



Engineering

[Back](#)

Open source

# Twitter's Recommendation Algorithm

By

[Twitter](#)

Friday, 31 March 2023

Twitter aims to deliver you the best of what's happening in the world right now. This requires a recommendation algorithm to distill the roughly 500 million Tweets posted daily down to a handful of top Tweets that ultimately show up on your device's **For You** timeline. This blog is an introduction to how the algorithm selects Tweets for your timeline.

Our recommendation system is composed of many interconnected services and jobs, which we will detail in this post. While there are many areas of the app where Tweets are recommended—Search, Explore, Ads—this post will focus on the home timeline's For You feed.

## How do we choose Tweets?

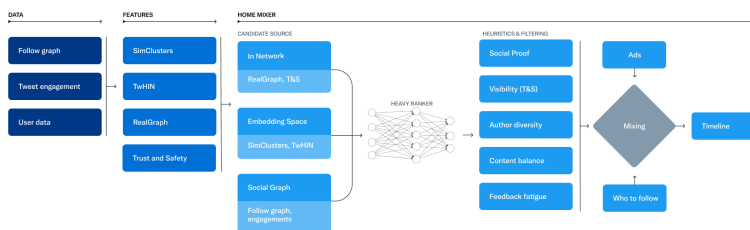
The foundation of Twitter's recommendations is a set of core models and features that extract latent information from Tweet, user, and engagement data. These models aim to answer important questions about the Twitter network, such as, *"What is the probability you will interact with another user in the future?"* or, *"What are the communities on Twitter and what are trending Tweets within them?"* Answering these questions accurately enables Twitter to deliver more relevant recommendations.

The recommendation pipeline is made up of three main stages that consume these features:

1. Fetch the best Tweets from different recommendation sources in a process called **candidate sourcing**.
2. **Rank** each Tweet using a machine learning model.
3. Apply **heuristics and filters**, such as filtering out Tweets from users you've blocked, NSFW content, and Tweets you've already seen.

The service that is responsible for constructing and serving the For You timeline is called Home Mixer. Home Mixer is built on Product Mixer, our custom Scala framework that facilitates building feeds of content. This service acts as the software backbone that connects different candidate sources, scoring functions, heuristics, and filters.

This diagram below illustrates the major components used to construct a timeline:



Let's explore the key parts of this system, roughly in the order they'd be called during a single timeline request, starting with retrieving candidates from **Candidate Sources**.

## Candidate Sources

Twitter has several Candidate Sources that we use to retrieve recent and relevant Tweets for a user. For each request, we attempt to extract the best 1500 Tweets from a pool of hundreds of millions through these sources. We find candidates from people you follow (**In-Network**) and from people you don't follow (**Out-of-Network**). Today, the For You timeline consists of 50% In-Network Tweets and 50% Out-of-Network Tweets on average, though this may vary from user to user.

## In-Network Source

The In-Network source is the largest candidate source and aims to deliver the most relevant, recent Tweets from users you follow. It efficiently ranks Tweets of those you follow based on their relevance using a logistic regression model. The top Tweets are then sent to the next stage.

The most important component in ranking In-Network Tweets is Real Graph (<https://www.ueo-workshop.com/wp-content/uploads/2014/04/sig-alternate.pdf>). Real Graph is a model which predicts the likelihood of engagement between two users. The

higher the Real Graph score between you and the author of the Tweet, the more of their tweets we'll include.

The In-Network source has been the subject of recent work at Twitter. We recently stopped using Fanout Service, a 12-year old service that was previously used to provide In-Network Tweets from a cache of Tweets for each user. We're also in the process of redesigning the logistic regression ranking model which was last updated and trained several years ago!

## Out-of-Network Sources

Finding relevant Tweets outside of a user's network is a trickier problem: *How can we tell if a certain Tweet will be relevant to you if you don't follow the author?* Twitter takes two approaches to addressing this.

### Social Graph

Our first approach is to estimate what you would find relevant by analyzing the engagements of people you follow or those with similar interests.

We traverse the graph of engagements and follows to answer the following questions:

- *What Tweets did the people I follow recently engage with?*
- *Who likes similar Tweets to me, and what else have they recently liked?*

We generate candidate Tweets based on the answers to these questions and rank the resulting Tweets using a logistic regression model. Graph traversals of this type are essential to our Out-of-Network recommendations; we developed GraphJet (<https://www.vldb.org/pvldb/vol9/p1281-sharma.pdf>), a graph processing engine that maintains a real-time interaction graph between users and Tweets, to execute these traversals. While such heuristics for searching the Twitter engagement and follow network have proven useful (these currently serve about 15% of Home Timeline Tweets), embedding space approaches have become the larger source of Out-of-Network Tweets.

