

**Symposium  
on  
Media and Internet Use  
During General Election 2015**

**Wednesday, 27 January 2016**

**Conference Room, Level 1, Oei Tiong Ham Building**

# Using Text Analytics in Analysing GE2015 Online Content

**Professor Lim Ee-Peng**

Director

Living Analytics Research Centre  
School of Information Systems  
Singapore Management University

# Using Text Analytics in Analysing GE2015 Online Content

**LIM Ee-Peng**

School of Information Systems  
Singapore Management University

# The Team



- Aek Palakorn, SMU LARC
- Dai Bingtian, SMU LARC
- Adrian Vu, SMU LARC
  
- Tan Tarn How, NUS IPS
- Carol Soon, NUS IPS
- Siti Nadzirah Binte Samsudin, NUS IPS
- Shawn Goh Ze Song, NUS
- Valerie Yeo Jiewen, NUS

# Motivation

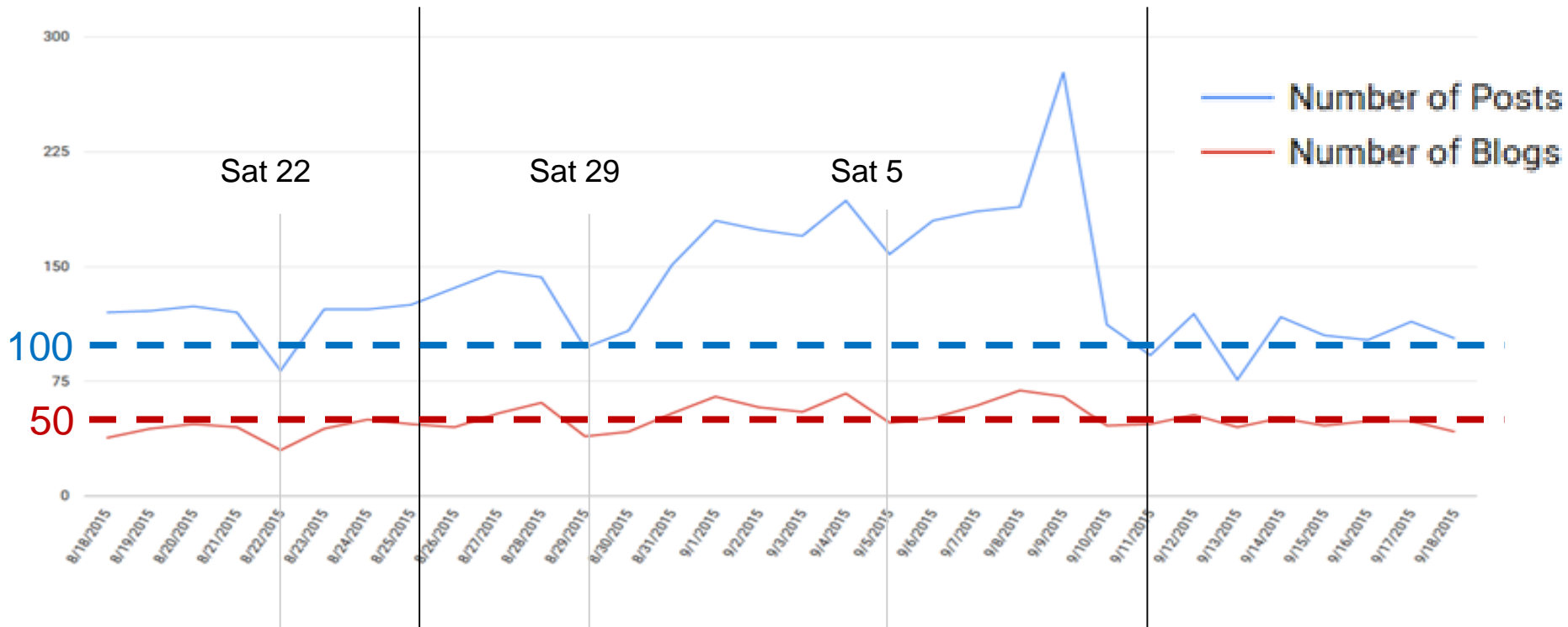


- Manual judgement requires much time and effort.
- Online content is generated at a fast pace.
  - More than 200 blogs relevant to Singapore
  - More than 100 posts each day
- During GE2015, as many as 275 posts were generated in one single day.

