VAM: Supplemental Materials

Fred Mubang and Lawrence O. Hall

Supplemental information to the main VAM paper

University of South Florida
Department of Computer Science

Email: fmubang@usf.edu (Fred Mubang), lohall@usf.edu (Lawrence O. Hall)

CONTENTS

I. INTRODUCTION

This document contains supplementary material for the VAM paper.

II. ADDITIONAL DATA COLLECTION INFORMATION

A. New and Old User Average Daily Frequencies

Table I on page 8 shows the percentages and Figure 1 shows the bar plot for the average daily new and old user ratio per topic. These ratios were obtained by calculating the average number of new users per day, and then calculating the average old users per day, per topic. These values were then normalized between 0 and 100%. In Figure 1, the orange bars represent the old user average frequencies, and the blue bars represent the new user average frequencies. In the plot it can be seen that on average per day for each topic, most of the users are old, however for some topics there are still a considerable amount of new users. For example, the *other/censorship_outage* topic has about 60% new users on average per day, and the *other/anti_socialism* topic has roughly 50% new users on average per day.

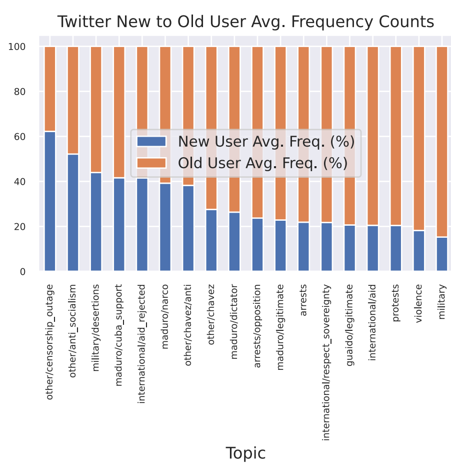


Figure 1: The new/old user proportion plots for Twitter.

B. Twitter Topic Graph Information

Table II on page 8 contains the network statistics for all 18 topic networks in Twitter.

III. VOLUME MODULE ADDITIONAL METHODOLOGY INFORMATION

This section contains additional implementation details from VAM's Volume Prediction Module.

A. Feature Configuration By Platform and Volume Lookback Factor

As mentioned in the paper, the time series features differed across all 64 models. Table III on page 8 describes the 16 potential time series features for a given sample, and Table IV on page 9 describes which model had which time series features.

To better understand how the features are configured we shall describe an example in Table IV on page 9. Take, for example, the first row for the model, *VAM-TR-24*. This is a *VAM* model trained on Twitter and Reddit time series features. The *Time Series Used* column illustrates which time series were fed in from Table III on page 8. It says that features $1, 2, 3, 7, 8, 9, 16$ were used. If you look at Table III on page 8 you will see that these are all time series related to Twitter and Reddit, which explains the "TR" in the model tag. Note that each platform in IV on page 9 is represented with a letter. "T" stands for "Twitter", "R" stands for "Reddit", "G" stands for "GDELT", and "Y" stands for "Youtube".

The *volume lookback factor* column for *VAM-TR-24* indicates it's "24 hours". So, for each time series category listed in the *time series used* category, a time series of 24 elements is placed into the feature set for the *VAM-TR-24*. Since there are 7 time series, a *volume lookback factor* of 24, and 18 topic features, the calculation for number of features is 7 * 24 + 18, which equals 186. Therefore, the dataset for the *VAM-TR-24* model was comprised of 186 features (as shown by the *total_fts* column).

Table V on page 9 illustrates how the features were used to create each sample.

B. Log Normalization in Volume Prediction Module

For all Twitter-related features, we rescaled the data by first taking the natural log of all samples twice. Before taking the logs, we added 1 to all values in order to avoid taking the natural log of 0. Adding 1 is especially important for the Youtube data, which was more sparse. For the Youtube, Reddit, and GDELT NumMentions features, we only took the natural log once, because the magnitude of those features was not as large as the Twitter ones.

We did not take the natural log of the AvgTone or Goldsteinscale GDELT features because those features had a lower bound of negative infinity. It is not possible to take the natural log of negative numbers.

IV. VOLUME MODULE ADDITIONAL PREDICTION RESULTS

This section contains additional results for the Volume Prediction Module.

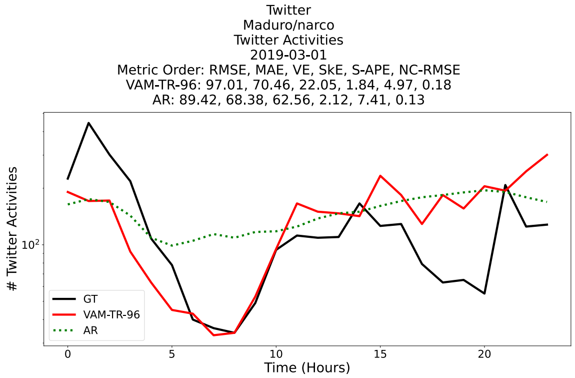
A. Volume Prediction Full Model Rank Tables

Table VI on page 10 contains all Twitter VAM and baseline models ranked by Overall Normalized Metric Error (ONME).

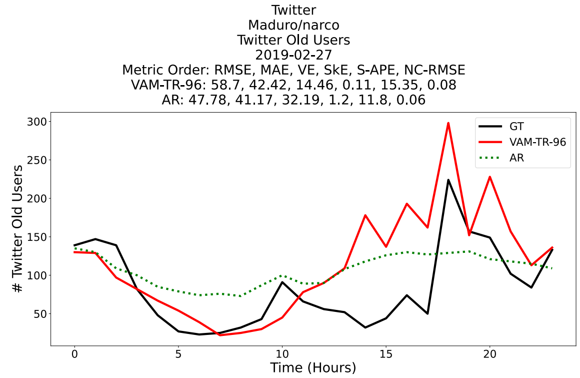
B. Volume Prediction Module Results Across Topics

Figure 3 on page 4 contains bar plots comparing the *VAM-TR-96* model against the best baseline model per each topic and metric pair. Tables VII on page 10, VIII on page 11, IX on page 11, X on page 12, XI on page 12, and XII on page 12 show per-topic results for the RMSE, MAE, NC-RMSE, S-APE, SkE, and VE metrics, respectively.

For the RMSE metric, VAM won against the best baselines 18 out of 18 times. For MAE, VAM won 17 times; for



(a) Maduro/Narco Activities on March 1st



(b) Maduro/Narco Old Users on Feb 27th

Figure 2: Here are some examples in which a baseline model had better RMSE and MAE performance than the *VAM-TR-96*, but worse performance in other metrics. In (a), the RMSE and MAE of the AR model is 89.42 and 68.38, respectively, whereas for VAM it's 97.01 and 70.46, respectively. However, as one can see, visually the VAM model prediction (red) looks more similar to the ground truth (black) curve within the first 15 hours or so of the simulation. The VAM prediction manages to capture the major dip in the ground truth, unlike the AR prediction. This might be captured in the prediction metrics in which VAM had a Volatility Error and Skewness Error of 22.05 and 1.84, versus the AR model's results of 62.56 and 2.12, respectively. In 2b, a similar phenomenon can be observed. The AR model has better RMSE and MAE metrics than VAM (47.78 and 41.17 vs. VAM's 58.7, and 42.42, respectively), however, VAM has better VE and SkE metric results (VAM has 14.46 and 0.11 vs. the AR's 32.19 and 1.11, respectively).

Normalized Cumulative RMSE (NC-RMSE), VAM won 12 times; for Symmetric Absolute Percentage Error (S-APE), VAM won 12 times; for Skewness Error (SkE), VAM won 9 times; and for Volatility Error (VE), VAM won 13 times.

Overall, VAM outperformed the best baselines 81 out of 108, or 75% of the time. VAM performed particularly well at the "exact volume over time" metrics (RMSE and MAE). It performed decently for the "magnitude" or "scale" metric (S-APE). It also performed decently on the Volatility Error metric, which measures how well the volatility, or standard deviation of the time series matches that of the ground truth. It struggled the most with the Skewness Error metric, which measures the assymetry of the time series.

C. Issues with Using Only RMSE and MAE as Metrics

RMSE and MAE are commonly used metrics for time series regression problems, however we found that they can have their limitations. When plotting the VAM models against the baseline models, we found that there are some instances in which the baseline time series has better RMSE and MAE results than the VAM prediction, however when visually inspecting the time series plots, the VAM models seem to better match the ground truth time series. For this reason, we also utilize 4 more metrics that measure other elements of time series prediction performance besides just the "exact timing" measurement of RMSE and MAE. Figure 2 contains 2 examples of this phenomenon.

V. User-Assignment Additional Methodology Information

A. User-Assignment: Various User Definitions

In order to better understand how the User-Assignment algorithm works, in this section we will further describe the differences between *new* and *old* users, as well as *child* and *parent* users.

B. Child and Parent Users

For any given edge, $(u, v, w(u, v, t))$, user u is the *child user*. This means that u either created a post (such as a tweet), or u reacted to a post created by user v. Likewise for an edge, $(u, v, w(u, v, t))$, user v is the *parent* user. This means that v has had her post reacted to in some way by u. Note that if u creates a post, she is both a child and parent user. In the edge list, E_t, this would would be a self loop, which could be written as $(u, u, w(u, u, t)) == (u, v, w(u, v, t))$.

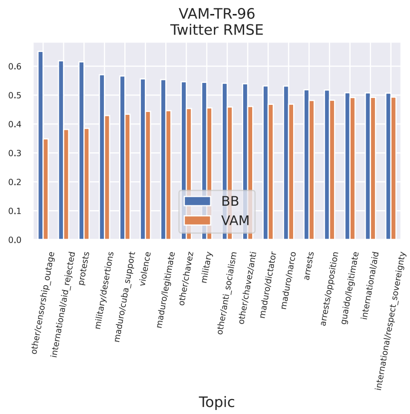
C. New and Old Users

An *old* and *new* user can be described as follows. A user, u is *old* at time step t if u appears in G_t in the temporal graph sequence G, and it also appeared in at least one of the previous graphs spanning G_1 up to G_{t-1}.

