CONVERSATION DYNAMICS IN E-COMMERCE SPACE

PRESENTED BY: UNNIKRISHNAN T A (14IT60R07) | MENTOR: PROF. NILOY GANGULY

COLLABORATION PROJECT WITH FLIPKART DATASCIENCE

Contents

A brief overview

User Sentiments & Classification

Network Analysis

Conversation Graph

Askew Bowtie

Social Relationship Management

Popularity Dynamics

Motivation

- E- Commerce Promotional Events
- Conversations in Twitter
- Customer Queries and Sentiments
- Popularity of Conversations

Dataset Description

EVENT	DURATION	# TWEETS	# USERS
Big Billion Day (BBD)	Oct 6, 2014	135,593	40, 891
Big App Shopping Day 2 (BASD2)	Dec 8 – 12, 2014	51, 323	10, 926
Big App Shopping Day 3 (BASD3)	June 22 – 24, 2015	160, 322	19, 155
Amazon Prime Day (APD)	July 15, 2015	67, 212	39, 779

Profile Enrichment

- **Basic profile info**
- **Up to 200 recent tweets**

Sentiment Classification

Hashtags
hashtags

Smileys
smileys

Punctuation marks
punctuation marks

Positive and Negative words
Count from Sentiment 140 lexicon and NRC emoticon lexicon sets

Elongated words
elongated words (“goooood”)

Negated Context
negated context.

Words in Capital
words completely typed in caps

Sentiment Lexicon Based
Features based on NRC Emotion Lexicon, MPQA, Bing Liu Lexicon

Classifier

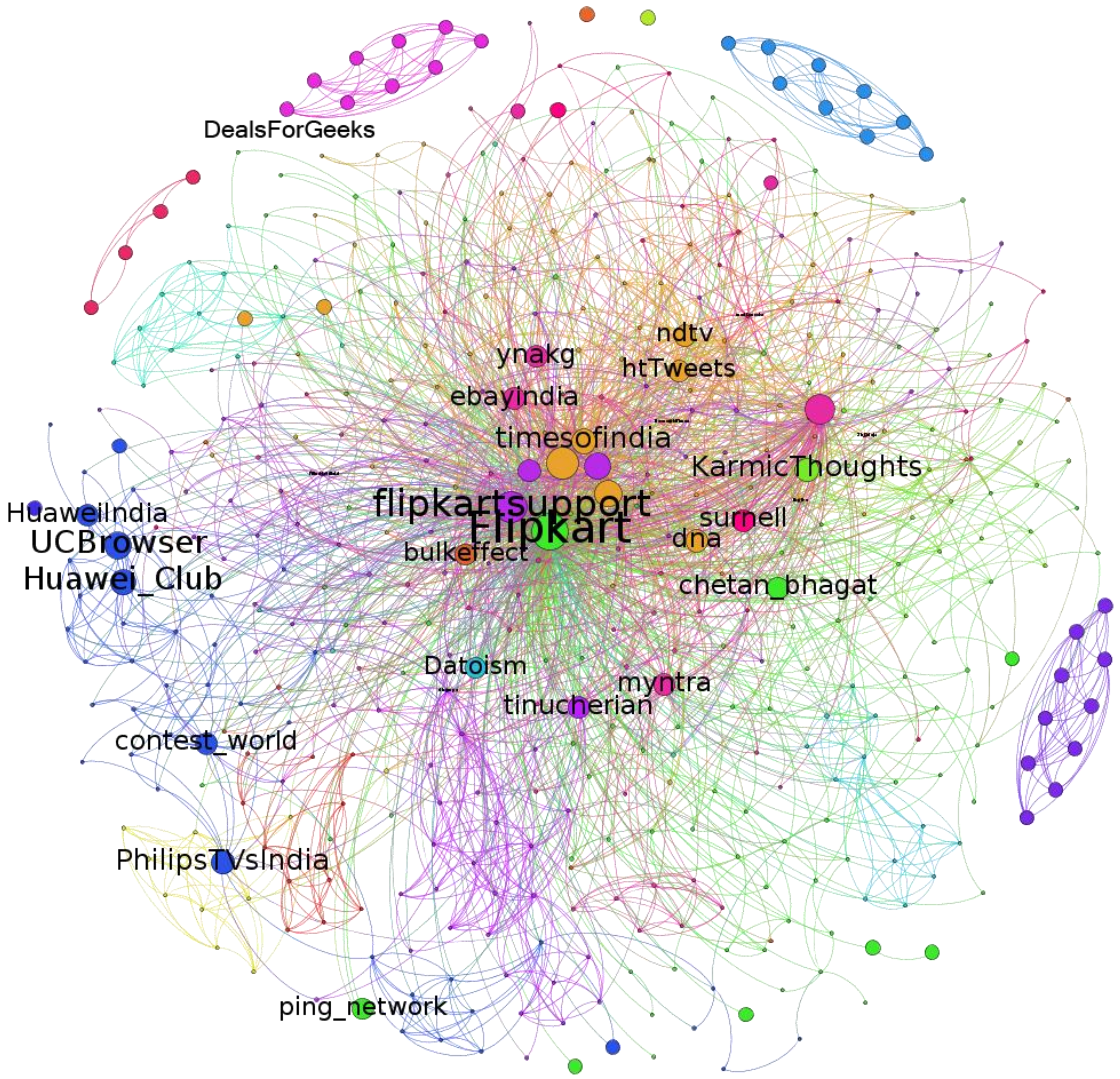
We used SVM classifier with 1000 manually annotated tweets for positive and negative sentiment

Performance

We obtain average precision and recall of 90% across the dataset

Co-Mention Graph

Analysis of co-mention graph from BBD



Observations

Communities

The graph shows the existence of a set of communities

Promotion Campaigns

Some of these communities are promotion al campaigns.