# Overall Trends

Dissolution of Parliament  
(25 Aug)

Polling Day  
(11 Sep)

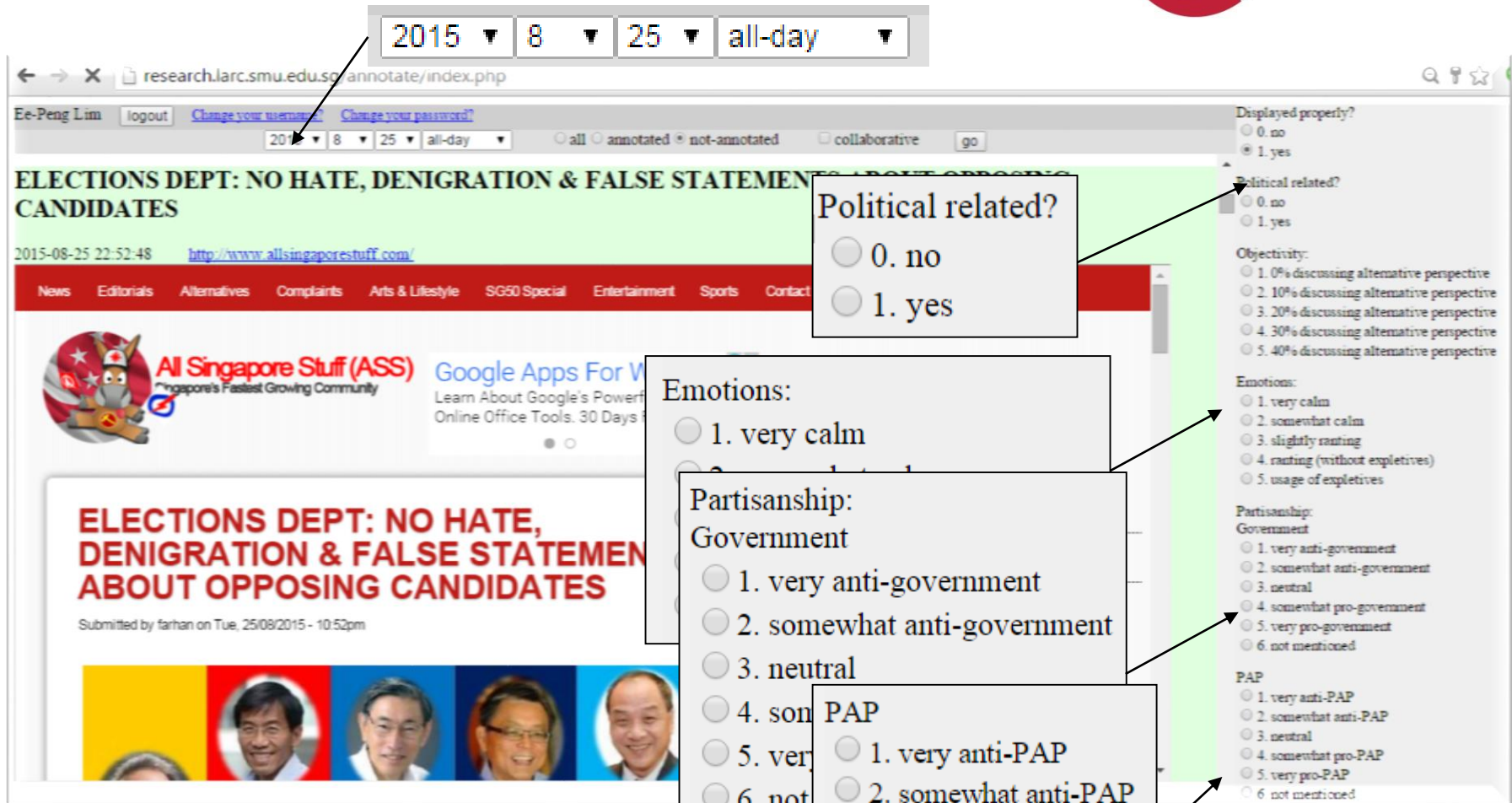


4365 blog posts were generated during the above period.

# Research Questions

- Can text analytics help?
  - Need human experts to provide some training data
- Can we derive useful insights quickly from online content using text analytics?
- Can we create text analytics systems to help us find interesting content?