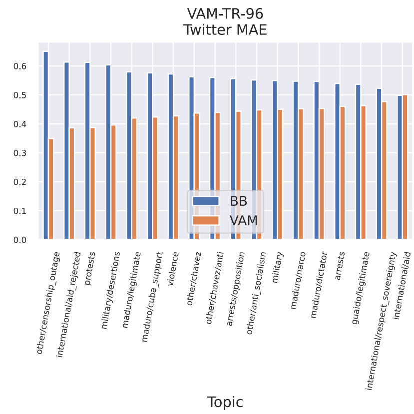
A user, u is *new* at time step t if u appears in G_t in the temporal graph sequence G, but *has not* appeared in any of the previous graphs from G_1 up to G_{t-1}.

D. User-Assignment Symbols

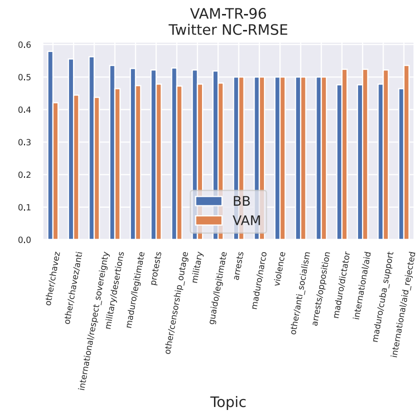
Table XIII on page 13 contains various symbols used in the User-Assignment algorithm for quick reference.



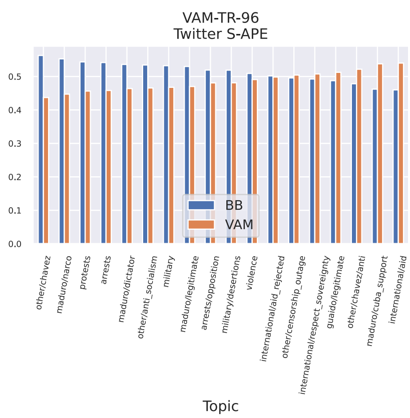
(a) RMSE (Lower is Better)



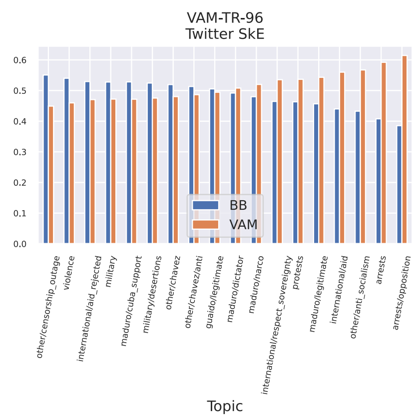
(b) MAE (Lower is Better)



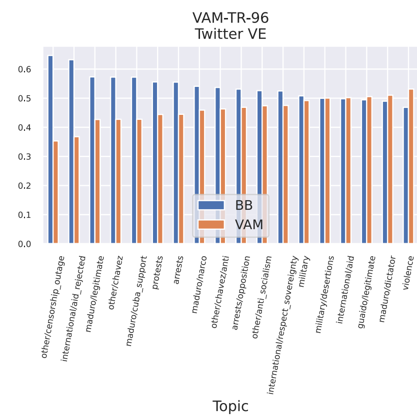
(c) NC-RMSE (Lower is Better)



(d) S-APE (Lower is Better)



(e) Skewness Error (Lower is Better)



(f) Volatility Error (Lower is Better)

Figure 3: These barplots show VAM (orange) vs. the best baseline, BB, (blue) across various topics. The metric result per topic for both models were normalized between 0 and 1 for easier visualization.

E. Additional User-Assignment Implementation Details

If the *VP-Module* predicts more users than actions at some future timestep, $T+1$, then VAM simply changes the number of actions to equal the number of users. For example, if 2 users and 1 action were predicted to occur at $T+1$, then VAM simply changes the number of predicted actions to be 2.

If, for some future timestep $T+1$, the *VP-Module* predicted more old users than exist in the recent history table, H^{recent}, then VAM simply changes the number of predicted active old users to equal the number of old users that do exist in H^{recent}. For example, if the *VP-Module* predicts 100 old users will be active at $T+1$, but H^{recent} only contains 90 old users, then VAM will change the number of predicted old users to be 90, and then sample all 90 active old users from H^{recent} for user-activity assignment.

VI. User-Assignment Algorithm Step-by-Step in Detail

In this section we describe in more detail the steps to assign a user to an action via the user-assignment module.

A. User Assignment - Inputs and Outputs

The inputs to the algorithm are the full temporal graph G, the number of output timesteps to be predicted, S, the user-assignment lookback factor, L^{user}, the volume prediction matrix $\hat{\boldsymbol{Y}} \in \mathbb{R}^{3\times S}$, the old user index *old_idx*, the new user index new_idx, and the activity index act_idx. The three indices are used to access the predicted number of old users, new users, and activities on the predicted timestep of interest from $\hat{\boldsymbol{Y}}$. The output of the *Assign_Users* algorithm is the temporal graph sequence, $\hat{G}^{future}$.

B. Initializations

At the beginning of the algorithm, G^{recent} is constructed with the $Get\_Recent\_Temporal\_Graph$ function using the full temporal graph, G, and the lookback parameter, L^{user}. As mentioned in the main paper, the lookback factor parameter

L^{user} is used to determine the number of snapshots to use. For example, if $L^{user} = 5$, then only the 24 most recent graphs in sequence G will be used to make H^{recent}. The assumption here is that recent history is all that is needed to make temporal network predictions.

Furthermore, note that the length of the full temporal graph sequence G is T. So, for example, if G contains $T = 100$ graphs, and $L^{user} = 5$, then only graphs 96 up to 100 will be in the sequence G^{recent}. Or, in other words, only graphs $T - L^{user} + 1$ up to T are included in G^{recent}. Then, an empty array, $\hat{G}^{future}$ is created.

Recall that the number of prediction timesteps of interest is S. There is a for loop that iterates S times for each $s \leq S$. This s represents the current prediction timestep of interest. At each timestep s, the number of old users, new users, and activities are retrieved from their respective location in the $\hat{\boldsymbol{Y}}$ matrix.

C. The History Table

The sequence G^{recent} is used to construct a history table, H^{recent}, that contains the event tuples from G. For each iteration s, the *UA-Module* first uses G^{recent} to construct a recent history table called H^{recent}. This table can be thought of as a hash table. Each key into the table is an integer representing a time step, t, spanning from $T - L^{user} + s$ up to $T + s - 1$. This range can be also be defined as the *current lookback period of interest* Each time step key maps to an event table, H_t^{recent}. This table can be thought of as an array of "event tuples", each of which with the following form:

$$(u, v, w(u, v, t), \mathbb{I}_{V_t^{new}}(u), \mathbb{I}_{V_t^{new}}(v))$$

. u is the child user and v is the parent user. $w(u, v, t)$ is a numerical value that represents how many times u interacted with v at time step t in G_t.

V_t is the set of all users in G at some particular timestep t. V_t^{new} is the set of all new users in V_t. $\mathbb{I}_{V_t^{new}}$ is an indicator function that returns 1 if a particular user, u was in the new user set, V_t^{new} at time step t, and 0 otherwise.

D. Retrieving Old User Candidates

The *Get_Active_Old_User_Candidates* function is used to construct a table of each old user candidates (W^{old_cand}) from H^{recent} and their likelihood of being active at time s.

This table can be thought of as an array of tuples, each with the following form: $(u, p^{old\_act}(u, T + s))$. The term $p^{old\_act}(u, T + s)$ represents the probability that user u will be active at time step $T + s$. This probability is obtained by calculating, for each user u, the normalized average activity frequency of u during the *current lookback period of interest*.

The assumption here is that a user's future probability of acting is equivalent to his past probability of acting.

E. Retrieving Most Likely Old Users From Candidates

The *Get_Most_Likely_Active_Old_Users* function is then used to retrieve the set of most likely old users ($\hat{O}_s$) from this table of candidates, as well as a table containing their re-weighted probabilities (W^{old}).

In order to retrieve ($\hat{O}_s$) and W^{old}, *VAM* performs a weighted random sample on the W^{old_cand} table in order to create the set of most likely active old users at time $T + s$. The weights used for this random sample are the activity probabilities ($p^{old\_act}(u, T + s)$) for each old user that were calculated in the previous step. We call the set of predicted active users, $\hat{O}_s$. Furthermore, *VAM* creates a new weight table, called W^{old}, which is the same as W^{old_cand} minus any users that were not chosen by the weighted random sample. This new table, W^{old} is needed when the time comes for *VAM* to predict how many actions each old user will perform. Old users with a higher probability of acting, a.k.a. $p^{old\_act}(u, T + s)$, are more likely to perform more actions.

F. Creating the New User Set

Next, the set of new users ($\hat{N}_s$) is generated using the *Generate_New_Users* function. Recall that the number of new users is known because that was predicted from the *Volume-Prediction Module* and it is contained in the matrix $\hat{\boldsymbol{Y}}$.

G. Assigning Attributes to New Users

The question that remains at this point is "How does one decide what actions the new users will perform?" *VAM* does this by constructing a *New User Archetype Table*. This table is comprised of *recently active users*. These are users that have appeared as new within the lookback factor period of $T + s - L^{user}$ up to $T + s - 1$, which can also be referred to as the "recent history". The opposite of a *recently active* user would be a *long-acting user*, which would be a user who has appeared in G before timestep $T + s - L^{user}$. The assumption behind the archetype table is that new users in the future will behave in a similar manner to previously new users from the recent past.