Positive Sentiments

The communities formed by promotional campaigns help spead positive sentiments in the network

User Classification

Classifying users into four categories

Satisfied

Users with at lest 60% positive tweets

Unsatisfied

Users with less than 60% positive tweets

Company accounts

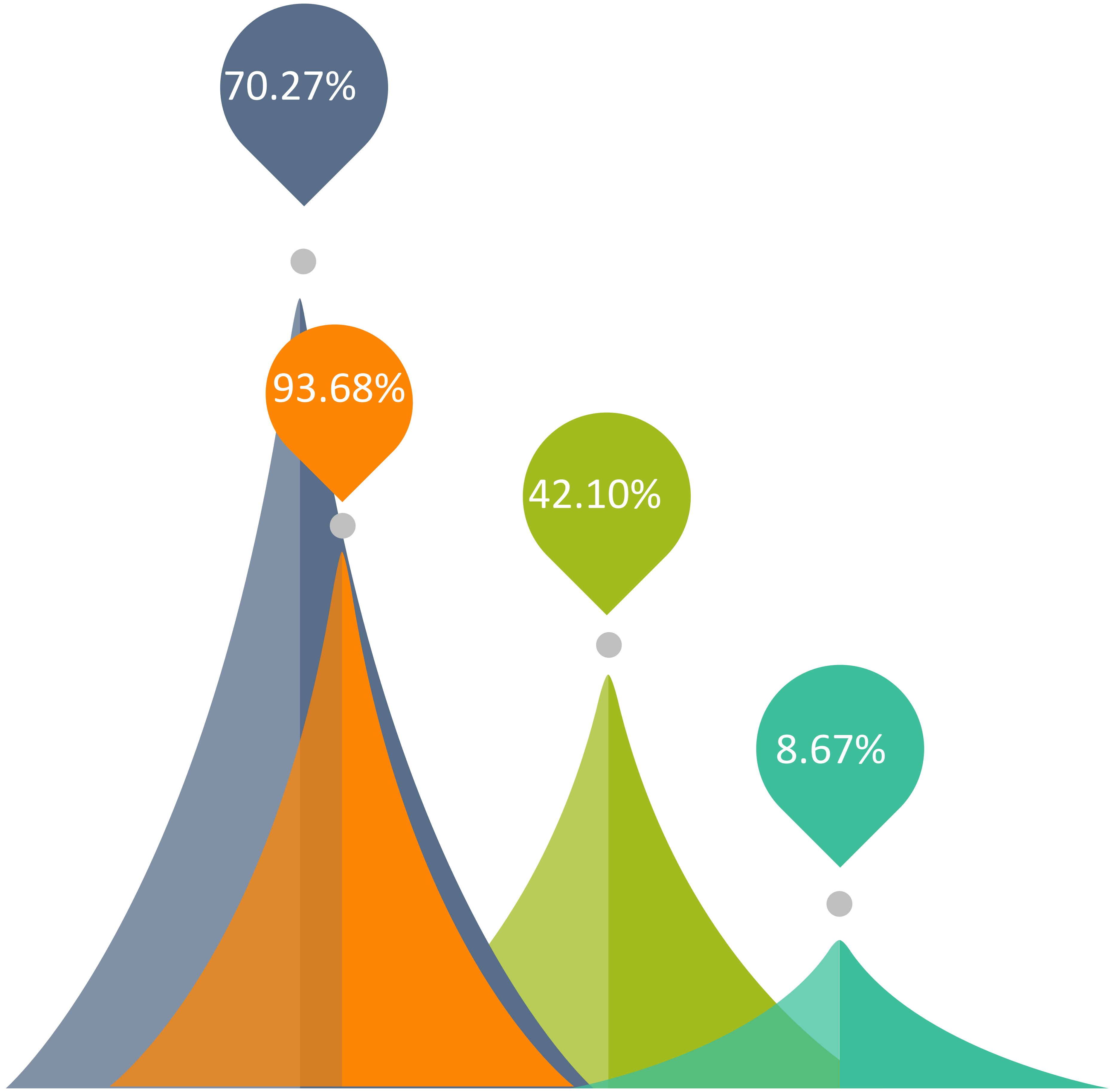
Accounts running promotional campaigns

Adversarial accounts

Accounts targeted at spreading negativity

Conversation Graph: Motivation

Motivation behind analyzing the conversation graph



- Tweets with mentions
- Negative sentiment tweets with mention
- Tweets with @flipkart as mention
- Tweets with more than one mention

Conversation Graph: Construction

A formal definition for conversation graph

We define conversation graph as a directed graph where the nodes are users and two nodes u, v are connected by an edge if u mentions b in a tweet. This captures

