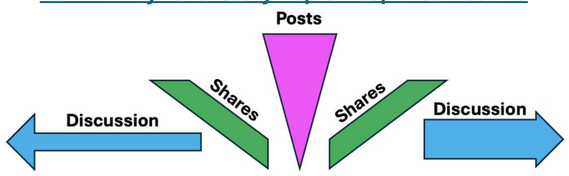
PRIVACY PRESERVING DATA ACCESS PRIMITIVES FOR SOCIAL PLATFORM DATA

Kyle Resnik(1), Christine Task(2), Doyle Groves(1,3), Bennett Hillenbrand(4)
(1) Chattersome Labs, (2) Knexus Research, (3) Indiana University, (4) Georgetown University

Social media platforms are essential environments for studying human behavior, information diffusion, and community formation. **We introduce Privacy Preserving Data Access Primitives (PPDAP), a proposed standardized framework for high-level analysis of social platform dynamics that maintains individual privacy**. PPDAP consists of a prespecified collection of rich summary statistics designed to be computable with minimal sensitivity to individual records. This framework is built from the ground up using an abstract, cross-platform taxonomy of network actions and operates over aggregated count data derived from those actions, satisfying formal privacy for anonymized data release.

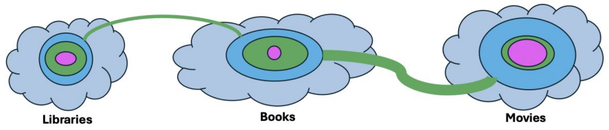
With PPDAP, we are now equipped to generate a general-purpose, privacy-preserving window into social platform activity. For a given interest-defined community (for example, a Subreddit on Reddit, or a hashtag-defined community on Bluesky), and a popular story topic identified by our privacy-preserving Word Adjacency Graph algorithm, we can summarize the impact of that story as the counts of users interacting with it across our taxonomy-defined user archetypes. Different stories will be received differently by different communities, and this comprises the dynamics of the information flow. **Because our base statistic is a count of users active on a topic, changing an individual changes these counts by at most one, and thus differential privacy requires only a small amount of added noise**.

Community-level Story Topic Response Actions



By evaluating at the community rather than individual level, we are able to produce meaningfully granular statistical information about that community's behavior without violating the privacy of the individual users who comprise those communities. PPDAP aims to support researchers in answering our hardest social science questions by improving the quality of data that underpins those answers.

Story Response and Network Behavior Between Communities



Scan here to check out the prototype application "Words of the Hour," built on the PPDAP principles by Chattersome Labs.

To enable these analyses, we introduce a unified framework combining a cross-platform taxonomy of user actions, a privacy-preserving topic identification process, and a novel approach to collecting and analyzing social network data in aggregate. **These components come together to form the PPDAP framework, which shifts analysis away from individual nodes and edges toward high-level community dynamics**. These community dynamics are informed by abstract social actions common to all platforms; users create original content through posts, broadcast content through sharing, then create discussion by replying. PPDAP enables privacy-preserving access to community activity patterns on trending topics, using aggregation across user roles and action types.

Cross-Platform Taxonomy

A cross-platform taxonomy enables the identification of universal user actions – such as posting, sharing, discussion, and reaction – and provides a basis for defining cross-platform behavioral roles. Although platforms employ distinct terminology and features (e.g., "retweets," "shares," or "reposts"), these actions are functionally equivalent and serve analogous social purposes. This fragmentation has historically impeded the development of standardized approaches to social platform data analysis. **The taxonomy serves as a translation layer that maps observed user behavior to universal actions and associated behavioral roles**. Users exhibit consistent tendencies in how they employ platform actions (e.g., prioritizing original content creation over redistribution) captured by roles such as the "Content Broadcaster." This observation enables privacy-preserving measurement of community responses by aggregating across roles of the individuals that interact with it: How much interaction does a topic get from a community's habitual posters, sharers, or discussants?