The *New User Archetype Table* table contains the following information: (1) the name of the *recently active* a.k.a. *archetype user*, (2) the probability that this archetype would be active in any given timestep (e.g. via tweeting or retweeting), and (3) the probability that this archetype will be "influential" in any given timestep. In Twitter, probability of influence is measured by how often a user is retweeted.

With this in mind we now define a *new user archetype* record as follows:

$$(u^{arch}, p^{act\_arch}(u^{arch}), p^{infl\_arch}(u^{arch}))$$

. The term $p^{act\_arch}(u^{arch})$ is a weight that describes how likely it is that a user of archetype, u^{arch} will be active in future time step. The term, $p^{infl\_arch}(u^{arch})$ describes how likely it is that a user of archetype u^{arch} will be influential in some future time step. We define influence as the quantity by which other users will respond to a post created by u^{arch} in some social media platform with regards to topic q.

As an example, say there is some recently active user named Carol with an action probability of 0.2 and an influence probability of 0.1. Since Carol is a recently active user, she will be considered an archetype of new user and added to the new user archetype table, W^{new_arch}. The record for "Carol" will be as follows:

$$(``Carol'', 0.2, 0.1)$$

. Now, "Carol" archetype could be applied to a new generated user arbitrarily given the identifier "Phil". Even though Phil has been generated as a new user, he still needs to be assigned a probability of activity and probability of influence. *VAM* will randomly sample a user archetype from W^{new_arch} in order to assign attributes to Phil. If *VAM* randomly selects user archetype Carol, then *VAM* will assign Phil an activity probability of 0.2 and an influence probability of 0.1. These values are, of course, then normalized relative to the other generated new users so that all users' probabilities lie between 0 and 1.

VAM then iterates over every new user $u^{new} \in \hat{N}_s$ and performs a weighted random sample to select a new user archetype tuple from W^{new_arch}. This process then yields a new table, called the *new user attribute table*, or W^{new}. This table can be thought of as any array of tuples of the following form.

$$(u^{new}, u^{arch}, p^{act\_arch}(u^{new}), p^{infl\_arch}(u^{new}))$$

u^{new} is the new user of interest. u^{arch} is the archetype that this new user was created from. $p^{act\_arch}(u^{new})$ and $p^{infl\_arch}(u^{new}))$ are u^{new}'s probabilities of activity and influence, respectively. They are equivalent to the probability and influence probabilities of user archetype, u^{arch}. In other words, each user $u^{new} \in \hat{N}_s$ was assigned probability and activity attributes from some user archetype, $u^{arch} \in U^{arch}$. Note that U^{arch} is the set of all user archetypes.

The *New User Archetype Table* is then used to assign the activity and influence probabilities to each new user in $\hat{N}_s$. These probabilities are stored in W^{new}.

H. Creating the Old and New User Parent Tables

Now, *VAM* needs the most likely sets of parents that the old and new users will interact with. To that end, it creates what we call *parent distribution tables*.

Firstly, using the $Create\_Old\_User\_Parent\_Table$ function, the old parent distribution table, D^{old_parent}, is created. This table can be thought of as a hash table, in which each key is a user, u, and each user maps to a parent distribution table for that particular user. Each table has the form: $(v, p^{edge}(u, v, T+s))$. The term $p^{edge}(u, v, T+s))$ represents the probability that an edge will form between u and v at time $T+s$.

Next, VAM must create a parent distribution table for the new users using the *Create_New_User_Parent_Table*. In order to do so, it iterates over every new user record in W^{new}. It then checks which user archetype, u^{arch} the new user u^{new} was created from. Recall that each new user archetype was created from users who have actually existed in the data, so it is possible for *VAM* to collect information regarding who their previous parents were. *VAM* will then create a parent distribution table for each u^{arch} and will assign this parent distribution table to the appropriate new user u^{new}. The final table created, D^{new_parent}, will be a new user parent distribution table, similar to D^{old_parent}. Each key of the hash table is a new user, u^{new}, and each key hashes into a new user parent distribution table of the form $(v, p^{edge}(u^{new}, v, T+s))$. The term $p^{edge}(u^{new}, v, T+s))$ represents the probability that an edge will form between u^{new} and v at time $T+s$.

I. Creating the Links

At this point, *VAM* now has the information it needs to perform link prediction. To that end, it uses the function, $Create\_Links$ to perform link prediction and create the final graph, G_s^{future}. The arguments to $Create\_Links$ are $\hat{O}_s$,$\hat{N}_s$,num_acts , W^{old}, W^{new}, D^{old}, and D^{new}. Note that *VAM* "knows" the total edge weight of all links in G_s^{future} because the *Volume-Prediciton Module* predicted the total number of activities for each timestep $s \leq S$, hence the use of the argument, num_acts.

J. Updating the Recent Temporal Graph Sequence

The predicted graph, $\hat{G}^{future}$ is then used to update G^{recent}. The user-assignment for-loop then continues $S-1$ more times until the full $\hat{G}^{future}$ graph is predicted such that $\hat{G}^{future} = \{\hat{G}_1^{future}, \hat{G}_2^{future}, ..., \hat{G}_S^{future}\}$.

K. User-Assignment Graphic

Figure 4 on page 14 is an illustration of all 7 steps of the User-Assignment Algorithm.

VII. User-Assignment Additional Results

This section contains additional results of the User Assignment Module from the main paper.

A. Unweighted Jaccard Similarity Results

In addition to the Weighted Jaccard Similarity results shown in the main paper, we also calculated the Unweighted Jaccard Similarity results for the full user set, highly influential user cluster, and lowly influential user cluster. Figure 5 on page 15 contains bar plots for these results. As one can see, VAM also outperformed the Persistence Baseline in these 3 categories as well.

Let A represent the set of the actual old users within a particular hour, and let P represent the predicted set of old users within a particular hour. The unweighted Jaccard Similarity is as follows:

$$J(A, P) = \frac{|A \cap P|}{|A \cup P|}$$

.

B. Jaccard Similarity Tables

Table XIV on page 16 contains the unweighted Jaccard Similarity results for the full set of old users. Table XV on page 16 contains the unweighted JS results for the highly influential cluster of old users. Table XVI on page 16 contains the unweighted JS results for the lowly influential user cluster.

Table XVII on page 16 contains the weighted JS results for the full set of old users. Table XVIII on page 17 contains the weighted JS results for the highly influential cluster of old users. Table XIX on page 17 contains the weighted JS results for the lowly influential cluster of old users.

C. Earth Mover's Distance Tables

Tables XX on page 18, XXI on page 18, and XXII on page 18 contain the Earth Mover's Distance results for the full set, highly-influential cluster, and lowly-influential cluster of users.

D. Relative Hausdorff Distance Tables

Tables XXIII on page 18, XXIV on page 19, and XXV on page 19 contain the contain the Relative Hausdorff Distance results for the full set, highly-influential cluster, and lowly-influential cluster of users.

Twitter Avg. New and Old User Frequencies Per Day		
Topic	**New User Avg. Freq. (%)**	**Old User Avg. Freq. (%)**
other/censorship_outage	62.22	37.78
other/anti_socialism	52.15	47.85
military/desertions	43.94	56.06
maduro/cuba_support	41.61	58.39
international/aid_rejected	41.53	58.47
maduro/narco	39.21	60.79
other/chavez/anti	38.24	61.76
other/chavez	27.55	72.45
maduro/dictator	26.34	73.66
arrests/opposition	23.7	76.3
maduro/legitimate	22.84	77.16
arrests	21.88	78.12
international/respect_sovereignty	21.74	78.26
guaido/legitimate	20.64	79.36
international/aid	20.49	79.51
protests	20.43	79.57
violence	18.2	81.8
military	15.27	84.73

Table I: Twitter Avg. New and Old User Frequencies Per Day

Twitter Graph Information			
Topic	**# Nodes**	**#Edges**	**Total Edge Weight (a.k.a. Total # of Activities)**
military	457,200	2,458,703	4,580,984
international/aid	484,405	2,018,902	3,530,265
protests	451,542	2,058,608	3,083,175
violence	400,141	1,957,442	3,031,137
guaido/legitimate	355,381	1,437,221	2,122,211
international/respect_sovereignty	205,180	815,250	1,635,717
maduro/dictator	355,552	1,137,656	1,528,799
other/chavez	222,025	697,542	1,154,887
arrests	175,685	687,628	935,191
arrests/opposition	147,454	551,617	718,539
international/aid_rejected	211,168	518,668	662,886
maduro/legitimate	94,424	351,705	655,588
other/chavez/anti	142,556	300,346	398,892
military/desertions	125,257	285,934	365,718
maduro/narco	92,208	190,973	244,958
other/anti_socialism	119,519	184,152	238,342
maduro/cuba_support	62,904	112,281	153,640
other/censorship_outage	62,603	110,097	122,581

Table II: Twitter Network Statistics

Time Series Index Label	**Time Series Description**
1	New user volume time series for a given topic in Twitter.
2	Old user volume time series for a given topic in Twitter.
3	Activity volume time series for a given topic in Twitter.
4	New user volume time series for a given topic in Youtube.
5	Old user time series for a given topic in Youtube.
6	Activity volume time series for a given topic in Youtube.
7	Activity volume time series across all topics in Twitter.
8	New user volume time series across all topics in Twitter.
9	Old user volume time series across all topics in Twitter.
10	Activity volume across all topics in Youtube.
11	New user volume time series across all topics in Youtube.
12	Old user volume time series across all topics in Youtube.
13	The GDELT AvgTone time series.
14	The GDELT GoldsteinScale time series.
15	The GDELT NumMentions time series.
16	Activity volume time series in Reddit.