1. Mentions

Situations where *a* mentions *b*

2. Reply

If *a* replies to *b*, a mention from *a* to *b* is added by twitter

3. Retweet

When a tweet is retweeted, mention edge is duplicated

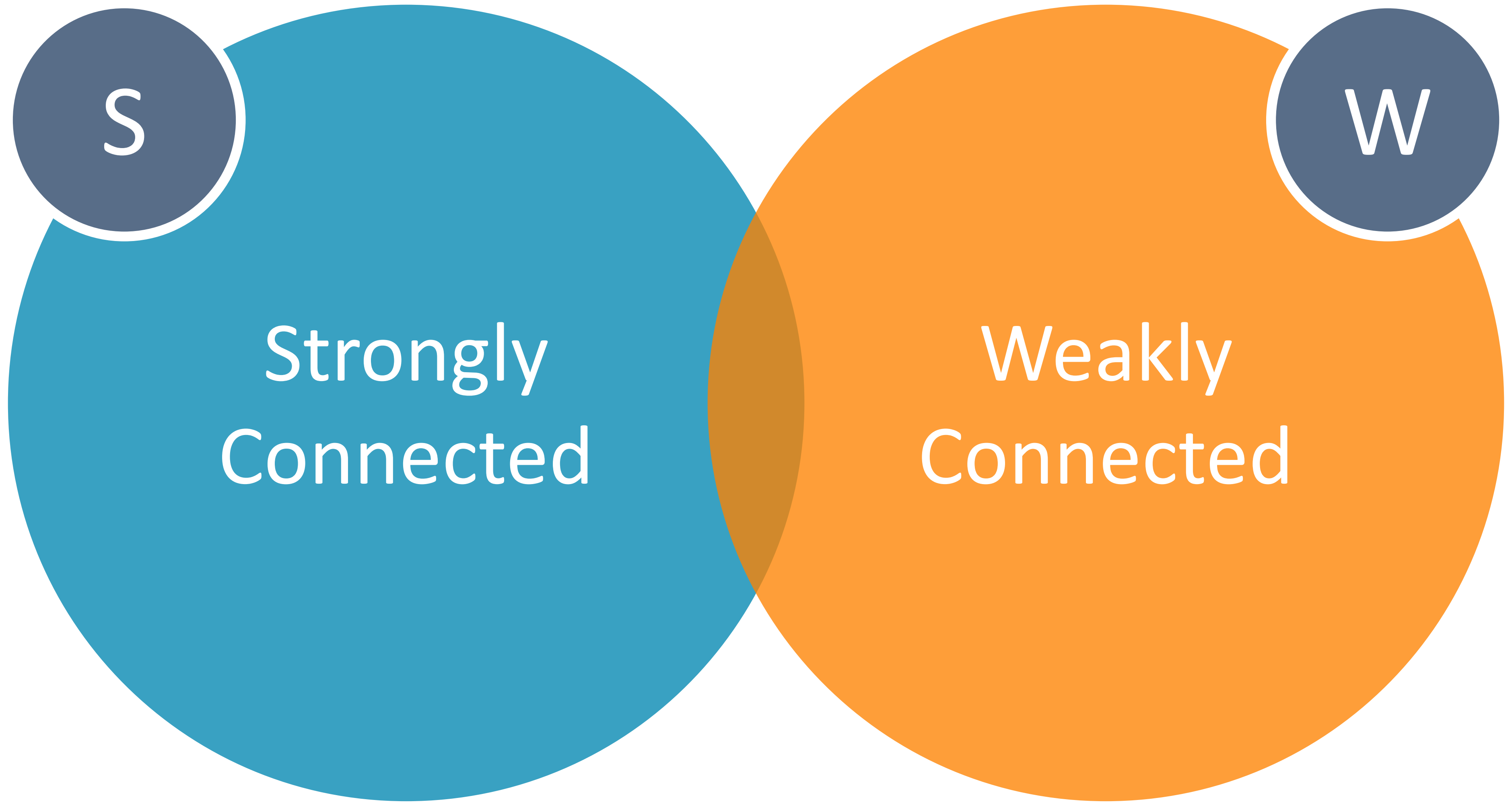
4. Time Slice

Graph is often constructed for time slice between t_i and t_{i+1}

Conversation Graph: Components

A component-wise analysis of conversation graph

Nodes in Largest Strongly Connected component(LSCC)
4418



Nodes in Largest Weakly Connected Component (LWCC)
34331

We treat LSCC and LWCC to be mutually exclusive, although by definition LSCC ⊆ LWCC

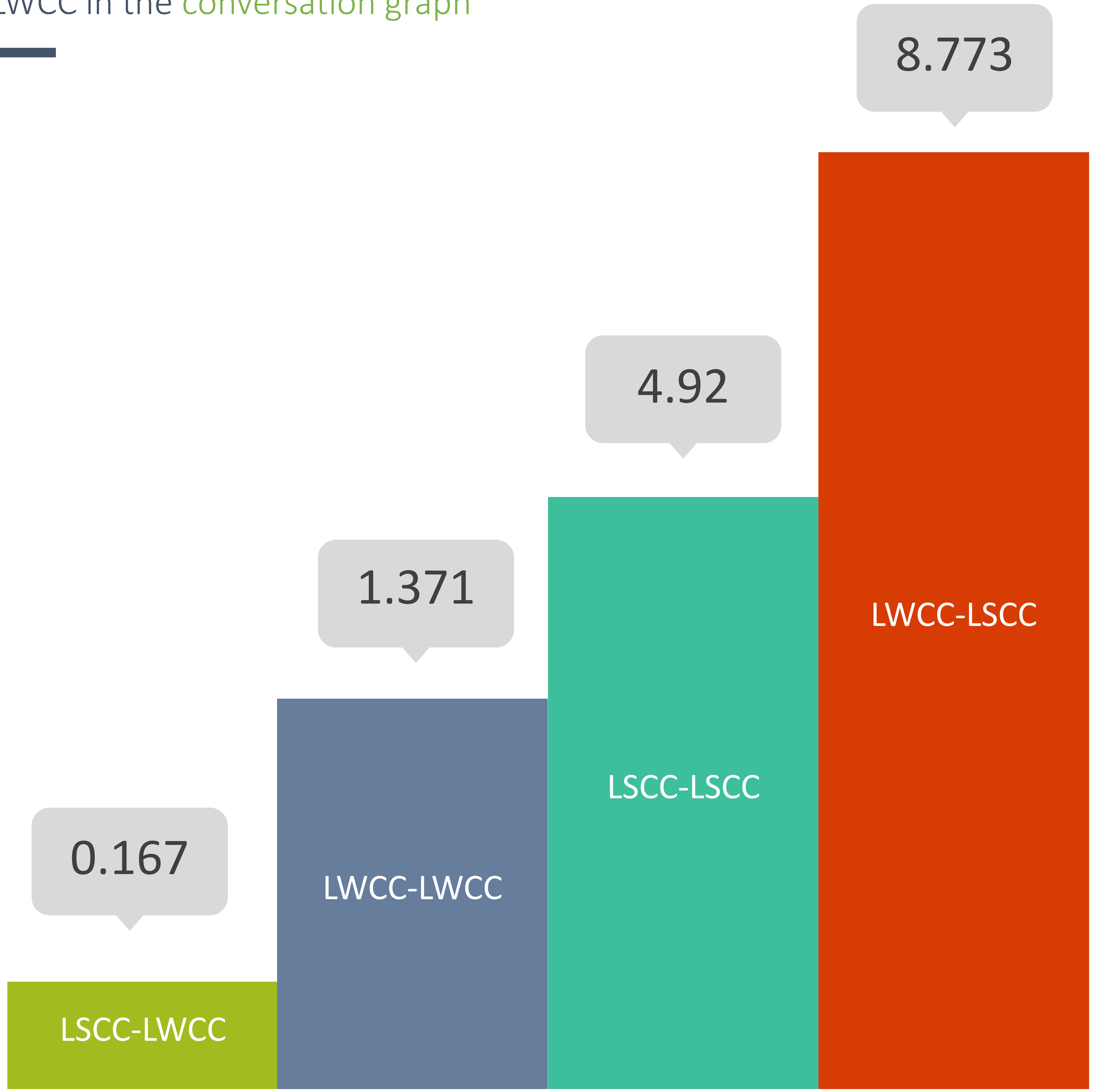
Mean Mentions Among Components

Mean mentions among LSCC and LWCC in the conversation graph

Dominance of LSCC

Our initial analysis reveal that LSCC drives the conversation volume.