### Embedding Spaces

Embedding space approaches aim to answer a more general question about content similarity: *What Tweets and Users are similar to my interests?*

Embeddings work by generating numerical representations of users' interests and Tweets' content. We can then calculate the similarity between any two users, Tweets or user-Tweet pairs in this embedding space. Provided we generate

accurate embeddings, we can use this similarity as a stand-in for relevance.

One of Twitter's most useful embedding spaces is SimClusters (<https://dl.acm.org/doi/10.1145/3394486.3403370>). SimClusters discover communities anchored by a cluster of influential users using a custom matrix factorization algorithm (<https://github.com/twitter/sbf>). There are 145k communities, which are updated every three weeks. Users and Tweets are represented in the space of communities, and can belong to multiple communities. Communities range in size from a few thousand users for individual friend groups, to hundreds of millions of users for news or pop culture. These are some of the biggest communities:

We can embed Tweets into these communities by looking at the current popularity of a Tweet in each community. The more that users from a community like a Tweet, the more that Tweet will be associated with that community.

## Ranking

The goal of the For You timeline is to serve you relevant Tweets. At this point in the pipeline, we have ~1500 candidates that may be relevant. Scoring directly predicts the relevance of each candidate Tweet and is the primary signal for ranking Tweets on your timeline. At this stage, all candidates are treated equally, without regard for what candidate source it originated from.

Ranking is achieved with a ~48M parameter neural network that is continuously trained on Tweet interactions to optimize for positive engagement (e.g. Likes, Retweets, and Replies). This ranking mechanism takes into account thousands of features and outputs ten labels to give each Tweet a score, where each label represents the probability of an engagement. We rank the Tweets from these scores.

## Heuristics, Filters, and Product Features

After the Ranking stage, we apply heuristics and filters to implement various product features. These features work together to create a balanced and diverse feed. Some examples include:

- **Visibility Filtering:** Filter out Tweets based on their content and your preferences. For instance, remove Tweets from accounts you block or mute.
- **Author Diversity:** Avoid too many consecutive Tweets from a single author.
- **Content Balance:** Ensure we are delivering a fair balance of In-Network and Out-of-Network Tweets.
- **Feedback-based Fatigue:** Lower the score of certain Tweets if the viewer has provided negative feedback around it.
- **Social Proof:** Exclude Out-of-Network Tweets without a second degree connection to the Tweet as a quality safeguard. In other words, ensure someone you follow engaged with the Tweet or follows the Tweet's author.
- **Conversations:** Provide more context to a Reply by threading it together with the original Tweet.
- **Edited Tweets:** Determine if the Tweets currently on a device are stale, and send instructions to replace them with the edited versions.

## Mixing and Serving

At this point, Home Mixer has a set of Tweets ready to send to your device. As the last step in the process, the system blends together Tweets with other non-Tweet content like Ads, Follow Recommendations, and Onboarding prompts, which are returned to your device to display.

The pipeline above runs approximately 5 billion times per day and completes in under 1.5 seconds on average. A single pipeline execution requires 220 seconds of CPU time, nearly 150x the latency you perceive on the app.

The goal of our open source endeavor is to provide full transparency to you, our users, about how our systems work. We've released the code powering our recommendations that you can view [here](https://github.com/twitter/the-algorithm) (<https://github.com/twitter/the-algorithm>) (and [here](https://github.com/twitter/the-algorithm-ml) (<https://github.com/twitter/the-algorithm-ml>)) to understand our algorithm in greater detail, and we are also working on several features to provide you greater transparency within our app. Some of the new developments we have planned include:

- A better Twitter analytics platform for creators with more information on reach and engagement
- Greater transparency into any safety labels applied to your Tweets or accounts
- Greater visibility into why Tweets appear on your timeline

# What's Next?

Twitter is the center of conversations around the world. Every day, we serve over 150 billion Tweets to people's devices. Ensuring that we're delivering the best content possible to our users is both a challenging and an exciting problem. We're working on new opportunities to expand our recommendation systems—new real-time features, embeddings, and user representations—and we have one of the most interesting datasets and user bases in the world to do it with. We are building the town square of the future. If this interests you, please consider [joining us](#).

*Written by the Twitter Team.*



[.\(https://www.twitter.com/Twitter\)](https://www.twitter.com/Twitter)

Twitter