# Blog Post Coding Tool



The screenshot shows a web browser window with the URL `research.larc.smu.edu.sg/annotate/index.php`. The article title is "ELECTIONS DEPT: NO HATE, DENIGRATION & FALSE STATEMENTS ABOUT OPPOSING CANDIDATES". The article content includes a red header with the same title and a row of five candidate portraits. The coding tool interface on the right side of the page includes several sections with radio button options:

- Political related?**
  - 0. no
  - 1. yes
- Objectivity:**
  - 1. 0% discussing alternative perspective
  - 2. 10% discussing alternative perspective
  - 3. 20% discussing alternative perspective
  - 4. 30% discussing alternative perspective
  - 5. 40% discussing alternative perspective
- Emotions:**
  - 1. very calm
  - 2. somewhat calm
  - 3. slightly ranting
  - 4. ranting (without expletives)
  - 5. usage of expletives
- Partisanship: Government**
  - 1. very anti-government
  - 2. somewhat anti-government
  - 3. neutral
  - 4. somewhat pro-government
  - 5. very pro-government
  - 6. not mentioned
- PAP**
  - 1. very anti-PAP
  - 2. somewhat anti-PAP
  - 3. neutral
  - 4. somewhat pro-PAP
  - 5. very pro-PAP
  - 6. not mentioned

Coding Period: May 20



# Coding Schemes

Scale	Political?	Label
0	No	Non-Political
1	Yes	Political

Scale	Partisanship	Label
1	Very anti-government/PAP	Partisan
2	Somewhat anti-government/PAP	Partisan
3	Neutral	Non-Partisan
4	Somewhat pro-government/PAP	Partisan
5	Very pro government/PAP	Partisan

# Coding Schemes

Scale	Emotionality	Label
1	Very calm	Non-Emotional
2	Somewhat calm	Non-Emotional
3	Slightly ranting	Emotional
4	Very ranting	Emotional
5	Usage of expletives	Emotional

# Training The Classifiers

- Using the human labeled posts before GE2015 was announced, we trained classifiers to assign **political**, **partisanship** and **emotionality** labels using blog post content.

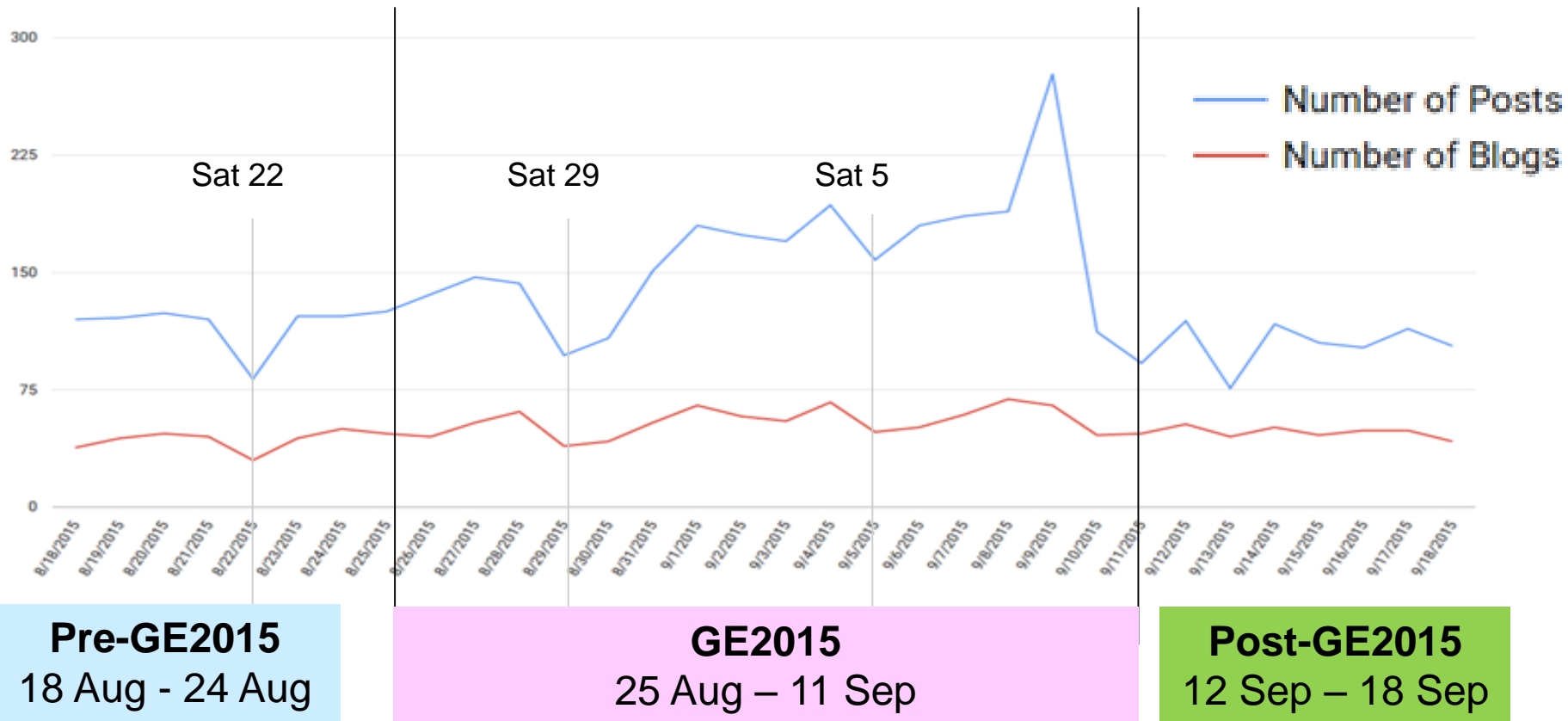
Labeled Data	Size (num of docs)	# Positive Examples	# Negative Examples	Accuracy (F1,10-CV)
Political	914	476	438	0.86
Partisan	461	278	183	0.78
Emotionality	462	174	289	0.71