Table III: The table of all possible time series feature categories.

Model	**Platforms Used**	**Volume Lookback Factor** (L^{vol})	**Time Series Used**	**Num Time Series Used**	**Num Static Fts.**	**Total Features**
VAM-TR-24	('Twitter', 'Reddit')	1 day (24 hours)	(1, 2, 3, 7, 8, 9, 16)	7	18	7 * 24 + 18 = 186
VAM-TY-24	('Twitter', 'Youtube')	1 day (24 hours)	(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12)	12	18	12 * 24 + 18 = 306
VAM-TRG-24	('Twitter', 'Reddit', 'GDELT')	1 day (24 hours)	(1, 2, 3, 7, 8, 9, 13, 14, 15, 16)	10	18	10 * 24 + 18 = 258
VAM-TG-24	('Twitter', 'GDELT')	1 day (24 hours)	(1, 2, 3, 7, 8, 9, 13, 14, 15)	9	18	9 * 24 + 18 = 234
VAM-TGY-24	('Twitter', 'GDELT', 'Youtube')	1 day (24 hours)	(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15)	15	18	15 * 24 + 18 = 378
VAM-TRGY-24	('Twitter', 'Reddit', 'GDELT', 'Youtube')	1 day (24 hours)	(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16)	16	18	16 * 24 + 18 = 402
VAM-T-72	('Twitter',)	3 days (72 hours)	(1, 2, 3, 7, 8, 9)	6	18	6 * 72 + 18 = 450
VAM-TRY-72	('Twitter', 'Reddit', 'Youtube')	3 days (72 hours)	(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 16)	13	18	13 * 72 + 18 = 954

Table IV: Some examples of what the features for each model were for Twitter.

Twitter Sample, Feature, and Output Info						
model_tag	**x_train**	**y_train**	**x_val**	**y_val**	**x_test**	**y_test**
VAM-TR-24	(17730, 186)	(17730, 72)	(2610, 186)	(2610, 72)	(378, 186)	(378, 72)
VAM-TY-24	(17730, 306)	(17730, 72)	(2610, 306)	(2610, 72)	(378, 306)	(378, 72)
VAM-TRG-24	(17730, 258)	(17730, 72)	(2610, 258)	(2610, 72)	(378, 258)	(378, 72)
VAM-TG-24	(17730, 234)	(17730, 72)	(2610, 234)	(2610, 72)	(378, 234)	(378, 72)
VAM-TGY-24	(17730, 378)	(17730, 72)	(2610, 378)	(2610, 72)	(378, 378)	(378, 72)
VAM-TRGY-24	(17730, 402)	(17730, 72)	(2610, 402)	(2610, 72)	(378, 402)	(378, 72)
VAM-T-72	(17730, 450)	(17730, 72)	(2610, 450)	(2610, 72)	(378, 450)	(378, 72)
VAM-TRY-72	(17730, 954)	(17730, 72)	(2610, 954)	(2610, 72)	(378, 954)	(378, 72)

Table V: Some examples of the Twitter datasets used. The terms "x" and "y" are used to refer to "inputs" and "outputs". For example, for *VAM*-TR-24, the training set had 17,730 samples. Each sample had 186 input features, and 72 outputs.

Twitter Overall Volume Prediction Results									
Rank	**Model**	**RMSE**	**MAE**	**VE**	**SkE**	**S-APE**	**NC-RMSE**	**Overall Normalized Metric Error**	**ONME PIFBB (%)**
1	**VAM-TR-96**	675.08053	482.97939	358.63956	0.99388	26.91419	0.11566	**0.02477**	**17.53362**
2	VAM-TY-96	687.28588	491.65766	366.44394	0.9095	27.62297	0.1172	0.02479	17.45628
3	VAM-TR-24	665.88435	467.68103	351.6489	0.95903	28.04565	0.12311	0.02479	17.45479
4	VAM-TR-48	666.11639	472.04979	356.92729	0.98988	27.50714	0.11939	0.02481	17.39281
5	VAM-TRY-96	683.30911	489.27646	365.33833	0.96186	27.19465	0.1184	0.02495	16.94269
6	VAM-TR-72	681.71863	483.31309	369.61879	0.97503	27.51952	0.11992	0.0251	16.43668
7	VAM-T-96	682.29561	488.55522	370.65342	0.99218	27.63607	0.11627	0.02512	16.3687
8	VAM-TRY-24	675.10884	477.63081	363.32588	0.96143	29.3425	0.12553	0.02536	15.57669
9	VAM-T-48	691.26144	490.17236	376.60796	0.95685	28.29505	0.12152	0.02539	15.47802
10	VAM-TY-48	687.22783	492.10927	375.16132	0.94802	29.09011	0.1208	0.02542	15.37886
11	VAM-T-24	676.82741	479.52227	360.73755	0.99542	28.99582	0.12494	0.02543	15.33253
12	VAM-TRY-48	681.46398	487.05464	369.27407	0.97438	29.0756	0.12263	0.02546	15.25608
13	VAM-TY-24	672.23035	478.00015	363.38869	0.97239	29.87538	0.12512	0.02546	15.24233
14	VAM-T-72	696.14708	494.04453	376.05549	0.97999	28.0642	0.12094	0.02549	15.14001
15	VAM-TRY-72	694.38245	493.19926	374.43818	0.97327	28.25191	0.12379	0.02556	14.92232
16	VAM-TY-72	704.26575	502.91803	382.18591	0.94582	28.731	0.12212	0.02567	14.52696
17	VAM-TGY-96	708.20652	513.72549	378.1994	0.96512	29.58298	0.12157	0.02594	13.65561
18	VAM-TG-72	702.48892	508.86155	376.30395	1.02945	29.29186	0.11839	0.02597	13.55621
19	VAM-TGY-24	709.13439	506.33194	366.01256	0.94612	29.76757	0.12996	0.02599	13.4829
20	VAM-TGY-72	711.09112	515.09256	379.65466	0.95047	29.81337	0.123	0.026	13.4328
21	VAM-TG-96	704.71508	512.29108	377.15018	1.02737	29.57371	0.11845	0.02605	13.26695
22	VAM-TRGY-24	710.29206	508.83061	370.79312	0.95675	30.21941	0.12853	0.02613	12.99762
23	VAM-TG-24	707.60466	509.9741	368.84375	1.00972	30.44914	0.12416	0.02622	12.72294
24	VAM-TRG-96	710.73371	513.49985	380.95541	1.03355	29.56703	0.12101	0.02626	12.57452
25	VAM-TGY-48	720.18705	521.13202	381.14999	0.9766	30.06877	0.12253	0.02626	12.57155
26	VAM-TRGY-96	716.40519	518.43713	383.11759	0.99343	29.76854	0.12422	0.02633	12.34702
27	VAM-TRG-24	716.32768	514.60813	369.99587	0.98663	30.5367	0.12812	0.02637	12.19785
28	VAM-TRGY-72	717.91641	520.06849	384.3774	1.00422	30.33378	0.12461	0.02651	11.73426
29	VAM-TRG-72	716.30463	516.74673	381.03	1.05843	29.83445	0.1222	0.02652	11.7269
30	VAM-TRGY-48	724.80274	523.01936	387.91648	1.00794	30.7914	0.12383	0.02668	11.1905
31	VAM-TG-48	729.39187	529.68721	385.59907	1.01644	30.72096	0.12333	0.02674	10.97272
32	VAM-TRG-48	725.92116	526.16667	380.91197	1.03678	30.86233	0.1236	0.02676	10.91415
33	Persistence_Baseline	888.9082	619.26606	454.85759	0.96809	29.42484	0.15699	0.03004	0.0
34	MA	922.13789	701.64627	444.88704	1.38811	37.35475	0.14152	0.0333	-10.86283
35	ARMA	1068.57479	823.89253	531.86923	1.24775	34.36105	0.1353	0.03489	-16.14924
36	AR	1174.3006	904.72248	605.11153	1.37021	34.54514	0.12422	0.0372	-23.83783
37	ARIMA	1321.44676	1034.54112	658.98357	1.21444	37.36525	0.14517	0.04026	-34.04431

Table VI: Overall Twitter volume prediction results.

Twitter VAM RMSE Baseline Comparisons by Topic									
Topic	**VAM-TR-96 RMSE**	**ARIMA RMSE**	**ARMA RMSE**	**AR RMSE**	**MA RMSE**	**PB RMSE**	**Best Baseline RMSE**	**Best Baseline Name**	**PIFBB (%)**
other/censorship_outage	**231.58**	2974.22	1604.84	2780.86	851.13	432.57	432.57	PB	**46.46**
international/aid_rejected	**947.85**	5350.74	3185.52	4087.05	1837.08	1538.11	1538.11	PB	**38.38**
protests	**698.22**	1182.41	1163.84	1127.14	1189.02	1115.51	1115.51	PB	**37.41**
military/desertions	**447.63**	1040.91	634.69	675.43	594.88	796.13	594.88	MA	**24.75**
maduro/cuba_support	**135.34**	303.48	229.91	192.1	176.6	207.06	176.6	MA	**23.36**
violence	**1733.91**	2650.92	2319.72	2172.55	2360.68	2444.25	2172.55	AR	**20.19**
maduro/legitimate	**167.25**	228.79	207.5	208.68	217.81	222.42	207.5	ARMA	**19.4**
other/chavez	**238.67**	312.87	297.8	287.43	317.53	296.74	287.43	AR	**16.96**
military	**2265.17**	3189.68	3489.88	3478.29	2705.04	2845.83	2705.04	MA	**16.26**
other/anti_socialism	**66.41**	81.2	79.26	78.31	85.95	80.38	78.31	AR	**15.19**
other/chavez/anti	**115.65**	143.9	135.37	139.24	140.63	147.11	135.37	ARMA	**14.57**
maduro/dictator	**490.78**	607.16	574.98	557.27	568.15	573.66	557.27	AR	**11.93**
maduro/narco	**119.25**	162.33	140.12	137.3	135.18	168.2	135.18	MA	**11.79**
arrests	**216.02**	232.6	243.52	235.2	281.7	308.89	232.6	ARIMA	**7.13**
arrests/opposition	**169.93**	189.9	192.97	182.32	241.76	221.65	182.32	AR	**6.8**
guaido/legitimate	**879.21**	914.83	923.39	908.76	926.29	1198.03	908.76	AR	**3.25**
international/aid	**2677.36**	3597.67	3231.64	3322.84	3128.41	2761.74	2761.74	PB	**3.06**
international/respect_sovereignty	**551.22**	622.45	579.39	566.65	840.64	642.09	566.65	AR	**2.72**