Mean mention from A-B refers to the mean number of mention (by dividing with number of nodes in B) that B is getting from A.



The Askew Bowtie

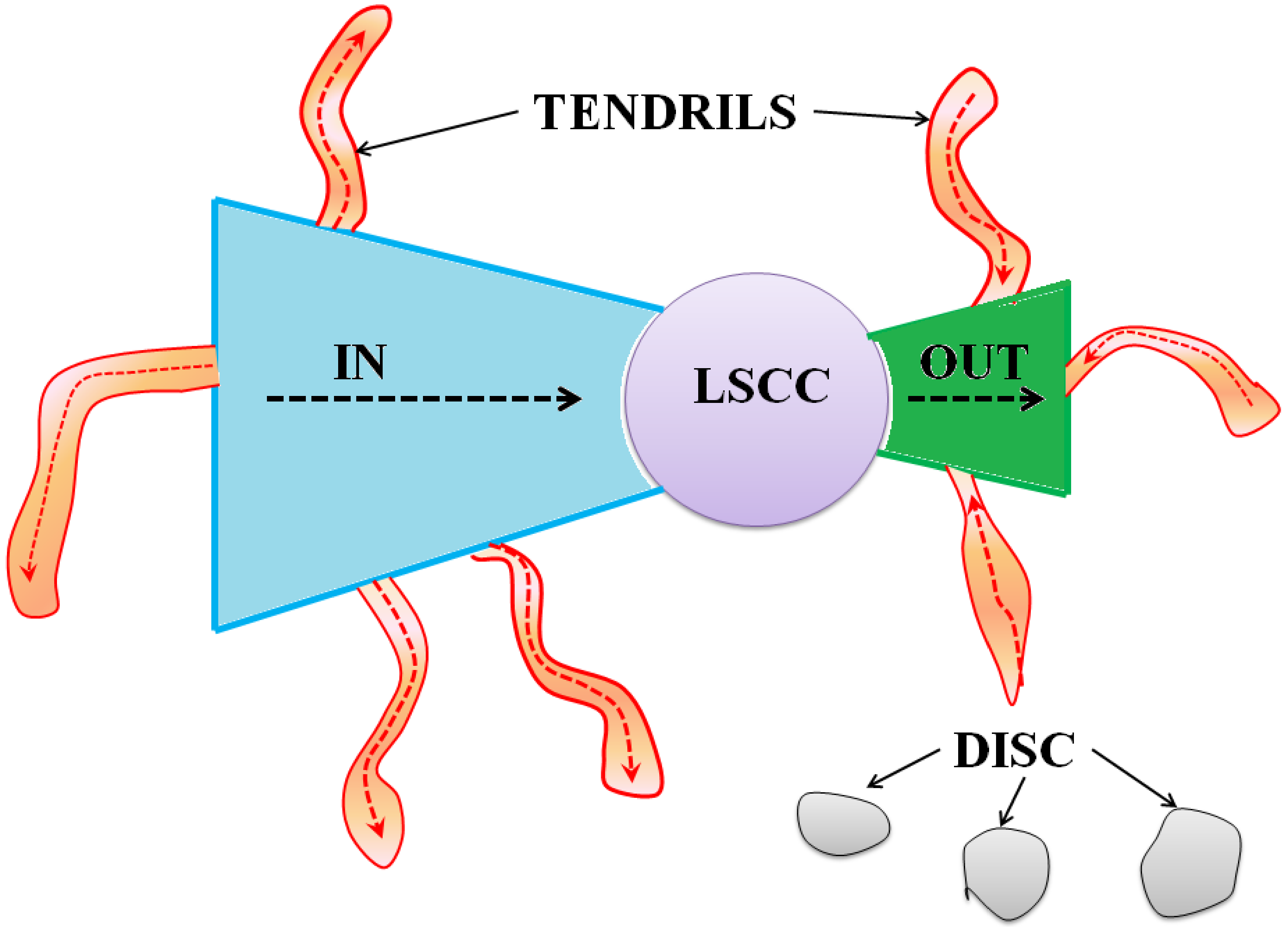
Evolution of Askew Bowtie Structure in Conversation Graph

LSCC
Follows the definition as before

IN
Set of nodes with a directed edge into LSCC, not vice versa

OUT
Set of nodes with a directed edge from LSCC, not vice versa

TENDRILS
Nodes with no direct connectivity to or from LSCC



DISCONNECTED
Nodes disconnected from the rest of the graph

Askew Bowtie: Components

Component size of Askew Bowtie Structure across our datasets

DATASET	IN	LSCC	OUT	TENDRILS	DISC
Big Billion Day	63%	12%	1.5%	21%	2.8%
Big App Shopping Day 2	55.3%	23.1%	3.5%	13.8%	6.7%
Big App Shopping Day 3	57.1%	21.1%	1.2%	14.3%	6.2%
Amazon Prime Day	0.45%	0.09%	0.08%	67%	32.4%

