Viewpoint Discovery and Understanding in Social Networks

May, 2018

Mainul Quraishi
L3S Research Center,
Leibniz University of Hannover
Hannover, Germany
quraishimainul@gmail.com

Pavlos Fafalios
L3S Research Center,
Leibniz University of Hannover
Hannover, Germany
fafalios@L3S.de

Eelco Herder
Radboud Universiteit Nijmegen,
Institute for Computing and
Information Sciences
Nijmegen, the Netherlands
eelcoherder@acm.org

Introduction & Motivation

➢ Social Media:
 ➢ Dominant platform to comment & discuss on breaking news and events
 ➢ Enable the expression of diverse opinions and **viewpoints**

➢ However:
 ➢ Media bias
 ➢ Echo Chamber
 ➢ Filter bubbles

http://theday.co.uk/technology/research-reveals-facebook-echo-chambers

➢ Social costs and effects
 ➢ As also observed in the US election 2016 (Allcott and Gentzkow, 2017)

➢ Need of methods to **detect** and **understand** the different viewpoints about a topic
 ➢ To better understand political or societal debates
 ➢ To better analyze and interpret course of historical events retrospectively

Viewpoint Discovery & Understanding in Social networks, WebSci'18, Amsterdam, The Netherlands, May 27-30, 2018.

1/19

Contribution

➢ **Viewpoint Discovery**
- Given a topic, detect the different viewpoints discussed in a Social Network (Twitter)
- Cluster the users based on their viewpoint on the topic (by exploiting retweets)

➢ **Viewpoint Understanding**
- What is a viewpoint about?
  - *by identifying a list of descriptive terms that characterize the viewpoint*
- What is a term about?
  - *by identifying other descriptive terms*

Viewpoint Discovery & Understanding in Social networks, WebSci'18, Amsterdam, The Netherlands, May 27-30, 2018.

2/19

Related Work

- ➤ Viewpoint discovery in Documents:
    - TAM (Paul and Girju,2010)
    - VODUM (Thonet et al, 2016)
- ➤ Viewpoint discovery in Social Media:
    - SNVDM-GPU (Thonet et al, CIKM 2017)
        - ▪ Detect viewpoints of subtopics related to a controversial topic
        - ▪ Show descriptive terms that characterize a viewpoint
    - ▪ Others:
        - ▪ Supervised models [Fang et al. 2015, Cohen et al. 2013]
        - ▪ Contrastive opinion mining (Ren et al,2016)
        - ▪ Controversy detection [Garimella et al. 2016]
        - ▪ Viewpoint discovery in forums [Qiu and Jiang, 2013]

Viewpoint Discovery & Understanding in Social networks, WebSci'18, Amsterdam, The Netherlands, May 27-30, 2018.

3/19

Our Approach

- Viewpoint Discovery
  - Use of **interaction graph** (retweet graph)
    - By first gathering tweets related to the topic
  - Apply **multilevel graph partitioning** (as done by Garimella et al, 2016)
  - Exploitation of **conductance** clustering quality metric
    - Find clusters representing concrete viewpoints
    - Identify noisy clusters

- Viewpoint Understanding
  - Apply a **Rank Difference** approach **iteratively**

- Contrary to existing works:
  - Applicable to unknown number of viewpoints
  - Detection of noisy groups with no clear viewpoint
  - Deep understanding of viewpoints (through focusing on specific terms)

Viewpoint Discovery & Understanding in Social networks, WebSci'18, Amsterdam, The Netherlands, May 27-30, 2018.

4/19

Viewpoints Discovery

Choosing the proper graph clustering method

  ➢ Main requirement: Ensuring <u>sparse connection</u> between different clusters
  ➢ No edge between the users ≠ different viewpoints
  • But: users with different viewpoints will definitely not endorse each other

  ➢ Various ways of clustering
  ▪ Node similarity
  ▪ The Louvain method (Community Detection)
  ▪ Multi-Level Graph Partitioning (MLGP)
    • Using METIS
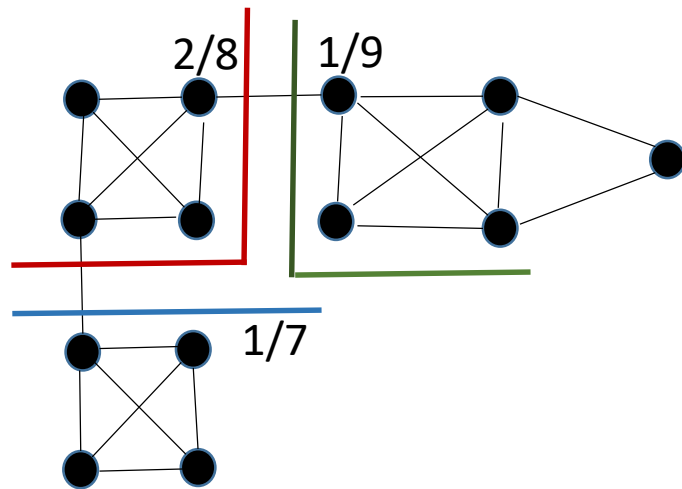   • Problem: what number of clusters to consider (**K** value)?

