

# **CORRELATING SENTIMENTS AND TOPICS WITH SPAM WAVES ON SOCIAL NETWORKS**

**Daniel Dichiu**

# Targeted attacks

- Manipulation ingredients
  - Ethos – be credible, be trustworthy
  - Pathos – appeal to user's emotion
  - Logos – reasonable facts and language

# Ethos

- Post as a friend
- Credibility transfer
- “This is how all the professionals do..  
They use this free software to add free credits to their accounts..  
So they can win all the time [http://bit.ly/wbf\\*\\*\\*](http://bit.ly/wbf***)”

# Pathos

- “Hey Checkout Who views your pr0file at -->>[http://goo.gl/H9\\*\\*\\*](http://goo.gl/H9***)”
- “What does your Birthday color say about you?  
-> -> [http://on.fb.me/H5c\\*\\*\\*](http://on.fb.me/H5c***)”
- “Amazing!  
we are giving 300 Credits for sharing this post to your friends!  
Claim your prizes here  
>>> [http://on.fb.me/xWv\\*\\*\\*](http://on.fb.me/xWv***) <<<  
Limited Time! Hurry!  
Please don't forget to click LIKE and SHARE. Thanks!”

# Logos

- “No Survey, No Personal Info, No Scam and No other Shit..  
Don't trust people who try to lie you with any bullshit that they are posting around..  
THE ONLY THING THAT HAS WORKED FOR ME SO FAR IS THIS PAGE..  
[http://goo.gl/Lk\\*\\*\\*](http://goo.gl/Lk***)”

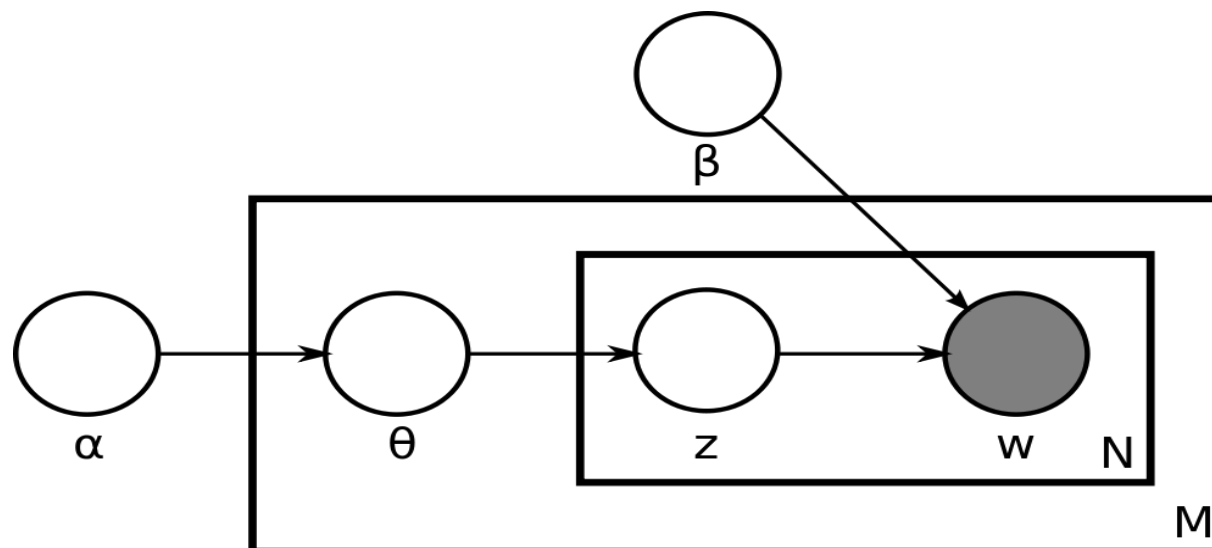
- “Wow I Just Got Free Facebook Credit  
Its not like other spam really claim you Facebook credit now  
[http://google\\*\\*\\*credit.blogspot.in/](http://google***credit.blogspot.in/)”

# Sentiment Analysis

- Features
  - Unigrams
  - Bigrams
- Generative models
  - Naïve Bayes
- Discriminative models
  - Maximum entropy
  - Support Vector Machines