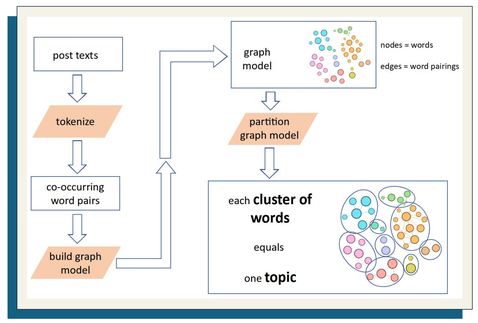
✅ - present on the platform | 🟨 - limited/partially implemented on the platform | ❌ - not present on the platform

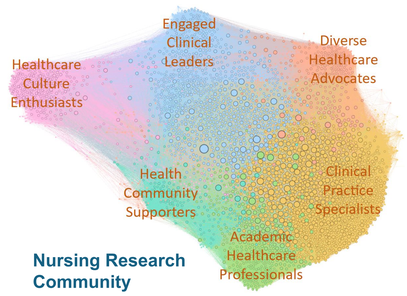
			Twitter (X)	Facebook	Reddit	TikTok	bluesky
CREATION ACTIONS	Original Content Creation	Text Posts	✅	✅	✅	✅	✅
		Content Drafts	✅	✅	✅	✅	✅
		Content Scheduling	✅	✅	❌	🟨	❌
		Interactive Content	✅	✅	✅	✅	❌
		Image Posts	✅	✅	✅	✅	✅
		Video Posts	✅	✅	✅	✅	✅
		Audio Posts	✅	✅	✅	✅	❌
	Profile Creation	Bio / Description Setup	✅	✅	✅	✅	✅
		Profile Photo	✅	✅	✅	✅	✅
		Account Settings / Customization	✅	✅	✅	✅	✅

Role Category	Specific Role	Engagement Level	Output Level	Description & Characteristics
Content Creators	Mega-Influencers	High	High	Millions of followers, professional content production
	Macro-Influencers	High	High	100K-1M followers, semi-professional or professional
	Micro-Influencers	High	High	10K-100K followers, niche expertise or local influence
	Nano-Influencers	High	High	1K-10K followers, highly engaged small communities. "Niche micro-celebrities" with consistent, passionate audiences
	Platform Strategists	Variable	High	Analyzes and predicts platform evolution, strategically manages multiple platforms, optimizes content for platform mechanics
	Emerging Creators	Variable	Variable	Building audience, experimental content phase. Measurable by interaction patterns and growth rate
	Accessibility Advocate	Variable	Variable	Creates content specifically designed for inclusive audiences conforming to ADA/W3C standards
Content Amplifiers	Super Connectors	High	Medium	Bridge multiple communities, high resharing activity. Re-shares content more than creating original
	Community Champions	High	Medium	Promote and defend specific communities or causes
	Trend Propogators	High	Medium	Early adopters who spread viral content
	Cross-Platform Distributors	High	Medium	Share content across multiple networks

Topic Detection

Word Adjacency Graph (WAG) modeling converts texts into word models represented by graphs of adjacent and co-occurring words. Frequent neighboring of words indicates topics, which in turn enables grouping of original texts by topic and noting demographic differentials among users behind each topic. **WAG modeling reframes topic detection as a graph problem rather than a probabilistic inference problem.** Instead of assuming documents are bags of independent words (as in LDA), WAG constructs a word adjacency graph in which nodes are terms (or *n*-grams) and edges represent local co-occurrence within a sliding context window. Crucially, WAG does not rank terms by raw frequency. Instead, it uses population-weighted filtering: a term is retained only if it is used by at least a minimum number of distinct authors. For privacy, random noise is added to weights to obfuscate the impact of adding or removing any individual's posts.





Privacy-Preserving Data Release

Our preliminary analysis shows that in a given community, a given individual often fills a predictable role, for instance preferring to mostly post or mostly react to posts. We can categorize users by their preferred distribution of activities in a given community, and we can then trace how a community responds to a story with privacy preserving activity statistics across different role types (see draft schema below). **If we change one individual, the counts of these community level activities change very little, which then requires little noise addition for differentially private protection.** If we include bridging activity statistics (ex: posts by individuals belonging to multiple communities, posts with hashtags spanning multiple communities), this data also gives us a high level view of network activity between communities. Put another way: we posit that how the Library and Book communities interact - and how information flows between them - is the most valuable signal when compared with how any one specific individual might engage across the Library and Book communities.