Viewpoint Discovery & Understanding in Social networks, WebSci'18, Amsterdam, The Netherlands, May 27-30, 2018.

5/19

- ➤ **Conductance**:
  - ▪ Ensures sparse connection between the different clusters!

$$\text{Conductance} = \frac{\text{size of cut (edges connecting the clusters)}}{\text{volume of smaller cluster devided by that cut}}$$

Volume= total number of edges starting from **nodes in the cluster**
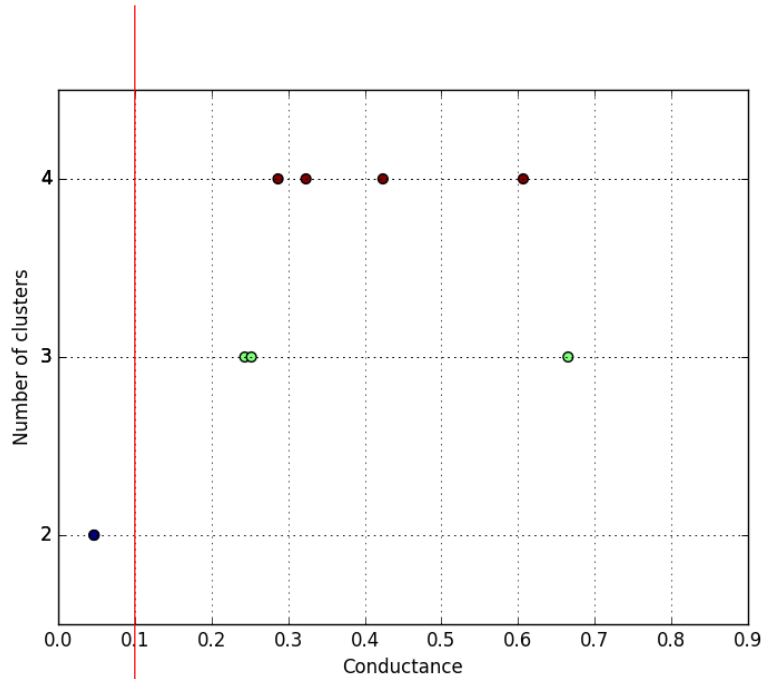

2/8   1/9

1/7

Advantages of Conductance:
- It considers the cut size
- It considers the volume of the cluster

Viewpoint Discovery & Understanding in Social networks, WebSci'18, Amsterdam, The Netherlands, May 27-30, 2018.

6/19

Viewpoints Discovery

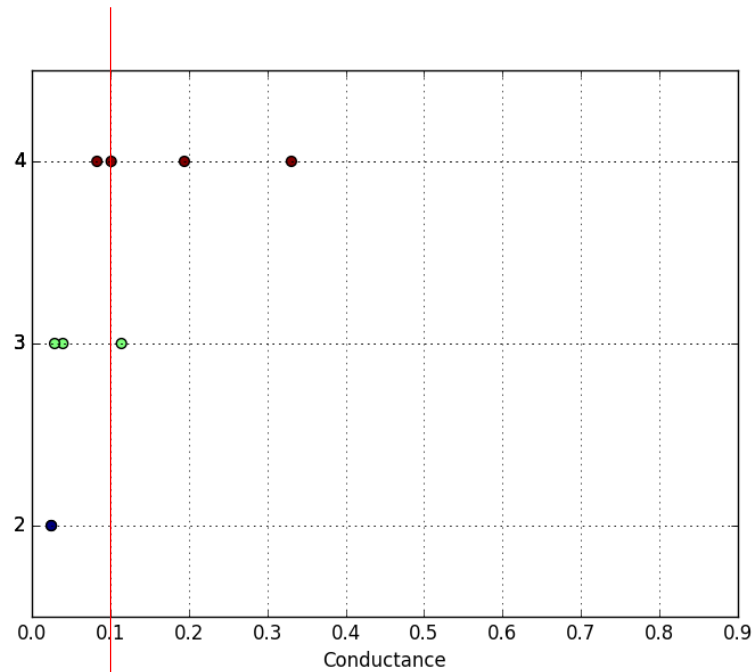## Deciding on the clusters that hold different viewpoints