- We then applied the classifiers to predict **political**, **partisanship** and **emotionality** labels to the posts during the GE2015 period.

# Overall Trends

Dissolution of Parliament  
(25 Aug)

Polling Day  
(11 Sep)



# Classified Posts

Classified Posts	Pre-GE2015 18 Aug - 24 Aug (# Posts)	GE2015 25 Aug - 11 Sep (# Posts)	Post-GE2015 12 Sep - 18 Sep (# Posts)
Political Posts	393	1756	353
Political Posts (% of all posts)	55%	74%	56%

Total Number of Blog Posts: 4365

- About 55% of all posts are political before and after GE2015.
- Proportion of political posts increases significantly to 74% during GE2015.

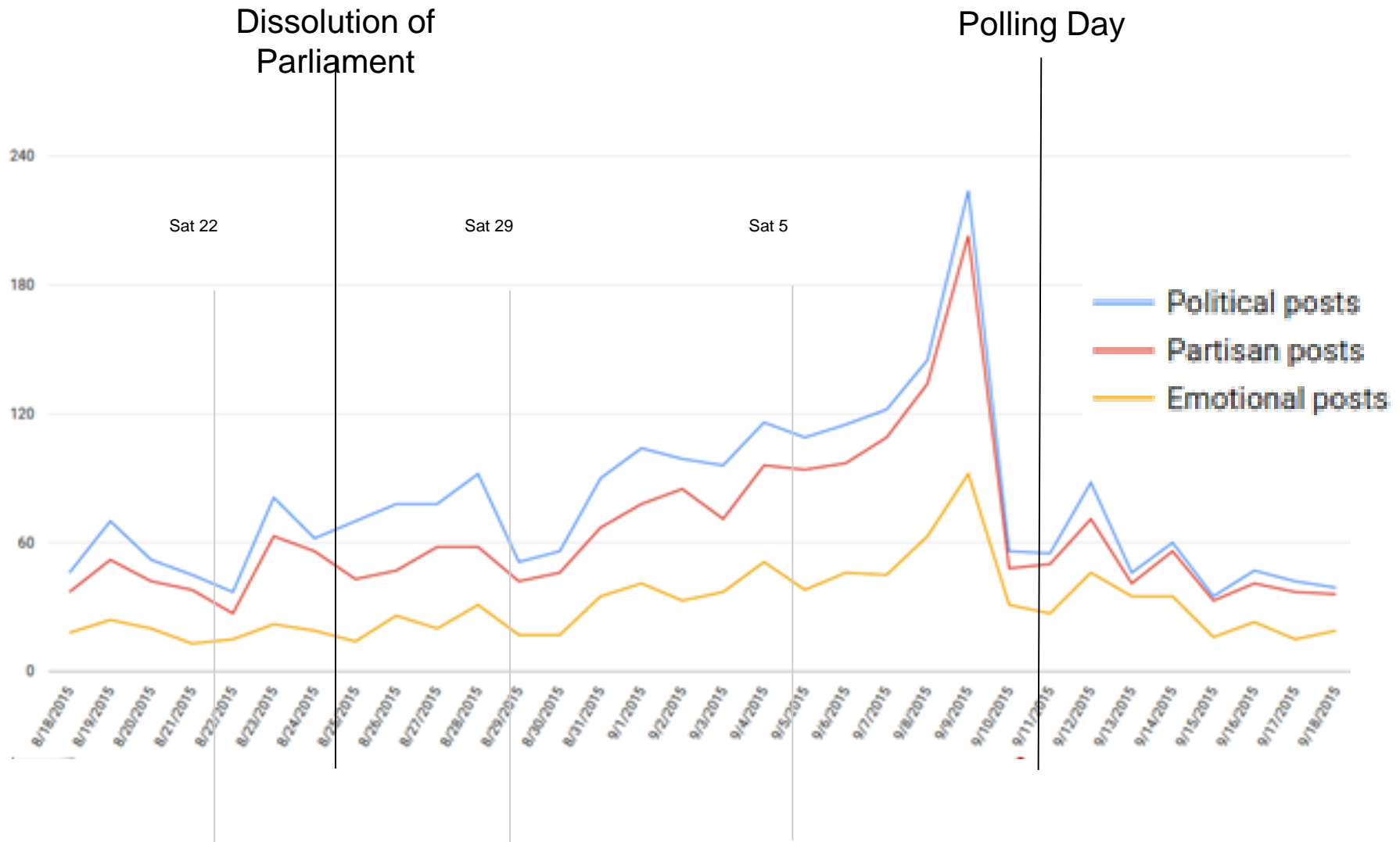
# Classified Posts

<b>Label</b>	<b>Pre-GE2015</b> 18 Aug - 24 Aug (# Posts)	<b>GE2015</b> 25 Aug - 11 Sep (# Posts)	<b>Post-GE2015</b> 12 Sep - 18 Sep (# Posts)
Political	393	1756	353
Partisan	315 (80%)	1426 (81%)	315 (89%)
Emotional	131 (33%)	664 (38%)	189 (54%)

(x% of classified political posts)

- Most political posts are partisan oriented.
  - Partisan posts increase slightly after GE2015.
- About 1/3 political posts are emotional.
  - Emotional posts increase significantly after GE2015.

# Trends: Political, Partisan, and Emotional Posts



# Political Post Classification: Top Features



- Minister
- PAP
- Singaporean
- Vote
