

# Mastodon Network Analysis

Ved Kothavade

University of Maryland, College Park  
College Park, MD  
vedk@terpmail.umd.edu

Phan Ngyuen

University of Maryland, College Park  
College Park, MD  
phan2003@terpmail.umd.edu

**Abstract**—The Mastodon network is a large area of research for sociologists and computer scientists alike, due to its distributed nature and unique audience. In this report, we crawl Mastodon network and measure statistics about the nature of the instances that make it up, and the connections between them. There exists research done on the content of the posts on Mastodon instances, but there is none to be found regarding the nature of the servers that run instances themselves, nor their connectivity with each other.

## I. INTRODUCTION

We want to explore characteristics of the Mastodon network related to the geographical distribution of nodes, cloud providers nodes are hosted on, and more.

## II. METHODOLOGY

While writing a crawler is simple in principle, to mitigate time constraints we began with a seed list of alive nodes created by an external, open-source Fediverse crawler. This crawler is updated every 6 hours. From the list created by this crawler, we filter down to Fediverse nodes hosting Mastodon or Mastodon-compatible software (as opposed to, for example, WordPress), and then to nodes which supported the `peers` API we leverage to determine node connectivity.

From this filtered list of nodes, we then collected data about each node leveraging the MaxMind GeoIP databases and by resolving the IPs for the domains, and looking up their ASN, AS organization names, country codes. Furthermore, we use an `instance` API to collect user and post counts for all instances.

Upon collecting this data about nodes in a database, we use the neo4j graph database to insert all nodes, the properties of these nodes, and finally create `PEERS_WITH` relationships between peers as defined by the aforementioned `peers` API.

We believe that this is a representative dataset of the Mastodon network, and thus feel that the analyses we conduct from this graph dataset represent the true network well.

## III. ANALYSIS

### A. Location of Instances

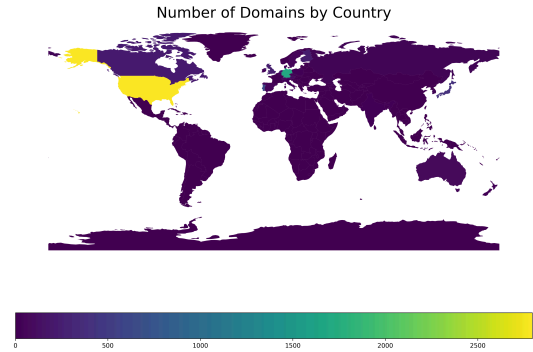


Fig. 1: Heatmap of Instance Locations

TABLE I: INSTANCE COUNTS IN TOP 10 COUNTRIES

Country	Count
US	2795
FR	1735
DE	1700
PT	652
JP	415
GB	249
FI	234
CA	229
NL	181
SG	138

## B. Cloud Providers

TABLE II: INSTANCE COUNTS BY CLOUD PROVIDER

Cloud Provider	Count
None	6200
OVH	2245
Hetzner	1207
DigitalOcean	493
AWS	200
GCP	68
Azure	37