➢ We inspect the conductance of the clusters for different values of **k**

➢ Conductance Threshold:
  • A value of 0.1 seem to be good threshold
    • *10% of the edges in the cluster point to nodes outside the cluster!*
  • Conductance < 0.1 means well separated therefore good clustering
  • Conductance > 0.1, means most likely noise

➢ For k=2, if the conductance >0.1, the topic is probably not controversial!

➢ Algorithm:
  • Start from k = 2 and check the conductance value
  • Continue for k > 2 until we get the maximum number of clusters below the threshold for the larger k value

Viewpoint Discovery & Understanding in Social networks, WebSci'18, Amsterdam, The Netherlands, May 27-30, 2018.
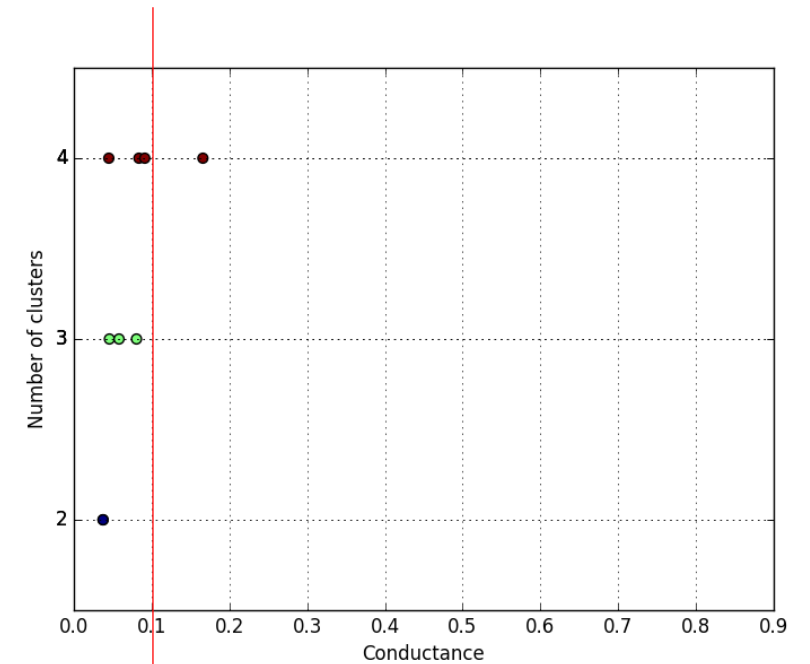
7/19

Viewpoints Discovery

Conductance of the cut of each cluster for various k values

Indyref

US Election

Brexit

Threshold: 0.10 (Red line)

Viewpoint Discovery & Understanding in Social networks, WebSci'18, Amsterdam, The Netherlands, May 27-30, 2018.

8/19

Viewpoints Understanding

➢ **Rank Difference**: automatic term recognition algorithm (Kit & Liu, 2008)
  - Operates over two ranked lists of terms: one extracted from the "**subject corpus**" (corpus of interest) and one extracted from a "**contrasting corpus**" (e.g., general corpus)
  - Ranking criterion for the lists: term frequency
  - Score of a term:

$$score(w) = \frac{rank(w, W_c)}{|W_c|} - \frac{rank(w, W_s)}{|W_s|}$$

  - Focuses on filtering out frequent terms occurring in both lists

Indian	European
Onion	Onion
Garlic	Garlic
Chili	**Black pepper**
Cumin	**Paprika**
Ginger	Ginger
Black pepper	Tomato
Paprika	Chili

Viewpoint Discovery & Understanding in Social networks, WebSci'18, Amsterdam, The Netherlands, May 27-30, 2018.

9/19

Viewpoints Understanding

➢ **Iterative Rank Difference (IRD)**

1. First Iteration
   - ▪ What is a viewpoint about?
   - ▪ Extract Top-n terms after running Rank Difference
     - • <u>Subject list</u>: frequent terms in tweets of a viewpoint
     - • <u>Contrasting list</u>: frequent terms in tweets from all **the other viewpoints**

2. Next Iterations
   - ▪ What is a term (w) about?
     - ▪ <u>Subject list</u>: frequent terms that **co-occur** with w in tweets about the viewpoint
     - ▪ <u>Contrasting list</u>: frequent terms that **do not co-occur** with w in tweets **about the same viewpoint**

Viewpoint Discovery & Understanding in Social networks, WebSci'18, Amsterdam, The Netherlands, May 27-30, 2018.

