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.

PPDAP supports a broad range of research applications. Social scientists can study subculture formation, community dynamics, and discourse patterns, while policy researchers can analyze how policy proposals are received across constituencies, how political narratives propagate, and which communities engage with specific government issues. 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.

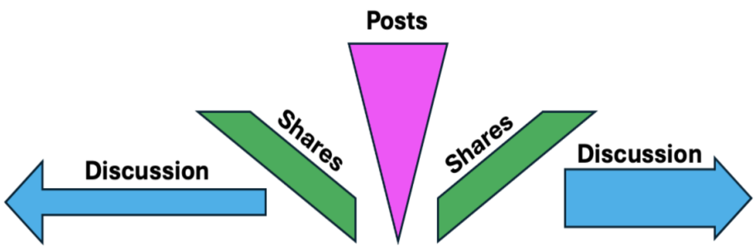

Figure 1: Social Activity on a Topic

Cross-platform Action 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.

Our taxonomy addresses this challenge by organizing user actions into cross-platform categories: *Creation Actions* (original content), *Engagement Actions* (approval and conversational behaviors), *Network-Building Actions* (follows, subscriptions, and direct communication), *Distribution Actions* (reposting, quoting, and sharing), *Governance Actions* (reporting, blocking, and privacy controls), and additional platform-specific clusters such as visual remixing, livestreaming, and community management.

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?

Privacy-preserving Topic Detection

The next step is identifying the topics under discussion on a platform at any given point in time. We use a Word Adjacency Graph topic detection approach, with contributions from individual users constrained to reduce sensitivity to posts from any particular individual.

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. This single threshold parameter replaces the many tunable hyperparameters common in traditional topic models and ensures that retained vocabulary reflects shared discourse rather than the habits of prolific authors, spam, or duplication. Topics are not predefined; they emerge naturally from collective usage patterns.

Once the graph is built, WAG applies ideas from network science, particularly network dismantling. Highly connected nodes – terms that bridge many parts of the graph – often correspond to domain-general or discourse-level vocabulary that obscures topical structure. By iteratively removing these high-centrality nodes and monitoring changes in graph modularity, WAG reduces noise while preserving meaningful structure. The remaining graph is then partitioned using a modern community-detection algorithm (Leiden), yielding clusters of tightly connected terms. Each cluster corresponds to a topic, defined structurally rather than probabilistically. Because the communities are guaranteed to be internally connected, the resulting topics tend to be coherent and interpretable.

Privacy-preserving Data Release

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 topic identified by our privacy-preserving WAG algorithm, we the 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.

Sports and Entertainment		Content Broadcaster	Pure Conversationalist	Engaged Discussant
Blue Jays celebration	Toronto Blue Jays reach World Series after defeating Seattle Mariners in thrilling comeback	-0.23	-0.60	0.26
George Springer playoff heroics	George Springer's dramatic postseason performance creates intense fan reactions and cheating allegations	0.11	-0.94	-0.02

Figure 2: Normalized Activity by Role

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

References

1. Miller, W. R., Groves, D., Knopf, A., Otte, J. O., Silverman, R.D. (2016). *Word Adjacency Graph Modeling: Separating Signal From Noise in Big Data*. *Western Journal of Nursing Research,* vol. 39, no. 1, 2017, pp. 166–185, https://doi.org/10.1177/0193945916670363
2. Task, Christine Marie. *Privacy-Preserving Social Network Analysis*. PhD diss., Purdue University, 2015. *Open Access Dissertations*, https://docs.lib.purdue.edu/open_access_dissertations/569/