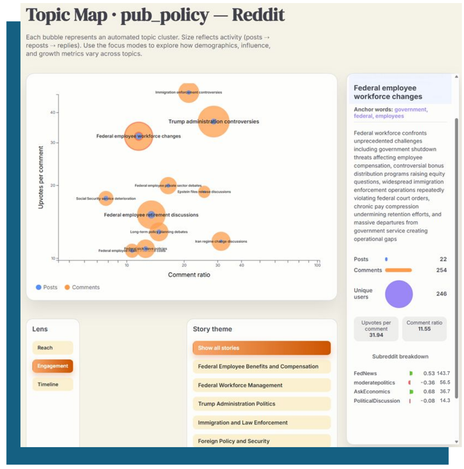
Community	Story Anchor	Story Expansion	Time	Role 1 Activity (Posters)	Role 2 Activity (Sharers)	...	Bridge Activity	Discussion Statistics

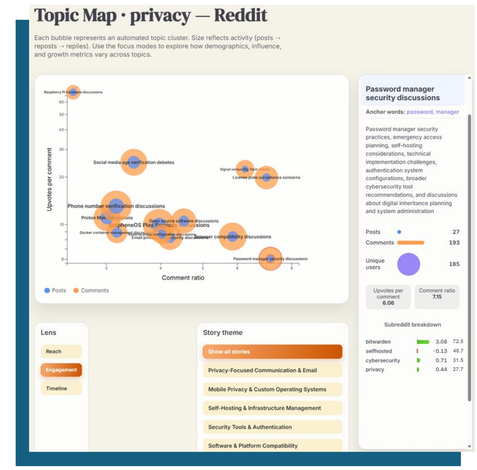
We are currently evaluating this approach on public data, using high-volume content from two platforms (Reddit and Bluesky) and have confidence in further developing a formally private generalized approach to work with any platform.

Once we've released the differentially private data in the raw story activity schema, there are many possibilities for downstream applications to make this data more accessible to non-technical users. **Like a weather reporter interprets barometric data, platform dashboards help us see social data.**

The figures below show two social platform observatory dashboards we've developed (prototypes not yet formally private): **Word Chipper** and **Words of the Hour**.

Word Chipper Reports





Words of the Hour

These dashboard prototypes provide a novel, high-level analysis of communities and topics of discussion on social platforms, built on the foundations of PPDAP. In **Word Chipper Reports**, users will be able to generate snapshots of discussions happening within particular communities on the social platforms. In the above examples, we can see the Reddit Public Policy and Privacy community's current discussion topics.

In the example to the left, the **Words of the Hour** show currently overactive words and phrases drawing from public data on the social platform BlueSky, as well as privacy controls that can be used to explore different noise addition and data suppression settings.

The PPDAP are not designed to address all dataset access needs. Our target user is a researcher with applications for basic platform activity data. **The PPDAP will enable access to that archetype of researcher as a single, generally available release - with no lengthy negotiation, complex contracts, or special skills needed**. Easier access and lower requirements for use, along with a widely-promoted public exercise, will encourage broader participation.

PPDAP doesn't focus on network analysis, but rather aggregate platform activity. We will be working to explicitly bring in new groups who can use this new pre-processed data for public good applications such as but not limited to: **economic analysis, public health, subculture analysis, and technology adoption**.

Welcoming more researchers, from more disciplines and perspectives, provides more opportunities for outside assistance on challenging problems facing social platforms, including real-time detection of important issues. **Because PPDAP are platform-curated and private by design, discoveries don't need to be limited to academic research publications: it can offer a broader opportunity for us to see and understand the online worlds we enjoy spending time in**.

Moving forward with our PPDAP project will require additional funding and collaboration. Collaborations with both social platforms themselves as well as social scientists and researchers will help us launch the next phase of PPDAP. Comprehensive experimentation and tuning of differential privacy, allowing us to understand privacy/utility trade-offs across a range of communities and platforms, will prepare this approach for real world deployment.

Contact us: Christine Task (christine.task@knexusresearch.com), Doyle Groves (deej@deej.com), Bennett Hillenbrand (bennett.hillenbrand@gmail.com), Kyle Resnik (kresnik.dev@gmail.com)