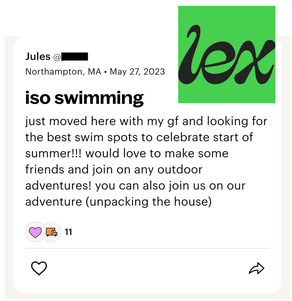
Exploring Online Tie Formation in Local Queer Communities

Jules Tucher

DACSS 758: Text-as-Data and DACSS790N: Network Inference, Spring 2025

Background

With homophobia and transphobia on the rise in American culture and politics, as well as increasing trends of loneliness in a post-pandemic world, there exists a strong need for community relationships among LGBTQ+ people. The geolocation-based mobile app Lex provides users with a platform to create and comment on text-based posts within their local community. Scraping posts, comments, and user profiles from the app within a 50-mile radius of Amherst, Massachusetts provides a robust corpus of what self-identified queer people are discussing in the local area.

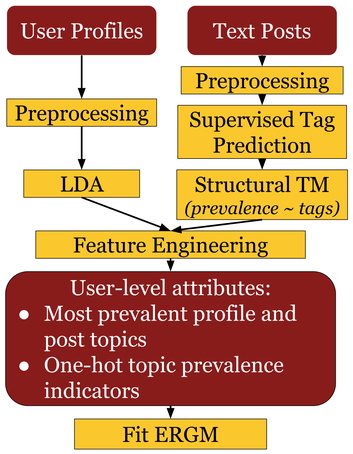


Research Hypotheses

H1: Tag usage will influence what topics local LGBT+ people are discussing on the alternative social media platform Lex.

H2: Users posting about similar topics are more likely to comment on each other's posts (homophily).

H3: Users who receive many comments will be more likely to receive additional comments (popularity). Users are more likely to comment on posts by users who have commented on their posts (reciprocity).



Preprocessing and Tag Classification

- 30% of posts in the corpus are tagged by the posting user
- Five most prevalent tags: *community*, *friends*, *event*, *hookup*, and *dating*
- Preprocessing decisions
  - Remove advertisement posts, stop words, and punctuation
  - Trim corpus-specific rare and common words from doc frequency matrix
  - Typos and alternative spellings were left unchanged
- Train supervised learning models with one-hot tag classification as outcome

Figure 1. Tag Classifier Performance by Outcome and Model Type

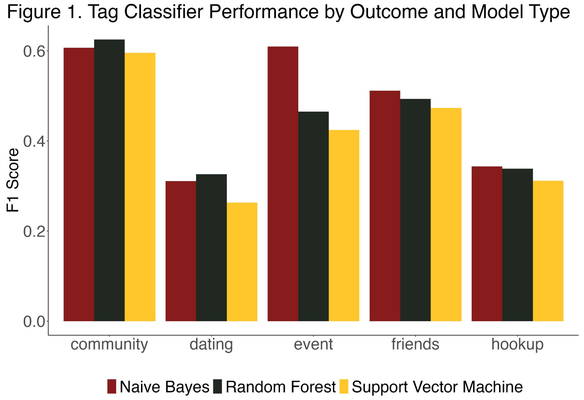


Figure 1 is a comparison of each model's performance (F1 score), used to select the model for predictions. Since the models for *dating* and *hookup* tags had poor performance, the predictions were not used in later steps.

Modeling Topics in User Profiles and Posts

Latent Dirichlet Allocation (LDA) for User Profiles

- LDA analysis applied to users' "About Me" section
- 5-fold cross-validation comparing perplexity scores optimal $k = 20$
- Topic themes: hobbies, relationship style (monogamy vs. polyamory), chronic illness and COVID precautions, gender and sexual identities, and location

Structural Topic Model STM for Text Posts

- STM modeled prevalence on predicted tag values
- Optimal number of topics $k = 20$ based on hold-out likelihood and residuals

Figure 2. Difference in Estimated Topic Proportion for Each Topic, by Tag

Network Description

- One-mode sociocentric directional network
- Node = users; tie = comment(s) by the sender on the receiver's post
- Subset to activity from June 2024
- 363 nodes and 505 edges (density = 0.004)
- Figure 3 plots the subnetwork, with nodes scaled based on the number of posts (activity) and shaded based on the number of incoming edges (popularity).

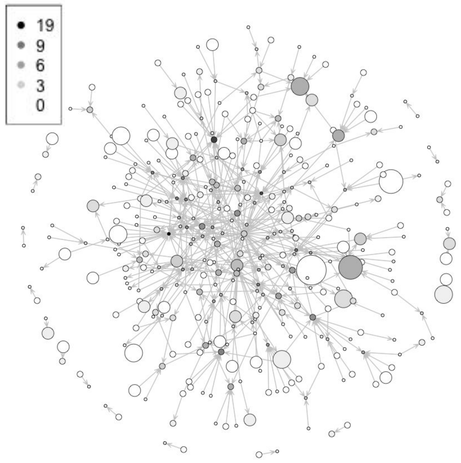


Figure 3. Lex Network Activity for June 2024

Feature engineering node (user) attributes:

- User profiles
  - Topic indicator: estimated topic proportion ≥ 20%
  - Highest topic: topic with highest estimated proportion
- Text posts
  - Topic indicator: any post by user has estimated topic proportion ≥ 20%
  - Highest topic: topic with highest average proportion across all posts

ERGM Results

- Exponential Random Graph Model (ERGM) chosen for its emphasis on tie formation (i.e. likelihood that users would leave a comment)

Table 1: ERGM Results

	Covariates (1)	Endogenous terms (2)
Edges	-6.408***	-4.393***
Number of posts	-0.141**	
Instagram	-0.185	
Age	0.017***	
Highest Profile Topic		
Topic = 3	-0.837*	
Topic = 8	0.808**	
Topic = 12	1.047***	
Topic = 18	1.089***	
Profile Topic (homophily)		
Topic 3	-0.401*	
Topic 6	0.279*	
Topic 8	0.403*	
Topic 13	-0.604***	
Topic 16	-0.438**	
Topic 18	0.461*	
Topic 20	-0.466**	
Highest Post Topic		
Topic = 3	0.679***	
Topic = 8	0.531*	
Topic = 16	0.273*	
Topic = 18	0.530*	
Post Topic (homophily)		
Topic 4	0.283*	
Mutual (reciprocity)		1.224**
GW In-degree, decay = 0.5 (popularity)		-2.795***
GW Out-degree, decay = 0.15 (sociality)		-0.252
GW Edgewise-shared partners, decay = 0.25		1.154***
GW Dyad-shared partners, decay = 0.25		-0.047*
AIC	6,541.193	6,252.694
BIC	7,353.435	6,311.410

*p < .05; **p < .01; ***p < .001

Select user profile topic word clouds:

Topic 8

Topic 12

Topic 18

ERGM with endogenous terms outperforms covariates model (smaller AIC).

Discussion

Lex provides its LGBTQ+/queer users a platform to express themselves and connect about shared community events and interests. Community-tagged posts include requests for housing connections, while event-tagged posts promote financial accessibility with terms like "sliding scale." Many people are seeking recommendations for local spots to get outside or new members to join their board game groups. This project demonstrates the variety of language that local queer people use to self-identify and how, in some cases, that language can lead to connection. Lex users who publicize their attention to COVID precautions or specific hobbies are more likely to form ties with other users who do the same. Counter to H3, network activity seems to improve the likelihood of tie formation more than post popularity, as evidenced by the negative coefficient for GW In-degree. Finally, users are more likely to comment if it's reciprocal or closes a triad. This project provides a peek into a rich local subculture that has its own priorities, lingo, and means of connecting.