- Prime
- Ministry
- Policy
- Wage
- Lee
- Parliament
- Amos
- Opposition
- Development
- Officer
- Society
- Yew
- Lim
- Authority
- Education
- Security



# Partisan Post Classification: Top Features



- Population
- Income
- Award
- Conservative
- Medical
- Active
- Breach
- Unfair
- Singaporean
- Expense
- Salary
- PAP
- Vote
- Prison
- Battle
- War
- Passenger
- Tax
- Game
- Power

# Emotional Post Classification: Top Features



## Words:

- Bed
- Malay
- Warning
- Bank
- Political
- Heart
- Taxpayer
- Singaporean
- Content
- PAP
- Game
- Age
- Vote
- Solution
- Council
- Island
- NEA
- Malaysia

## Stylometric:

### Punctuations

- #
- \*
- %
- ?
- &

## Stylometric:

### Repeating Chars

- \_
- h
- x
- w

## Stylometric:

### Political Sentiment Words

- Afraid
- Careful
- Warn
- Hate
- Fair
- Critic
- Strang\*
- Complain
- Well
- Shield
- Aristocra\*

# Exclusive Word Clouds: Political Posts

Election



Post Election



# GE2015 Analytics System





<http://research.larc.smu.edu.sg/ge2015/>




#GE2015   Persons ▾   Parties ▾   Constituencies ▾   Issues ▾   **Blog Posts**   Media Posts   ⓘ





the Bangla terrorists, I will take my chance to write about racial politics. I am a keen admirer of Mdm Speaker Halimah Jacob, and a strong advocate for her to ... Continue reading —