Table VII: VAM Twitter RMSE comparisons to baseline

VAM Twitter MAE Baseline Comparisons by Topic									
Topic	**VAM-TR-96 MAE**	**ARIMA MAE**	**PB MAE**	**ARMA MAE**	**AR MAE**	**MA MAE**	**Best Baseline MAE**	**Best Baseline Name**	**PIFBB (%)**
other/censorship_outage	**100.99**	2130.83	188.03	1093.88	2114.93	495.51	188.03	PB	**46.29**
international/aid_rejected	**549.71**	4020.93	873.29	2322.74	2800.13	1237.82	873.29	PB	**37.05**
protests	**509.04**	961.97	804.59	947.08	925.92	981.67	804.59	PB	**36.73**
military/desertions	**285.53**	870.57	547.49	470.45	510.42	435.16	435.16	MA	**34.38**
maduro/legitimate	**115.66**	182.74	159.71	159.5	165.7	172.51	159.5	ARMA	**27.49**
maduro/cuba_support	**85.93**	247.73	134.2	175.96	141.5	116.84	116.84	MA	**26.46**
violence	**1258.09**	2145.25	1694.71	1839.37	1684.15	1848.93	1684.15	AR	**25.3**
other/chavez	**175.38**	261.29	225.52	248.58	240.24	266.25	225.52	PB	**22.23**
other/chavez/anti	**83.63**	114.73	110.6	106.57	112.12	112.57	106.57	ARMA	**21.53**
arrests/opposition	**102.0**	127.69	136.39	133.26	128.81	189.5	127.69	ARIMA	**20.12**
other/anti_socialism	**46.48**	62.82	57.19	60.9	61.77	68.85	57.19	PB	**18.73**
military	**1726.54**	2625.28	2106.43	2852.07	2847.54	2163.28	2106.43	PB	**18.03**
maduro/narco	**86.52**	130.75	124.36	108.18	107.07	104.65	104.65	MA	**17.33**
maduro/dictator	**334.07**	457.53	403.28	419.58	412.6	425.42	403.28	PB	**17.16**
arrests	**141.29**	165.5	206.29	175.01	175.5	219.62	165.5	ARIMA	**14.63**
guaido/legitimate	**553.26**	645.55	757.79	650.36	641.34	653.12	641.34	AR	**13.73**
international/respect_sovereignty	**419.45**	499.17	505.66	466.33	459.55	673.77	459.55	AR	**8.73**
international/aid	2120.06	2971.4	2111.25	2600.25	2755.72	2464.18	**2111.25**	PB	-0.42

Table VIII: VAM Twitter MAE Baseline Comparisons by Topic

Twitter NC-RMSE Baseline Comparisons by Topic									
Topic	**VAM-TR-96 NC-RMSE**	**AR NC-RMSE**	**MA NC-RMSE**	**ARMA NC-RMSE**	**ARIMA NC-RMSE**	**PB NC-RMSE**	**Best Baseline NC-RMSE**	**Best Baseline Name**	**PIFBB (%)**
other/chavez	**0.08**	0.11	0.11	0.11	0.12	0.11	0.11	PB	**26.55**
other/chavez/anti	**0.08**	0.1	0.12	0.12	0.13	0.11	0.1	AR	**23.81**
international/respect_sovereignty	**0.07**	0.09	0.16	0.11	0.11	0.09	0.09	AR	**16.54**
military/desertions	**0.13**	0.16	0.15	0.15	0.21	0.27	0.15	MA	**13.42**
maduro/legitimate	**0.09**	0.1	0.11	0.13	0.12	0.13	0.1	AR	**9.93**
protests	**0.11**	0.12	0.12	0.12	0.14	0.16	0.12	AR	**9.5**
other/censorship_outage	**0.17**	0.19	0.19	0.19	0.19	0.26	0.19	MA	**9.16**
military	**0.11**	0.12	0.14	0.13	0.12	0.14	0.12	AR	**8.09**
guaido/legitimate	**0.13**	0.14	0.14	0.14	0.15	0.18	0.14	ARMA	**6.72**
arrests	**0.12**	0.12	0.13	0.13	0.13	0.17	0.12	AR	**2.98**
maduro/narco	**0.12**	0.12	0.12	0.13	0.19	0.17	0.12	AR	**2.49**
violence	**0.12**	0.12	0.15	0.15	0.15	0.14	0.12	AR	**0.56**
other/anti_socialism	0.11	0.11	0.12	0.11	0.12	0.14	**0.11**	AR	-0.24
arrests/opposition	0.15	0.15	0.17	0.16	0.18	0.2	**0.15**	AR	-1.41
maduro/dictator	0.11	0.11	0.1	0.12	0.12	0.12	**0.1**	MA	-3.44
international/aid	0.11	0.13	0.19	0.15	0.15	0.1	**0.1**	PB	-7.14
maduro/cuba_support	0.12	0.12	0.16	0.12	0.11	0.14	**0.11**	ARIMA	-10.19
international/aid_rejected	0.15	0.13	0.16	0.14	0.15	0.2	**0.13**	AR	-11.11

Table IX: Twitter NC-RMSE Baseline Comparisons by Topic

VAM Twitter S-APE Baseline Comparisons by Topic									
Topic	**VAM-TR-96 S-APE**	**ARIMA S-APE**	**PB S-APE**	**ARMA S-APE**	**AR S-APE**	**MA S-APE**	**Best Baseline S-APE**	**Best Baseline Name**	**PIFBB (%)**
other/chavez	**17.46**	25.73	22.48	23.08	23.62	26.24	22.48	PB	**22.34**
maduro/narco	**23.16**	41.66	29.69	30.11	28.62	29.22	28.62	AR	**19.08**
protests	**32.2**	51.61	38.35	53.17	53.83	56.21	38.35	PB	**16.04**
arrests	**25.14**	29.76	31.01	31.67	29.72	36.21	29.72	AR	**15.39**
maduro/dictator	**19.05**	25.08	22.01	24.89	25.24	27.48	22.01	PB	**13.44**
other/anti_socialism	**20.83**	29.11	23.9	29.07	28.83	31.11	23.9	PB	**12.87**
military	**26.8**	33.86	30.5	32.11	36.3	37.76	30.5	PB	**12.14**
maduro/legitimate	**17.12**	28.2	19.3	27.24	29.7	32.43	19.3	PB	**11.31**
arrests/opposition	**30.99**	37.04	37.74	39.46	33.46	42.92	33.46	AR	**7.38**
military/desertions	**36.12**	57.27	38.96	59.07	59.4	57.03	38.96	PB	**7.31**
violence	**31.39**	37.34	36.61	32.96	32.56	34.1	32.56	AR	**3.59**
international/aid_rejected	**38.04**	56.87	38.29	48.59	47.47	49.7	38.29	PB	**0.66**
other/censorship_outage	39.67	69.16	38.99	56.73	60.09	60.9	**38.99**	PB	-1.73
international/respect_sovereignty	16.74	19.92	16.24	18.9	20.35	29.39	**16.24**	PB	-3.06
guaido/legitimate	28.32	26.94	29.53	28.28	28.63	28.55	**26.94**	ARIMA	-5.1
other/chavez/anti	19.82	23.11	23.29	18.17	20.5	18.68	**18.17**	ARMA	-9.06
maduro/cuba_support	32.04	40.22	27.53	33.49	29.29	31.14	**27.53**	PB	-16.39
international/aid	29.58	39.69	25.19	31.49	34.2	43.31	**25.19**	PB	-17.42

Table X: Twitter S-APE Baseline Comparisons by Topic

VAM Twitter SkE Baseline Comparisons by Topic									
Topic	**VAM-TR-96 SkE**	**ARIMA SkE**	**AR SkE**	**MA SkE**	**ARMA SkE**	**PB SkE**	**Best Baseline SkE**	**Best Baseline Name**	**PIFBB (%)**
other/censorship_outage	**1.15**	2.11	2.72	2.63	2.09	1.41	1.41	PB	**18.43**
violence	**1.03**	1.27	1.37	1.23	1.24	1.21	1.21	PB	**15.12**
international/aid_rejected	**0.96**	1.88	1.56	1.54	1.6	1.08	1.08	PB	**10.85**
military	**0.93**	1.08	1.21	1.28	1.04	1.06	1.04	ARMA	**10.2**
maduro/cuba_support	**0.75**	0.94	1.03	1.09	0.91	0.84	0.84	PB	**9.89**
military/desertions	**0.97**	1.71	2.2	2.06	2.06	1.07	1.07	PB	**9.16**
other/chavez	**0.61**	0.66	0.84	0.72	0.68	0.77	0.66	ARIMA	**6.61**
other/chavez/anti	**0.73**	0.82	0.95	0.89	0.81	0.77	0.77	PB	**5.46**
guaido/legitimate	**1.36**	1.54	1.62	1.63	1.62	1.39	1.39	PB	**2.16**
maduro/dictator	0.94	1.02	1.16	1.19	1.07	0.91	**0.91**	PB	-2.31
maduro/narco	0.91	1.24	1.08	1.1	1.17	0.84	**0.84**	PB	-7.97
international/respect_sovereignty	0.68	0.59	0.66	0.91	0.62	0.8	**0.59**	ARIMA	-15.69
protests	1.17	1.26	1.6	1.89	1.64	1.01	**1.01**	PB	-15.81
maduro/legitimate	0.69	0.58	0.83	0.77	0.61	0.73	**0.58**	ARIMA	-18.69
international/aid	0.98	0.82	0.95	1.17	0.82	0.77	**0.77**	PB	-27.61
other/anti_socialism	1.18	1.27	1.39	1.46	1.18	0.9	**0.9**	PB	-30.67
arrests	1.22	1.31	1.46	1.44	1.44	0.84	**0.84**	PB	-44.85
arrests/opposition	1.61	1.74	2.04	1.98	1.86	1.01	**1.01**	PB	-58.97