Skewedness

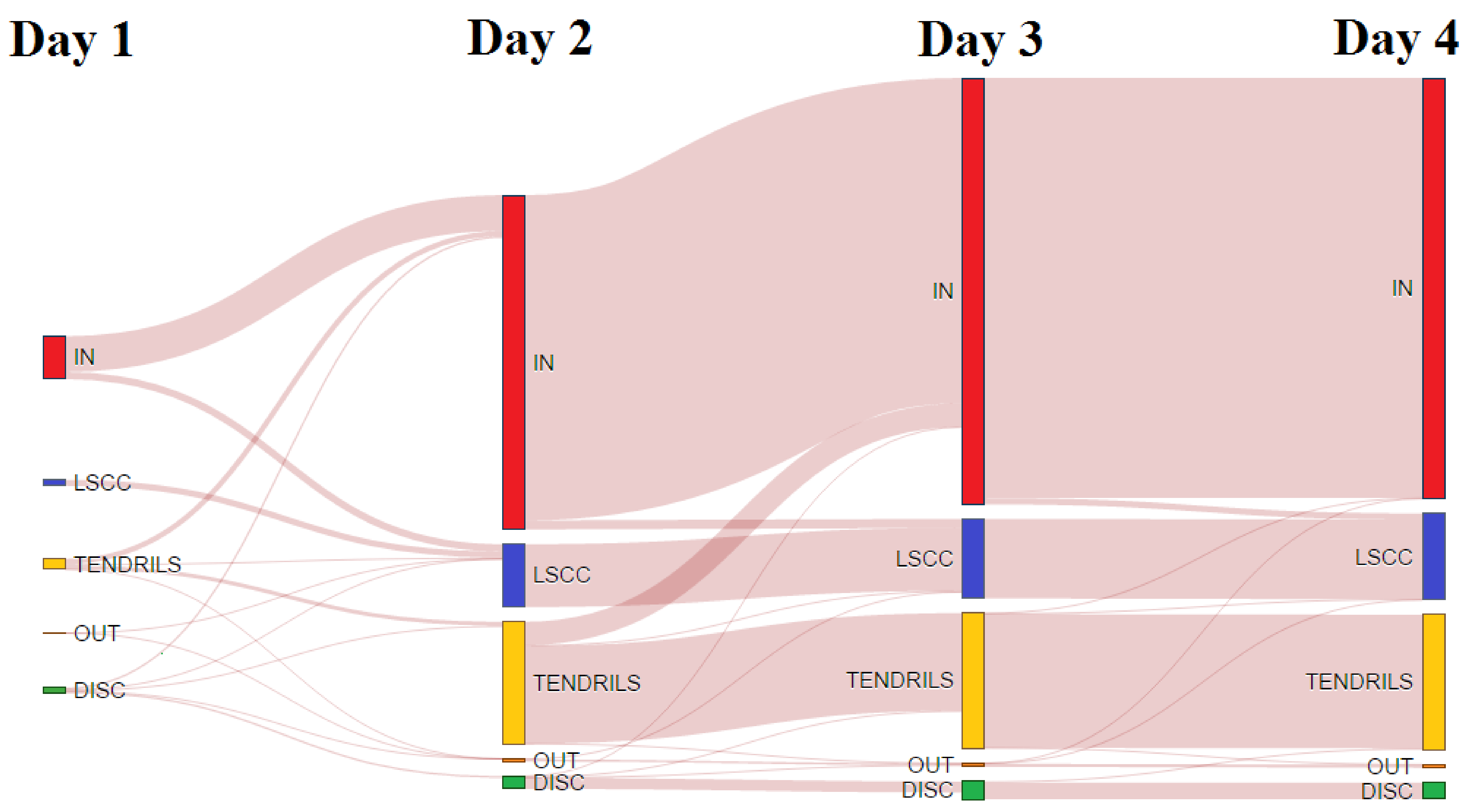
Ratio of the components are different from that of the web graph. Bowtie is skewed towards IN.

Consistency across datasets

Relative component size of the Askew Bowtie structure remains consistent across our dataset.

Askew Bowtie: Temporal Dynamics

User movements across different components of Askew Bowtie



Temporal Stability

As the event progresses, the user movements reduces and the over all structure tends to become stable.

Stability of Components

TENDRILS tend to be most unstable with an average of 35% users shifting to other components. LSCC is the most stable component across different days.

User accounts

Users belonging to adversarial category tend to move the most

Social Relationship Management

Social Relationship Management: Overview

Tweet Volume

Large amount tweets arrive in bursts.

Sentiment

Tweets carrying negative sentiment affects reputation

Popularity

As negative tweets get viral, more damage is caused.

Ranking

A ranking function that can prioritize the tweets based on expected damage.

SRM: Modelling

Features and cost function for modelling SRM

- If a user u mentions the company handle and post a negative tweet for which the waiting time is t_w , the damage score for the tweet is defined as the weighted sum (by the retweet count) of the sentiment score of all the tweets by u during this time period.
- The damage score caused by the SRM is the sum of damage scores of all the incoming tweets

Features used

- **Klout score of the user**
- **Fraction of followers in the component**
- **Average sentiment of user till reply**
- **Askew Bowtie component the user belongs to**
- **Core index of the user in the component**
- **Sentiment score of the tweet**
- **Waiting time for reply**

SRM Simulation: Performance Analysis

Feature analysis and performance of our SRM modelling

Rank	Feature	ΔMSE
1	Klout-score of user	3.69
2	Fraction of followers of the user inside the component	2.67
3	Core index of user inside the component	2.40
4	Waiting time for reply	2.14
5	Component of user inside Askew Bowtie	1.97

Results of Support Vector Regression

Performance

51% relative improvement in cost for BASD2, and APD.

83% increase in BASD3

Only **8%** in BBD which we believe is mainly due to the intrinsic negativity around it.

Importance of Bowtie Features

Bowtie based features prove to be important in assigning a priority score for the incoming tweets

Popularity Dynamics

Motivation for investigating popularity dynamics

Reinforced Poisson Process

We use RPP to predict the popularity dynamics

Point Processes

A point process that takes the arrival rate of items into consideration is better suited.

Retweeting Dynamics

Expected number of retweets for each individual tweets

Bowtie Based Ranking

Result of SVM regression.

Hashtag Dynamics

For a given hashtag, predicting the expected number of tweets containing it

Reinforced Poisson Process (RPP)

A brief introduction to Reinforced Poisson Process

A generative probabilistic framework to explicitly model the process through which individual items gain their popularity. The model simultaneously captures three parameters governing the popularity of an item.

Fitness of an Item

The intrinsic attractiveness by whith the item becomes popular

Reinforcement Mechanism

Captures the “rich gets richer” phenomena

Decay Function

Corresponds to temporal aging

RPP: Model Formulation

Modelling popularity dynamics using RPP

- The popularity dynamics of individual item d during time period [0, T] is characterized by a set of time moments $\{t_i^d\}$ $(1 \leq i \leq n_d)$

- For an individual item d, we model its popularity dynamics as a reinforced Poisson process characterized by the rate function $x_d(t)$ as

$$x_d(t) = \lambda_d f_d(t; \theta_d) i_d(t)$$

- Assume that all items are created equal and hence the effective number of attentions for all items has the same value, denoted by m. Then

$$i_d(t) = m + i - 1$$

RPP: Model Formulation Contd…

Modelling popularity dynamics using RPP

- Given that the $(i{-}1)th$ attention arrives at t_{i-1}^{d}, the probability that the i^{th} attention arrives at t_{i}^{d} follows

$$p_1\left(t_i^d \middle| t_{i-1}^d\right) = \lambda_d f_d\left(t_i^d;\theta_d\right)(m+i-1) \times e^{-\int_{t_{i-1}^d}^{t_i^d} \lambda_d f_d\left(t_i^d;\theta_d\right)(m+i-1)dt}$$

