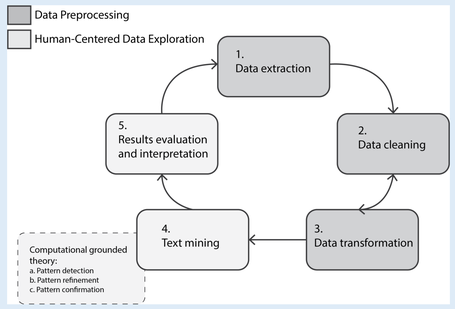

Figure 2. Analytical procedure following the five phases of Knowledge Discovery in Textual Databases (Feldman and Dagan, 1995) combined with Computational Grounded Theory (Nelson, 2017)

Variable	Date	weekday	message_ID	in_reply_to	references	num_discussants	from	to	cc	subject	content
Description	Date and time of message sent (UTC/GMT).	Day of the week for current message sent as an abbreviated three-character string.	A globally unique message identifier containing url-encoded characters.	Message-ID to which current message is replying, if any.	All appended Message-ID(s) involved in current message thread, if any.	The number of discussants for current message thread calculated by using references field.	Sender's Name along with email address.	Recipient's name along with email address.	Copied member(s) along with their email address(es).	Message subject.	Message content.
Example	2018-02-23 12:12:47 UTC	Fri	<002501d3ac3b$095e76b0$1c1b6410$@gmail.com>	<007802c3bz3b$095e76b0$1c1b6410$@pseudo.pseu>	<007802c3bz3b$095e76b0$1c1b6410$@pseudo.pseu> <008901z4dq3p$095e76b0$1c1b6410$@pseudo.pseu>	3	sjysky at gmail.com (JooYoung Seo)	NA	NA	[nfbcs] Pseudo Subject	This is message. Sincerely, JooYoung

Table 2. A structured sample of email texts

✓ Data Analysis

1. Descriptive Statistics

- Techniques: frequency (counting) & variation (central tendency)
- Variables: message_ID; from; subject; num_discussants
- Outcomes: frequency of message sent over time; M, SD, VAR of number of discussants; ranks of the most and least active members

2. Directional Network Analysis

- Techniques: exploratory network analysis
- Variables: message_ID (from-node); in_reply_to (to-node); frequency of the message exchange between the nodes (edge-weight)
- Outcomes: message exchange patterns; directionality; centrality; density

3. Structural Topic Modelling (Roberts, Stewart & Tingley 2019)

- Techniques: unsupervised probabilistic topic modelling; STM algorithms
- Variables: text (content); covariates (mailing_list_type; date; num_discussants)
- Outcomes: estimated topic clusters; topical prevalence; topical content

Evaluation and Interpretation

- Interpretive reflexivity
- Resting upon all the data-driven results
- Synthesized answer to the *qualitative* research questions

Expected Contributions

- Bringing meaningful discussion points on "How Blind People Learn STEM"
- Suggesting a novel methodology to investigate large corpora of texts in rigorous, reliable, and reproducible ways
- Describing how a blind learning scientist researches

References

- Feldman, Ronen, and Ido Dagan. 1995. "Knowledge Discovery in Textual Databases (Kdt)." In *KDD*, 95:112–17.
- Nelson, Laura K. 2017. "Computational Grounded Theory: A Methodological Framework." *Sociological Methods & Research*, 0049124117729703. https://doi.org/10.1177/0049124117729703.
- Roberts, Margaret, Brandon Stewart, and Dustin Tingley. 2019. *Stm: Estimation of the Structural Topic Model*. http://structuraltopicmodel.com.
- Seo, JooYoung, and Soyoung Choi. 2019. *Mboxr: Reading, Extracting, and Converting an Mbox File into a Tibble*. https://CRAN.R-project.org/package=mboxr.
- Shaffer, David Williamson. 2017. *Quantitative Ethnography*. Lulu. com.

Contact Info: jooyoung@psu.edu