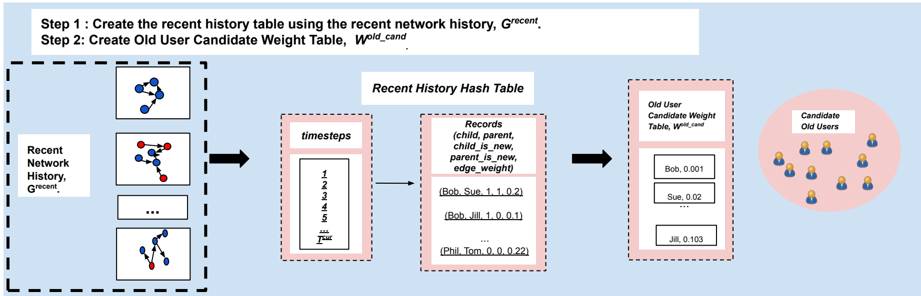
Table XI: VAM Twitter SkE Baseline Comparisons by Topic

VAM Twitter VE Baseline Comparisons by Topic									
Topic	**VAM-TR-96 VE**	**ARMA VE**	**AR VE**	**MA VE**	**ARIMA VE**	**PB VE**	**Best Baseline VE**	**Best Baseline Name**	**PIFBB (%)**
other/censorship_outage	**196.26**	1125.87	1714.58	664.93	2009.06	359.4	359.4	PB	**45.39**
international/aid_rejected	**643.88**	2059.78	2893.7	1249.18	3345.97	1106.68	1106.68	PB	**41.82**
maduro/legitimate	**70.94**	100.13	97.62	95.26	107.38	108.6	95.26	MA	**25.53**
other/chavez	**83.82**	121.73	125.09	112.25	150.29	119.72	112.25	MA	**25.33**
maduro/cuba_support	**80.97**	116.91	108.36	113.78	129.91	133.67	108.36	AR	**25.28**
protests	**370.8**	474.29	499.15	464.0	504.8	566.07	464.0	MA	**20.09**
arrests	**101.6**	127.86	127.44	134.04	133.27	126.81	126.81	PB	**19.88**
maduro/narco	**56.08**	66.08	70.48	70.39	67.88	81.79	66.08	ARMA	**15.13**
other/chavez/anti	**43.92**	63.01	66.61	63.43	69.71	50.92	50.92	PB	**13.74**
arrests/opposition	**85.57**	100.01	97.79	97.0	98.19	96.97	96.97	PB	**11.76**
other/anti_socialism	**25.69**	32.05	36.64	32.25	32.69	28.51	28.51	PB	**9.89**
international/respect_sovereignty	**231.09**	292.05	270.42	381.48	328.53	255.26	255.26	PB	**9.47**
military	**1177.11**	1594.55	1605.83	1215.86	1400.8	1414.2	1215.86	MA	**3.19**
military/desertions	323.45	338.19	344.77	323.03	417.22	453.57	**323.03**	MA	-0.13
international/aid	1228.38	1307.76	1218.71	1250.2	1401.07	1244.66	**1218.71**	AR	-0.79
guaido/legitimate	469.69	459.28	466.83	467.88	460.86	548.33	**459.28**	ARMA	-2.27
maduro/dictator	244.29	252.34	246.74	234.41	258.68	245.79	**234.41**	MA	-4.21
violence	1021.98	941.73	901.25	1038.6	945.37	1246.48	**901.25**	AR	-13.4

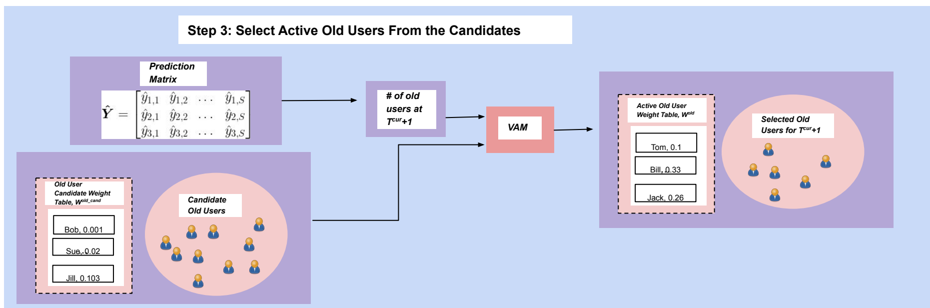
Table XII: VAM Twitter VE Baseline Comparisons by Topic

User-Assignment Module Symbols	
Symbol	***Meaning***
G	A temporal graph. It is a series of static graphs.
u	a child user in edge $(u, v, w(u, v, t))$
v	a parent user in edge $(u, v, w(u, v, t))$
$w(u, v, t)$	The number of times child user u interacted with parent v at time t.
T	The prediction time step of interest
S	The length of the output time series we wish to predict
$\hat{\boldsymbol{Y}}$	The time series matrix consisting of the (1) number of actions, (2) number of new users, and (3) number of old users from time $T + 1$ to $T + S$
$\hat{G}^{future}$	The temporal graph sequence that the user-assignment module must predict
L^{user}	The user-assignment lookback factor.
L^{vol}	The volume prediction lookback factor.
s	Any given time step between 1 and S, inclusive.
$T + s$	The current timestep of interest within the scope of the user assignment algorithm. $T + s = T + s - 1$.
p^{old_act}	The probability that an old user will be active in a given time step.
W^{old_cand}	The probability weight table for old user candidates.
W^{old}	The probability weight table for selected users from the W^{old_cand} table.
$\hat{O}_s$	The set of predicted active old users in time step $T + s$
$\hat{N}_s$	The set of newly generated active users in time step $T + s$
W^{new_arch}	The new user archetype weight table. Used to select the mostly likely archetypes a newly generated user will behave like.
u^{arch}	A user archetype.
p^{act_arch}	How likely a user of a particular archetype will be active in $T + s$
p^{infl_arch}	How likely a user of a particular archetype will be influential in $T + s$
u^{new}	A newly generated user.

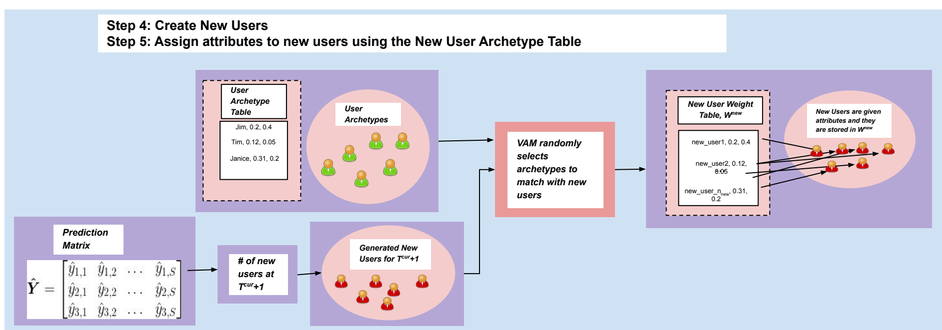
Table XIII: A table of the various symbols used in the user-assignment section of this work.



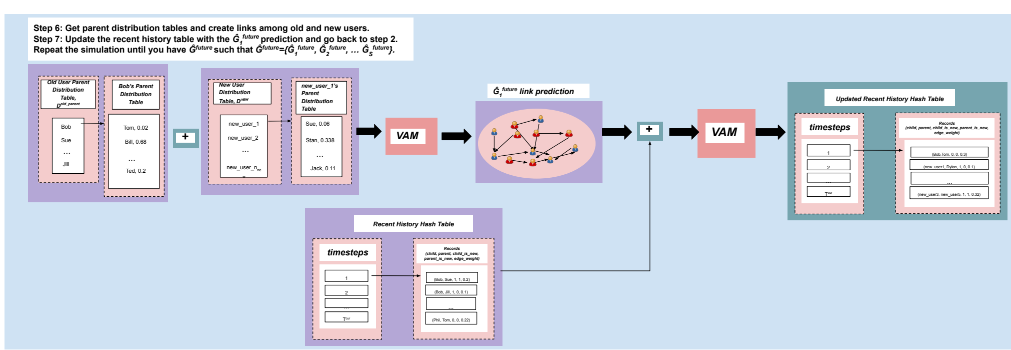
(a) Step 1: Use the recent network history, G^{recent} to create the recent history hash table, H^{recent}. Step 2: Use the recent history table to create the *Old User Candidate Weight Table*, W^{old_cand}.



(b) Step 3: Use the old user candidate weight table and the $\hat{Y}$ volume matrix to select the active old users from the candidates.