# Topic Extraction

- Algorithms
  - Latent Dirichlet Allocation



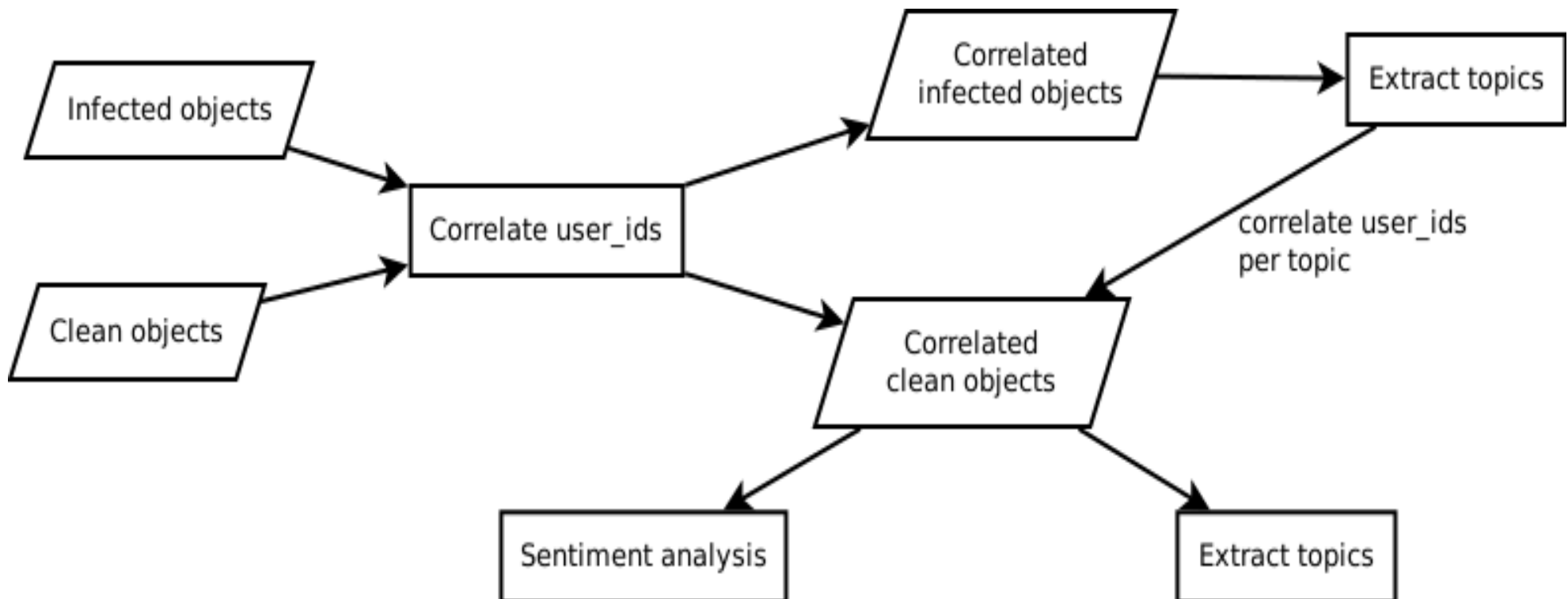
- Latent Semantic Indexing

# The Setup

- 210 657 infected objects
- 2 852 964 clean objects
- 16 834 users
- Pre-process objects
  - Tokenization
  - Stop-words removal



# The Setup



# The Setup

- Python
- Topic extraction
  - Gensim – both LDA and LSI
  - Machine Learning for Language Toolkit (MALLET)  
– only LDA
- Sentiment analysis
  - Shogun toolkit – SVM with Steepest Gradient Descent

# The Results

- Infected objects topic extraction

Method	Topics	Examples
LSI	19	<ul style="list-style-type: none"><li>- check, profile, views, viewing, viewed, today</li><li>- facebook, colors, themes, switch, goodbye, pink, blue</li><li>- dad, check, son, click, blocked, checked, timeline, credit</li></ul>
LDA	35 (5 from MALLET 30 from Gensim)	<ul style="list-style-type: none"><li>- free, credits, live, add, win, time, wow, software, professionals, accounts</li><li>- free, limited, time, giving, tickets, end, offer, mail, flags</li><li>- timeline, remove, back, users, facebook, profile, put, extensions</li><li>- pair, today, shoes, promotion, converse, ending, time, members, work, free</li><li>- boy, profile, girl, visit, checked, angry, stop, lying, visiting</li></ul>

# Topics Correlations

## Topics – spam objects

- happened, life, save, man, donate, work, release, companies, year, banner

- receive, read, show, opportunity, women, survey, tomorrow, miss, late

- hours, life, omg, wasted, facebook, cracked, check, easily, clothes

## Topics – clean objects

- contact, apply, joining, give, bonus, click, income, cityville, send  
- receive, contact, winners, apply, joining, twilight, boy, share, give, fight, bella, sag

- birthday, message, family, send, found, shoes  
- birthday, check, events, outerwin, message, butterfly, family, shoes

- days, black, doubt, cost, disorder, eating, couple, male, eating  
- doubt, factors, eating, disorders, add, male

# Sentiment Analysis

<b>Infected Topic</b>	<b>Clean NAPS</b>	<b>Clean NANS</b>	<b>Clean NASS</b>	<b>Clean NAOS</b>
Check, profile, views, viewing, viewed, today	0.25	0.28	0.27	0.38
Facebook, colors, themes, switch, goodbye, pink, blue	0.43	0.63	0.71	1
Check, total, views, facebook, photobucket, friend	0.65	0.89	1	0.72
Dad, check, son, click, blocked, checked, timeline, credit	0.54	0.53	0.48	0.47

# Sentiment Analysis

Infected Topic	Clean NAPS	Clean NANS	Clean NASS	Clean NAOS
Month, member, services, free, message, copy, blue	0.34	0.33	0.26	0.45
Free, credits, live, add, win, time, wow, software, professionals, accounts	1	1	0.24	0.46
Free, limited, time, giving, tickets, end, offer, mail, flags	0.8	0.72	0.24	0.5
Timeline, remove, clicl, back, users, facebook, profile, put, extensions	0.4	0.36	0.24	0.49

# Conclusions

- No conclusive evidence of targetted attacks
- Based on spam samples, attackers know how to manipulate
- Based on this study, attackers don't know who to manipulate

Thank you