- Probability that no attention arrives between $t_{n_d}^{d}$ and T is

$$p_0\left(T \middle| t_{n_d}^d\right) = e^{-\int_{t_{n_d}^d}^{T} \lambda_d f_d\left(t_i^d;\theta_d\right)(m+n_d)dt}$$

RPP: Model Formulation Contd…

Modelling popularity dynamics using RPP

- Likelihood of observing the popularity dynamics $\{t_i^d\}$ during time interval $[0, T]$ follows

$$\mathcal{L}(\lambda_d, \theta_d) = p_0(T|t_{n_d}^d) \prod_{i=1}^{n_d} p_1(t_i^d|t_{i-1}^d)$$

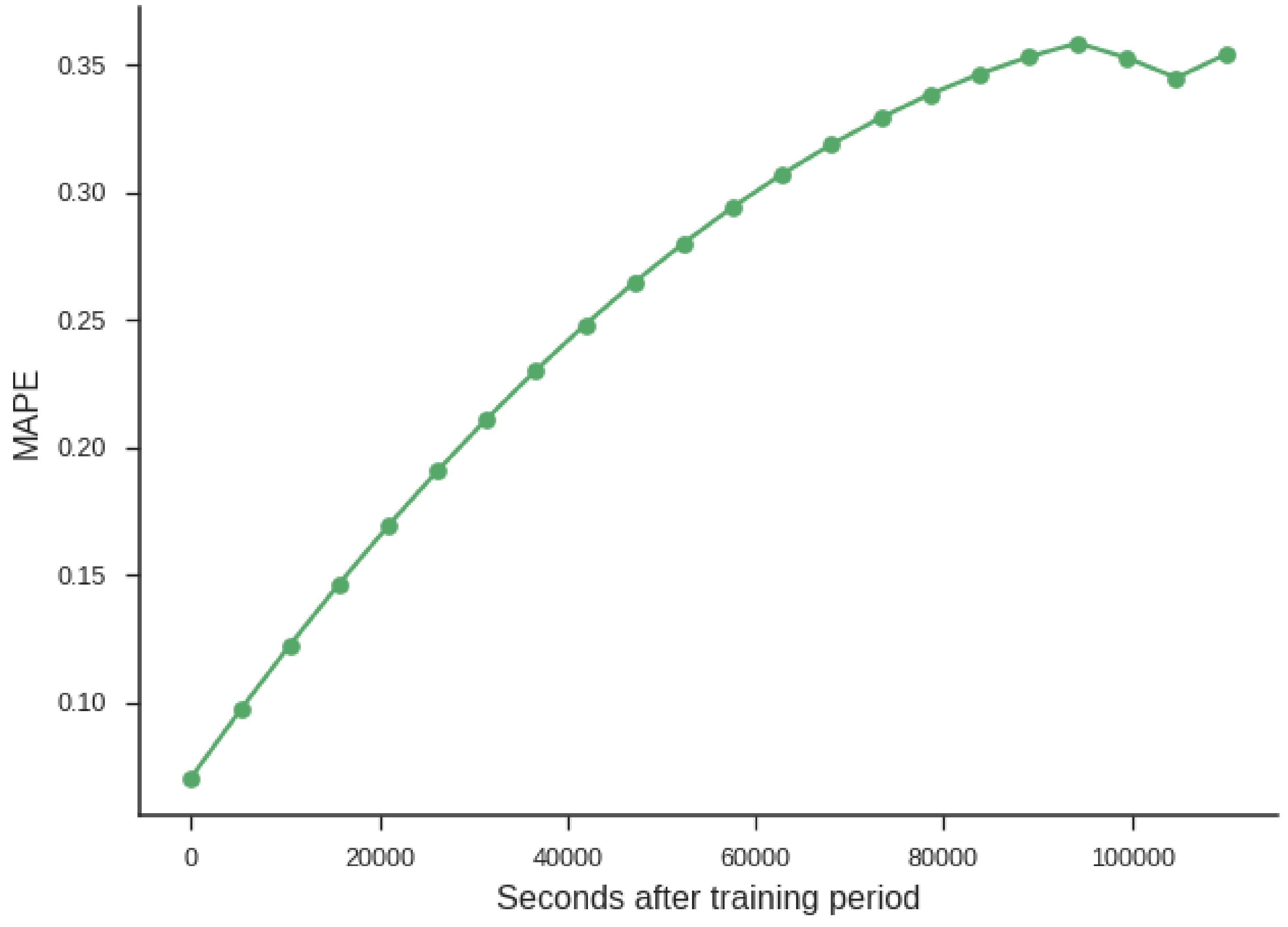
- By maximizing log likelihood function, we can obtain λ_d^* for item d and from the rate function we get