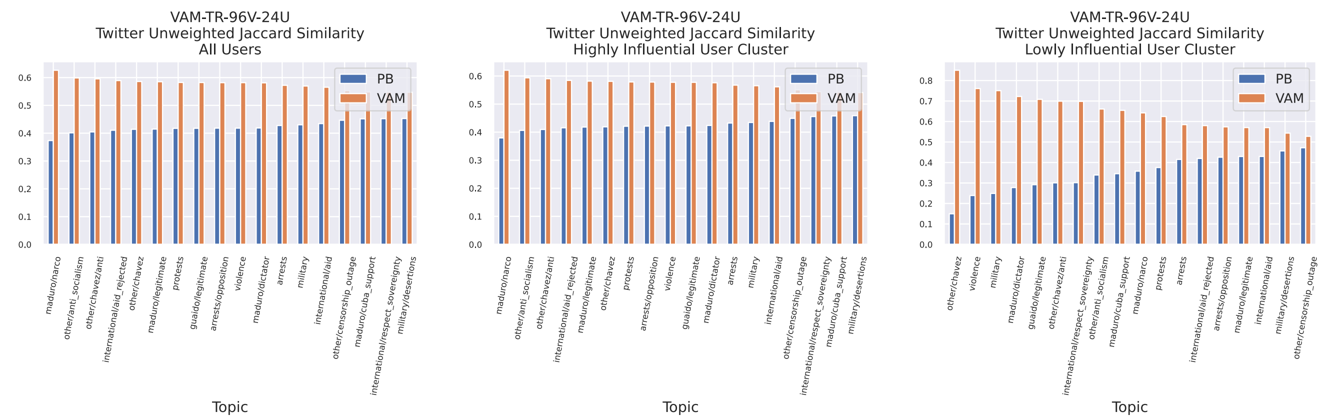


(c) Step 4: Use the $\hat{Y}$ volume matrix to generate new users. Step 5: Create the new user archetype table to assign attributes to the generated new users.



(d) Step 6: Create the new and old user parent distribution tables. Use these tables to assign edges among the old and new users using probabilities. This new set of links makes up the graph, $\hat{G}_1^{future}$. Step 7: Use $\hat{G}_1^{future}$ to update the recent history table, H^{recent}. Go back to step 2 and continue the simulation algorithm until *VAM* has created $\hat{G}_1^{future}$ such that $\hat{G}^{future} = \hat{G}_1^{future}, \hat{G}_2^{future}, ...\hat{G}_S^{future}$.

Figure 4: Steps 1-7 of the VAM User-Assignment Algorithm.



(a) All Users - Unweighted JS (Higher is Better) (b) Highly Influential Cluster - Unweighted JS (Higher is Better) (c) Lowly Influential Cluster - Unweighted JS (Higher is Better)

Figure 5: These barplots show weighted and unweighted Jaccard similarity results for the full user sets, highly influential clusters, and lowly influential clusters.

Twitter Unweighted Jaccard Similarity VAM-TR-96V-24U Old Users Full (Unweighted)			
Topic	PB	VAM-TR-96V-24U	PIFB
maduro/narco	0.0914	**0.1532**	**67.6657**
other/anti_socialism	0.0724	**0.108**	**49.1753**
other/chavez/anti	0.088	**0.1295**	**47.1248**
international/aid_rejected	0.1177	**0.1689**	**43.5021**
other/chavez	0.0896	**0.1268**	**41.537**
maduro/legitimate	0.095	**0.134**	**40.9659**
protests	0.1002	**0.1398**	**39.6176**
guaido/legitimate	0.1137	**0.1585**	**39.3256**
arrests/opposition	0.1348	**0.1875**	**39.1254**
violence	0.1265	**0.176**	**39.0475**
maduro/dictator	0.1084	**0.1504**	**38.7333**
arrests	0.1397	**0.1869**	**33.7607**
military	0.1325	**0.1754**	**32.3579**
international/aid	0.1301	**0.1692**	**30.0468**
other/censorship_outage	0.1259	**0.156**	**23.8988**
maduro/cuba_support	0.1026	**0.1245**	**21.2528**
international/respect_sovereignty	0.137	**0.1659**	**21.0456**
military/desertions	0.1554	**0.1878**	**20.8761**

Table XIV: Twitter VAM-TR-96V-24U Old Users - unweighted

Twitter Unweighted JS VAM-TR-96V-24U Old Users Lowly Infl.			
Topic	PB	VAM-TR-96V-24U	PIFB
other/chavez	0.0043	**0.0244**	**466.3653**
violence	0.0141	**0.0449**	**217.9803**
military	0.0144	**0.0434**	**201.2047**
maduro/dictator	0.0165	**0.0429**	**160.1118**
guaido/legitimate	0.0166	**0.0403**	**142.4018**
other/chavez/anti	0.0575	**0.1337**	**132.4138**
international/respect_sovereignty	0.0162	**0.0375**	**131.8387**
other/anti_socialism	0.1329	**0.2591**	**94.9254**
maduro/cuba_support	0.0915	**0.1734**	**89.5422**
maduro/narco	0.1567	**0.2813**	**79.4937**
protests	0.0478	**0.0795**	**66.4434**
arrests	0.139	**0.1959**	**40.9633**
international/aid_rejected	0.1836	**0.2537**	**38.1899**
arrests/opposition	0.3115	**0.4199**	**34.7953**
maduro/legitimate	0.2144	**0.2849**	**32.9002**
international/aid	0.0229	**0.0304**	**32.6733**
military/desertions	0.5281	**0.6286**	**19.0338**
other/censorship_outage	0.7282	**0.8143**	**11.8256**

Table XVI: Twitter VAM-TR-96V-24U Old Users Lowly Influential Cluster (Unweighted)

Twitter Unweighted JS VAM-TR-96V-24U Old Users Highly Infl.			
Topic	PB	VAM-TR-96V-24U	PIFB
maduro/narco	0.0986	**0.1612**	**63.4799**
other/anti_socialism	0.0777	**0.1135**	**45.9964**
other/chavez/anti	0.094	**0.1355**	**44.1014**
international/aid_rejected	0.1256	**0.1766**	**40.5989**
maduro/legitimate	0.0986	**0.1372**	**39.2171**
other/chavez	0.0962	**0.1334**	**38.6994**
protests	0.1063	**0.1461**	**37.4643**
arrests/opposition	0.1417	**0.1944**	**37.1511**
violence	0.1345	**0.1839**	**36.7106**
guaido/legitimate	0.1225	**0.1674**	**36.6382**
maduro/dictator	0.1177	**0.1599**	**35.8666**
arrests	0.1489	**0.1954**	**31.2357**
military	0.139	**0.1809**	**30.0771**
international/aid	0.1375	**0.1761**	**28.0816**
other/censorship_outage	0.13	**0.1594**	**22.6478**
international/respect_sovereignty	0.1428	**0.1706**	**19.4916**
maduro/cuba_support	0.1116	**0.1322**	**18.4438**
military/desertions	0.1646	**0.1942**	**17.9705**

Table XV: Twitter VAM-TR-96V-24U Old Users Highly Influential Cluster (Unweighted)

Twitter Weighted JS VAM-TR-96V-24U Full User Set			
Topic	PB	VAM-TR-96V-24U	PIFB
maduro/cuba_support	0.0465	**0.1023**	**120.1467**
international/aid_rejected	0.0329	**0.0716**	**117.7645**
other/anti_socialism	0.0329	**0.0683**	**107.6227**
maduro/narco	0.0502	**0.1021**	**103.3784**
maduro/legitimate	0.0497	**0.0767**	**54.2822**
other/chavez/anti	0.044	**0.0674**	**53.0084**
violence	0.0582	**0.0815**	**40.0279**
other/censorship_outage	0.0724	**0.1007**	**38.9522**
other/chavez	0.0459	**0.0634**	**38.1265**
maduro/dictator	0.0474	**0.0641**	**35.0402**
military/desertions	0.0595	**0.0798**	**34.1568**
protests	0.0473	**0.0594**	**25.5639**
arrests/opposition	0.0668	**0.0837**	**25.3776**
international/respect_sovereignty	0.0814	**0.0994**	**22.1611**
international/aid	0.0867	**0.1006**	**15.9882**
arrests	0.0742	**0.0809**	**9.0074**
military	0.0802	**0.0814**	**1.5735**
guaido/legitimate	0.0654	**0.0655**	**0.1112**

Table XVII: VAM-TR-96V-24U Old Users Weighted (Full)

Twitter Weighted JS VAM-TR-96V-24U Old Users Highly Infl. Cluster			
Topic	**PB**	**VAM-TR-96V-24U**	**PIFB**
maduro/cuba_support	0.0478	**0.1041**	**117.723**
international/aid_rejected	0.0332	**0.0719**	**116.413**
other/anti_socialism	0.0336	**0.0689**	**105.3473**
maduro/narco	0.0508	**0.1026**	**101.7409**
maduro/legitimate	0.0502	**0.0771**	**53.5985**
other/chavez/anti	0.0448	**0.0679**	**51.5688**
violence	0.059	**0.082**	**39.0396**
other/censorship_outage	0.0734	**0.1011**	**37.8058**
other/chavez	0.0468	**0.0639**	**36.6019**
maduro/dictator	0.048	**0.0645**	**34.232**
military/desertions	0.0618	**0.0801**	**29.5377**
protests	0.048	**0.0598**	**24.7245**
arrests/opposition	0.0675	**0.0842**	**24.603**
international/respect_sovereignty	0.0821	**0.0998**	**21.6542**
international/aid	0.0872	**0.1009**	**15.8175**
arrests	0.0752	**0.0814**	**8.2824**
military	0.081	**0.0818**	**1.0194**
guaido/legitimate	**0.0662**	0.0658	-0.6144