[thesingaporebeacon.com](http://thesingaporebeacon.com)  
21 Jan, 01:03 am

   0    0

 **MEDIA LITERACY COUNCIL**





**Doubt on the necessity of Media Literacy Council's existence - The Online Citizen**


   0    0


 **My nephew the newly minted police officer | Gintai\_昇泰**

Last Saturday (16/1/16), my nephew graduated from the 3-month residential Police Officer Basic Course (NS) at Home Team Academy. He was amongst the 600 over National Servicemen enlisted to serve in the police undergoing the 3-month training. There ...

[gintai.wordpress.com](http://gintai.wordpress.com)  
20 Jan, 11:28 pm

   0    0

 **WP's airborne trooper MAJ (NS) Pritam Singh shows up in court in No. 4 | Mothership.SG**

 **Are you (nationality) first, and (ethnicity)**

# GE2015 Analytics System

<http://research.larc.smu.edu.sg/ge2015/>




#GE2015 Persons ▾ Parties ▲ Constituencies ▾ Issues ▾ ⓘ

Democratic Progressive Party 9 National Solidarity Party 36 People's Action Party 142 People's Power Party 6 Reform Party 28

Singapore Democratic Alliance 16 Singapore Democratic Party 26 Singapore People's Party 27 Singaporeans First 19 Workers' Party 110

1 Sep, 08:11 am

 **SDA can have national impact, says party chief, Politics News & Top Stories - The Straits Times**

The Singapore Democratic Alliance (SDA) may be contesting only one ward, but it believes it can have a national impact through the issues it is championing, said party chief Desmond Lim Bak Chuan yesterday.. Read more at [straitstimes.com](http://straitstimes.com).

1 Sep, 08:09 am

Photo not available

**'We'll run a constructive and positive campaign', Politics News & Top Stories - The Straits Times**


The flags and red balloons are ready. An army of tiny teddy bears, declaring "I Love SDP" on their T-shirts, is raring to go.. Read more at [straitstimes.com](http://straitstimes.com).

1 Sep, 08:09 am


 **Theatre Uncle Lim Kay Siu Blasts SG Ministers, Demands PM Lee's Resignation - RedWire Times Singapore | RedWire Times Singapore**

One of the old guard of Singapore theatre, Lim Kay Siu, has lashed out the Singapore government, and Prime Minister Lee Hsien Loong. In a recent Facebook post, Lim accused PM Lee failing to serve Singaporeans, and losing their trust. He urged PM ...

1 Sep, 08:07 am

 **Five Stars And a MoonUnlikely Candidates in an Apolitical Party - Five Stars And a Moon**

Children of dissidents in the People's Action Party? Is it true?

 **Independent candidate, Tan Lam Siong backs out from 3 cornered fight at Potong Pasir | The Online Citizen**

1 Sep, 07:59 am

# Conclusion

- Can text analytics help?
  - Accuracy depends on amount of training data.
  - Training data requires some efforts but software can help to reduce these efforts
- Insights from text analytics?
  - During GE2015, there are significantly more and larger proportion of political posts.
  - The last day of GE2015 campaigning saw a significant increase in political blog posts.
  - Cooling day reduces political blog posts.
  - Right after GE2015 result announcement, partisan and emotional posts increased by proportion.

# Looking Forward



- Bloggers  $\neq$  User population
- Social media users
  - Social media users may be silent
  - Users may read topics different from what they write about.
- Combine surveys with observed social media analytics

Wei Gong, Ee-Peng Lim, Feida Zhu. *Characterizing Silent Users in Social Media Communities*. In International AAAI Conference on Web and Social Media (ICWSM), Oxford, 2015.

Wei Gong, Ee-Peng Lim, Feida Zhu. *Posting Topics Are Not Reading Topics: On Discovering Posting and Reading Topics in Social Media*, In International School and Conference on Network Science (NetSciX), 2016.



# Thank you

This research is supported by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM Programme Office, Media Development Authority (MDA)



**Symposium  
on  
Media and Internet Use  
During General Election 2015**

**Wednesday, 27 January 2016**

**Conference Room, Level 1, Oei Tiong Ham Building**