$$\frac{dc^d(t)}{dt} = \lambda_d f_d(t;\theta_d)\left(m + c^d(t)\right)$$

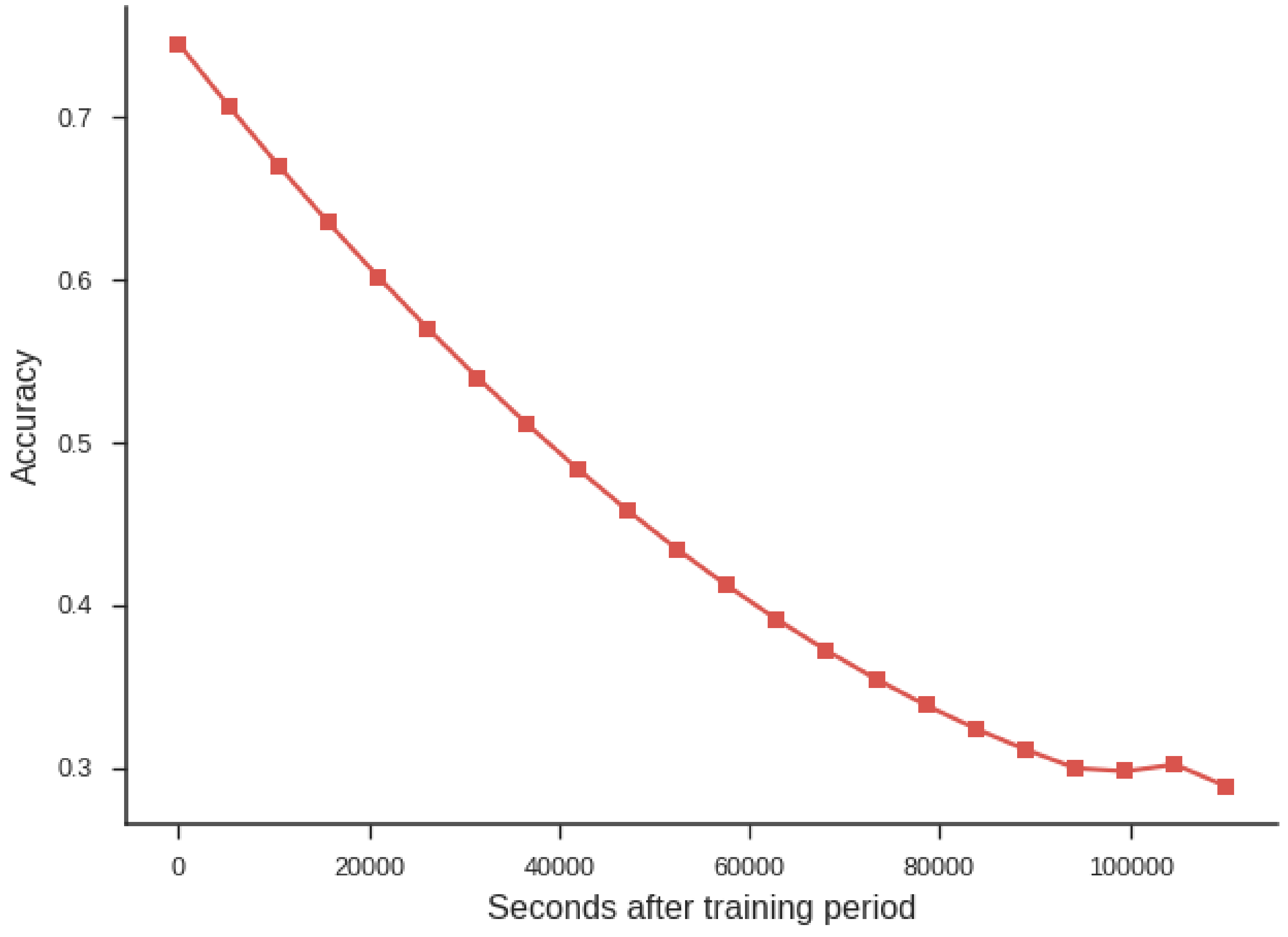
- By solving this equation, we can get the prediction function $c_d(t)$

RPP for Retweet Dynamics

Performance of RPP in predicting retweet dynamics



MAPE



Accuracy

MAPE

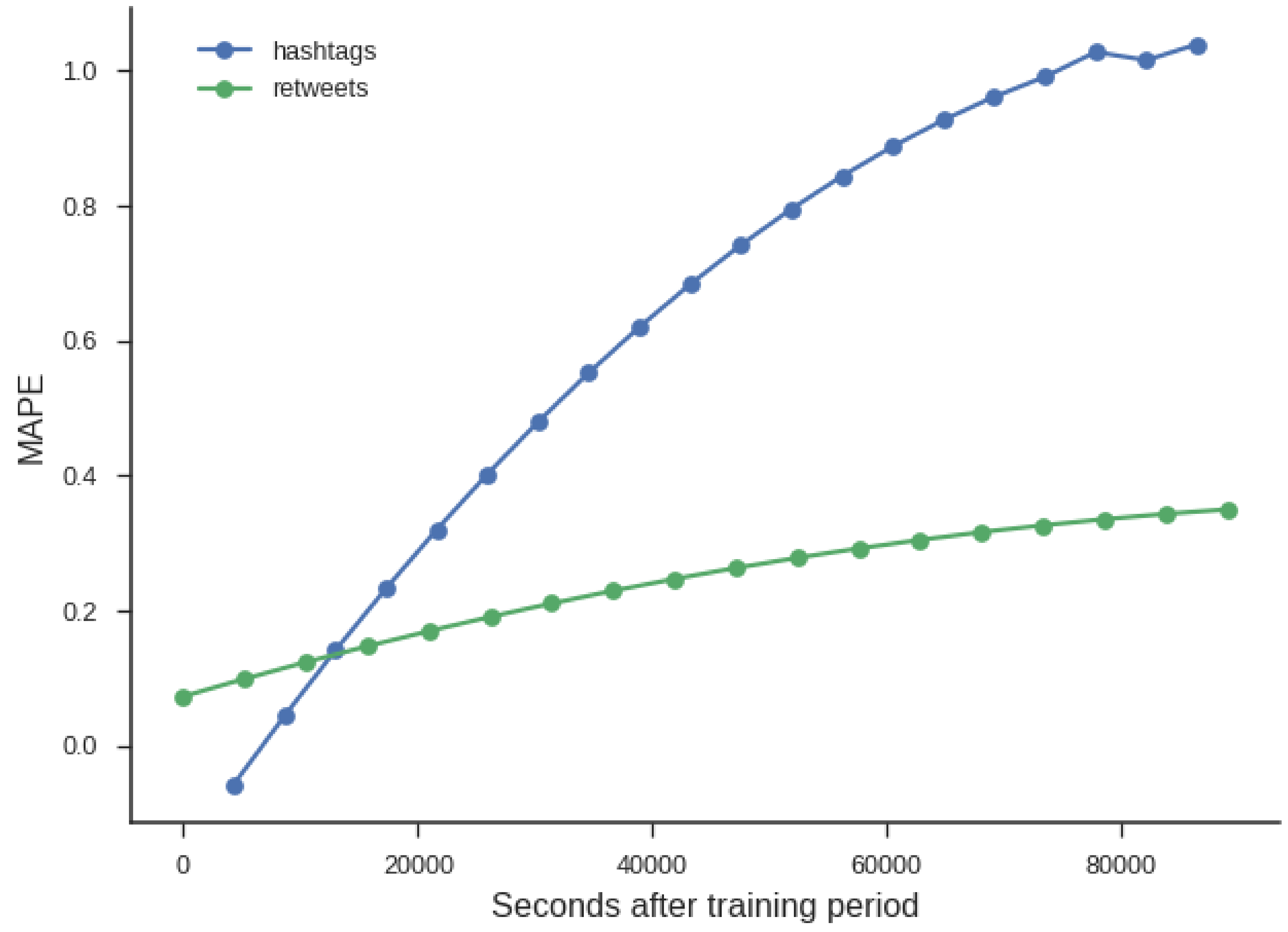
$$\frac{1}{N}\sum_{d=1}^{n}\left|\frac{c^d(t)-r^d(t)}{r^d(t)}\right|$$