Table XVIII: VAM-TR-96V-24U Old Users Weighted - High Cluster

Twitter Weighted JS VAM-TR-96V-24U Old Users Lowly Infl.			
Topic	**PB**	**VAM-TR-96V-24U**	**PIFB**
other/chavez	0.004	**0.0243**	**507.1694**
violence	0.0139	**0.0449**	**222.8647**
military	0.0139	**0.0433**	**211.1303**
maduro/dictator	0.0159	**0.0429**	**169.3563**
guaido/legitimate	0.016	**0.0402**	**150.9818**
international/respect_sovereignty	0.0159	**0.0374**	**135.193**
other/chavez/anti	0.0575	**0.1337**	**132.4138**
other/anti_socialism	0.1329	**0.2591**	**94.9254**
maduro/cuba_support	0.0913	**0.1734**	**89.9963**
maduro/narco	0.1567	**0.2813**	**79.4937**
protests	0.0478	**0.0795**	**66.3686**
arrests	0.1389	**0.1957**	**40.8761**
international/aid_rejected	0.1829	**0.2537**	**38.7267**
international/aid	0.022	**0.0302**	**36.8813**
arrests/opposition	0.3115	**0.4198**	**34.7777**
maduro/legitimate	0.2143	**0.2848**	**32.914**
military/desertions	0.5278	**0.6286**	**19.0961**
other/censorship_outage	0.7282	**0.8143**	**11.8256**

Table XIX: Twitter VAM-TR-96V-24U Old Users Lowly Influential Cluster (Weighted)

Twitter VAM-TR-96V-24U Earth Mover's Distance Full User Set Results			
Topic	**PB**	**VAM-TR-96V-24U**	**PIFB (%)**
maduro/narco	0.0497	**0.034**	**31.58**
military	0.0027	**0.002**	**25.94**
military/desertions	0.1817	**0.1365**	**24.85**
arrests/opposition	0.0609	**0.0476**	**21.8**
international/aid_rejected	0.035	**0.0276**	**21.1**
arrests	0.023	**0.0184**	**19.97**
violence	0.006	**0.0049**	**18.77**
maduro/legitimate	0.0198	**0.0161**	**18.65**
other/chavez	0.0066	**0.0054**	**18.18**
other/anti_socialism	0.0365	**0.0299**	**18.09**
other/chavez/anti	0.0173	**0.0144**	**17.16**
protests	0.0121	**0.0101**	**17.08**
international/aid	0.0028	**0.0024**	**15.98**
maduro/dictator	0.0059	**0.0051**	**12.8**
guaido/legitimate	0.0047	**0.0041**	**12.21**
other/censorship_outage	0.269	**0.241**	**10.43**
international/respect_sovereignty	0.0022	**0.0021**	**2.94**
maduro/cuba_support	**0.048**	0.0483	-0.59

Table XX: VAM Earth Mover's Distance Full User Set Results

Twitter VAM-TR-96V-24U Earth Mover's Distance Highly Influential Cluster			
Topic	**PB**	**VAM**	**PIFB (%)**
maduro/narco	0.0511	**0.0361**	**29.29**
military/desertions	0.1842	**0.1324**	**28.08**
military	0.003	**0.0023**	**23.73**
arrests/opposition	0.0617	**0.0494**	**19.94**
international/aid_rejected	0.0357	**0.0289**	**18.93**
arrests	0.0245	**0.0201**	**18.22**
maduro/legitimate	0.0204	**0.0167**	**18.07**
violence	0.0067	**0.0055**	**17.84**
other/chavez	0.0071	**0.0059**	**17.13**
other/chavez/anti	0.0182	**0.0152**	**16.48**
other/anti_socialism	0.0362	**0.0304**	**15.96**
protests	0.0128	**0.0108**	**15.48**
international/aid	0.0032	**0.0027**	**15.1**
maduro/dictator	0.0069	**0.006**	**12.29**
guaido/legitimate	0.0054	**0.0048**	**12.01**
other/censorship_outage	0.2687	**0.2368**	**11.87**
maduro/cuba_support	0.0494	**0.0483**	**2.18**
international/respect_sovereignty	0.0023	**0.0023**	**1.99**

Table XXI: Twitter VAM-TR-96V-24U Earth Mover's Distance Highly Influential Cluster

Twitter VAM-TR-96V-24U Earth Mover's Distance Lowly Influential Cluster			
Topic	**PB**	**VAM-TR-96V-24U**	**PIFB (%)**
military	0.0011	**0.0009**	**15.67**
international/aid	0.0013	**0.0012**	**9.76**
protests	0.0074	**0.0069**	**5.58**
guaido/legitimate	0.0021	**0.002**	**5.48**
international/respect_sovereignty	0.0014	**0.0014**	**4.46**
other/anti_socialism	0.0307	**0.0295**	**4.08**
maduro/narco	0.024	**0.0231**	**3.84**
violence	0.0028	**0.0028**	**2.62**
arrests/opposition	**0.0178**	0.0179	-0.5
other/chavez	**0.0046**	0.0046	-1.42
arrests	**0.0097**	0.01	-2.39
international/aid_rejected	**0.0178**	0.0182	-2.44
maduro/legitimate	**0.0161**	0.0171	-6.44
maduro/dictator	**0.0024**	0.0026	-8.02
military/desertions	**0.0394**	0.0439	-11.63
other/censorship_outage	**0.0616**	0.071	-15.25
other/chavez/anti	**0.0107**	0.0124	-15.31
maduro/cuba_support	**0.0197**	0.0248	-25.68

Table XXII: Twitter VAM-TR-96V-24U Earth Mover's Distance Lowly Influential Cluster

Twitter VAM-TR-96V-24U Relative Hausdorff Distance Full User Set Results			
Topic	**PB**	**VAM**	**PIFB (%)**
arrests	1.2153	**0.9883**	**18.68**
other/chavez	0.9996	**0.8205**	**17.92**
maduro/legitimate	0.9915	**0.8148**	**17.82**
protests	1.3615	**1.1192**	**17.8**
military/desertions	1.4612	**1.255**	**14.11**
arrests/opposition	1.1051	**0.9508**	**13.96**
guaido/legitimate	1.1751	**1.017**	**13.45**
maduro/narco	1.0597	**0.9269**	**12.53**
other/chavez/anti	0.9505	**0.8329**	**12.37**
other/censorship_outage	0.9662	**0.8483**	**12.19**
maduro/dictator	1.0244	**0.9213**	**10.06**
violence	1.3646	**1.2564**	**7.93**
other/anti_socialism	0.8227	**0.7764**	**5.63**
international/respect_sovereignty	0.9451	**0.8986**	**4.92**
maduro/cuba_support	0.8429	**0.8376**	**0.63**
international/aid_rejected	**1.1422**	1.2309	-7.77
military	**1.0861**	1.1989	-10.38
international/aid	**1.1475**	1.4215	-23.87

Table XXIII: Twitter VAM-TR-96V-24U Relative Hausdorff Distance Full User Set Results

Twitter VAM-TR-96V-24U Relative Hausdorff Distance Highly Influential Cluster			
Topic	**PB**	**VAM-TR-96V-24U**	**PIFB (%)**
arrests	1.2161	**0.9897**	**18.62**
protests	1.3625	**1.121**	**17.72**
other/chavez	1.0004	**0.8238**	**17.65**
maduro/legitimate	0.9922	**0.8172**	**17.63**
military/desertions	1.4662	**1.2556**	**14.36**
arrests/opposition	1.1065	**0.9536**	**13.82**
guaido/legitimate	1.1751	**1.0183**	**13.34**
other/censorship_outage	0.9713	**0.8424**	**13.27**
maduro/narco	1.0614	**0.9326**	**12.13**
other/chavez/anti	0.9499	**0.8358**	**12.01**
maduro/dictator	1.0246	**0.9233**	**9.89**
violence	1.3645	**1.257**	**7.88**
other/anti_socialism	0.8252	**0.7831**	**5.11**
international/respect_sovereignty	0.9453	**0.8998**	**4.81**
maduro/cuba_support	0.8411	**0.8389**	**0.27**
international/aid_rejected	**1.143**	1.2347	-8.02
military	**1.0863**	1.2	-10.47
international/aid	**1.1475**	1.4227	-23.98

Table XXIV: Twitter VAM-TR-96V-24U Relative Hausdorff Distance Highly Influential Cluster

Twitter VAM-TR-96V-24U Relative Hausdorff Distance Lowly Influential Cluster			
Topic	**PB**	**VAM-TR-96V-24U**	**PIFB (%)**
other/chavez	0.5363	**0.4406**	**17.85**
guaido/legitimate	0.5398	**0.47**	**12.93**
other/chavez/anti	0.453	**0.3957**	**12.63**
maduro/dictator	0.4676	**0.4135**	**11.57**
other/anti_socialism	0.3573	**0.3299**	**7.67**
protests	0.4845	**0.454**	**6.31**
other/censorship_outage	0.1012	**0.0967**	**4.37**
military/desertions	0.1978	**0.1929**	**2.44**
maduro/legitimate	0.3073	**0.301**	**2.04**
violence	0.4814	**0.4742**	**1.51**
arrests	0.3678	**0.3632**	**1.23**
maduro/cuba_support	**0.3701**	0.3784	-2.24
international/aid_rejected	**0.3535**	0.3678	-4.06
international/respect_sovereignty	**0.4867**	0.5111	-5.0
maduro/narco	**0.2137**	0.2306	-7.88
international/aid	**0.5191**	0.5628	-8.42
military	**0.4737**	0.5224	-10.29
arrests/opposition	**0.2264**	0.2519	-11.25

Table XXV: Twitter VAM-TR-96V-24U Relative Hausdorff Distance Lowly Influential Cluster