Accuracy

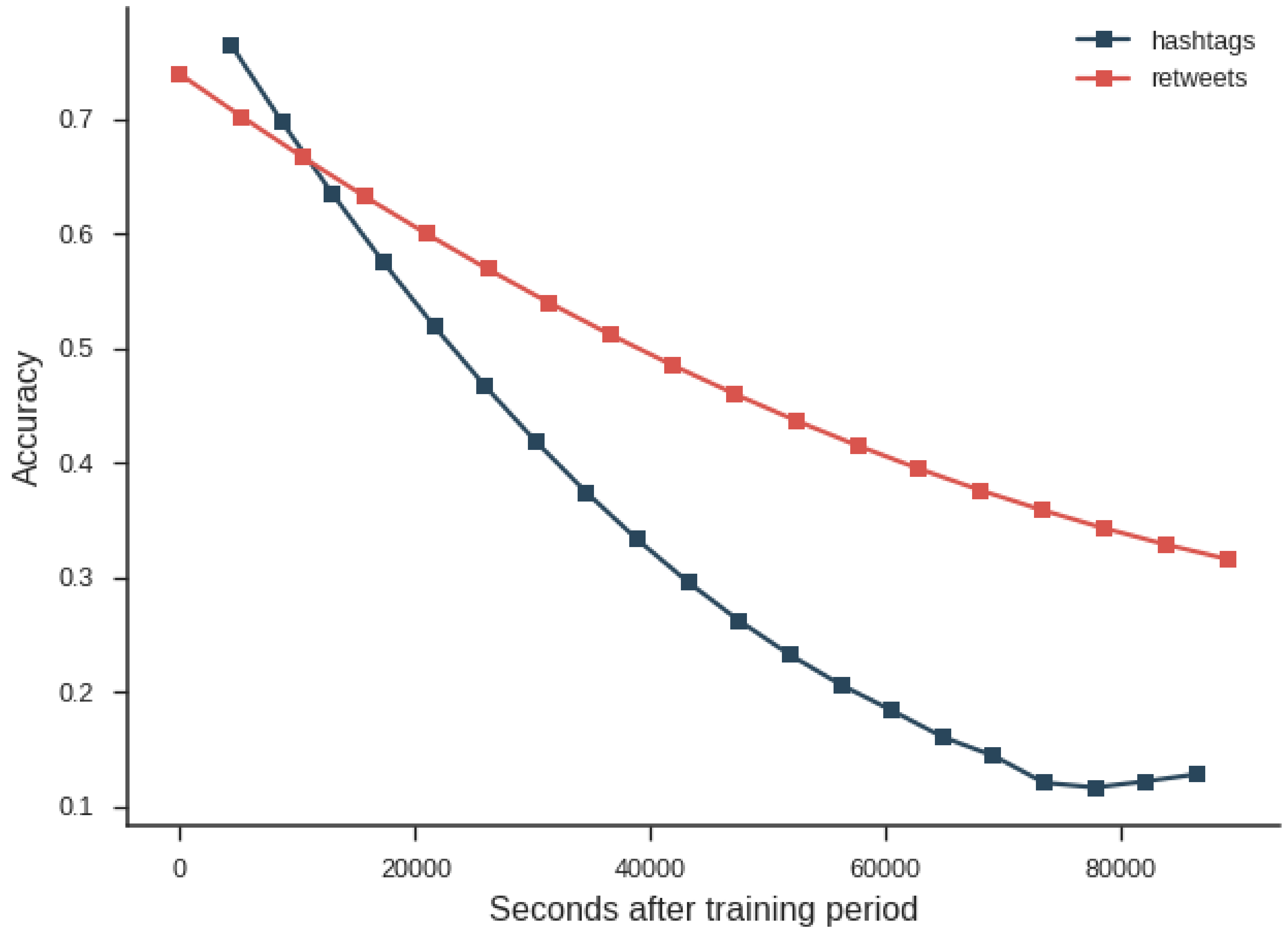
$$\frac{1}{N}\left|\left\{d\colon \frac{c^d(t)-r^d(t)}{r^d(t)}\le\epsilon\right\}\right|$$

RPP for Hashtag Dynamics

Performance of RPP in predicting hashtag dynamics



MAPE



Accuracy

Conclusion

Conversation Graph and Analysis

User categorization and sentiment analysis

Evolution of Askew Bowtie

We found that conversation graph evolves to a stable Askew Bowtie structure

Prioritization algorithm for SRM

How the Askew Bowtie structure can be used for prioritizing tweets for SRM

Retweet Dynamics and RPP

Showed that RPP can be used for efficiently predicting retweet dynamics

Hashtag Dynamics

Relatively less and needs improvement

Publications Based on Thesis

- Mathew, Binny, **T A, Unnikrishnan**, Tanmoy Chakraborty, Niloy Ganguly, and Samik Datta. "Mining Twitter Conversations around E-commerce Promotional Events." In Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion, pp. 345-348. ACM, 2016.

References

[1] Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., ... & Wiener, J. (2000). Graph structure in the web. Computer networks, 33(1), 309-320.

[2] Meusel, R., Vigna, S., Lehmberg, O., & Bizer, C. (2014, April). Graph structure in the web---revisited: a trick of the heavy tail. In Proceedings of the companion publication of the 23rd international conference on World wide web companion (pp. 427-432). International World Wide Web Conferences Steering Committee.

[3] Shen, H. W., Wang, D., Song, C., & Barabási, A. L. (2014). Modeling and predicting popularity dynamics via reinforced poisson processes. arXiv preprint arXiv:1401.0778.

[4] Barabasi, A. L. (2005). The origin of bursts and heavy tails in human dynamics. Nature, 435(7039), 207-211.

[5] Mohammad, S. M., Kiritchenko, S., & Zhu, X. (2013). NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets. arXiv preprint arXiv:1308.6242.