10/19

- Comparative evolution with topic modeling proposed by Thonet et al. [CIKM, 2017]
- Indyref: 2 viewpoints (YES and NO)
- Midterms: 2 viewpoints (Democrat and Republican)
- Evaluation metrics:
  - Purity: proportion of users whose are correctly clustered compared with the ground truth class
  - NMI: Based on Mutual Information and Entropy

Method	Indyref		Midterms	
	Purity	NMI	Purity	NMI
SNVDM-GPU	0.969	0.800	0.964	0.778
MLGP	**0.988**	**0.908**	**0.983**	**0.876**

Purity and NMI comparison between SNVDM-GPU and Multi-Level Graph Partitioning (MLGP)

Viewpoint Discovery & Understanding in Social networks, WebSci'18, Amsterdam, The Netherlands, May 27-30, 2018.

11/19

Qualitative Analysis of Viewpoint Understanding

- Indyref
  - Ground truth dataset (YES or NO)
  - First Iteration:

Viewpoint 1	Viewpoint 2
#voteyes, #yes,westminister,	#nothanks, #bettertogether,
Meeting, independent,	currency, #labourno,#scotdecides
#scotland, murphy, event	speech, alex, #voteno, part
national, folk	seperation

Viewpoint Discovery & Understanding in Social networks, WebSci'18, Amsterdam, The Netherlands, May 27-30, 2018.

12/19

Qualitative Analysis of Viewpoint Understanding

➢ Indyref

- ▪ Ground truth dataset (YES or NO)
- ▪ Second Iteration:

murphy	{#voteyes, #yes}
leader	Scoltland,
war	independent
tax	power,
labour	future,
really	country

Viewpoint Discovery & Understanding in Social networks, WebSci'18, Amsterdam, The Netherlands, May 27-30, 2018.

13/19

Qualitative Analysis of Viewpoint Understanding

➢ US Election
  ▪ First Iteration:

Viewpoint 1	Viewpoint 2
#nevertrump, #theresistance	#tcot, #trump2016, #pjnet,
#obamafarewell, #iamwithher	#trumptrain, #draintheswamp,
#resist, #trumps, gop, tweet	video,#makeamericagreatagain,
#gop, #trumpleaks, tax	#fakenews, breaking, god
#notmypresident, putin	#realdonaldtrump, usa, fbi

Viewpoint Discovery & Understanding in Social networks, WebSci'18, Amsterdam, The Netherlands, May 27-30, 2018.

14/19

Qualitative Analysis of Viewpoint Understanding

➤ US Election

- ▪ Second Iteration:

#trumpleaks	fbi
#trumprussia	investigation
#impeachtrump	comey
#dworkinreport	reopen
#amjoy	director
#muslimban	case

Viewpoint Discovery & Understanding in Social networks, WebSci'18, Amsterdam, The Netherlands, May 27-30, 2018.

15/19

Qualitative Analysis of Viewpoint Understanding

➢ Brexit

- First Iteration:

Viewpoint 1(Pro-Brexit)	Viewpoint 2 (Anti-Brexit)	Viewpoint 3 (Neutral)
Democracy,	#tory, tory	#ue, unido, reino, #maga
#nexit, try, #go,	Johnson,	#stocks, #cdnpoli, europa
#referendum	#theresamay	londres, #americafirst,
#leave, #britain	Boris, #votein	royaumeuni, #trumptrain
free, control	#stopbrexit	#healthinnovations,
#leaveeu	#nhs, cost ,nhs	#banking, #pharma

Viewpoint Discovery & Understanding in Social networks, WebSci'18, Amsterdam, The Netherlands, May 27-30, 2018.

16/19

Qualitative Analysis of Viewpoint Understanding

➢ Brexit

- ▪ Second Iteration:

control	{boris, johnson}	banking
#takecontrol,	Conservative,	#stock
Border, uncontrolled,	Biscuit,	#healthinnovations
#betteroffout,	#foreignsecratery, #mp,	#pharma
#takecontrol,	#borisjohnson	decrease, index

Viewpoint Discovery & Understanding in Social networks, WebSci'18, Amsterdam, The Netherlands, May 27-30, 2018.

17/19

Conclusion and Future Works

➢ Graph clustering using **multi-level graph partitioning** and **conductance** quality metric
  - Discover groups of users that represent clear viewpoints
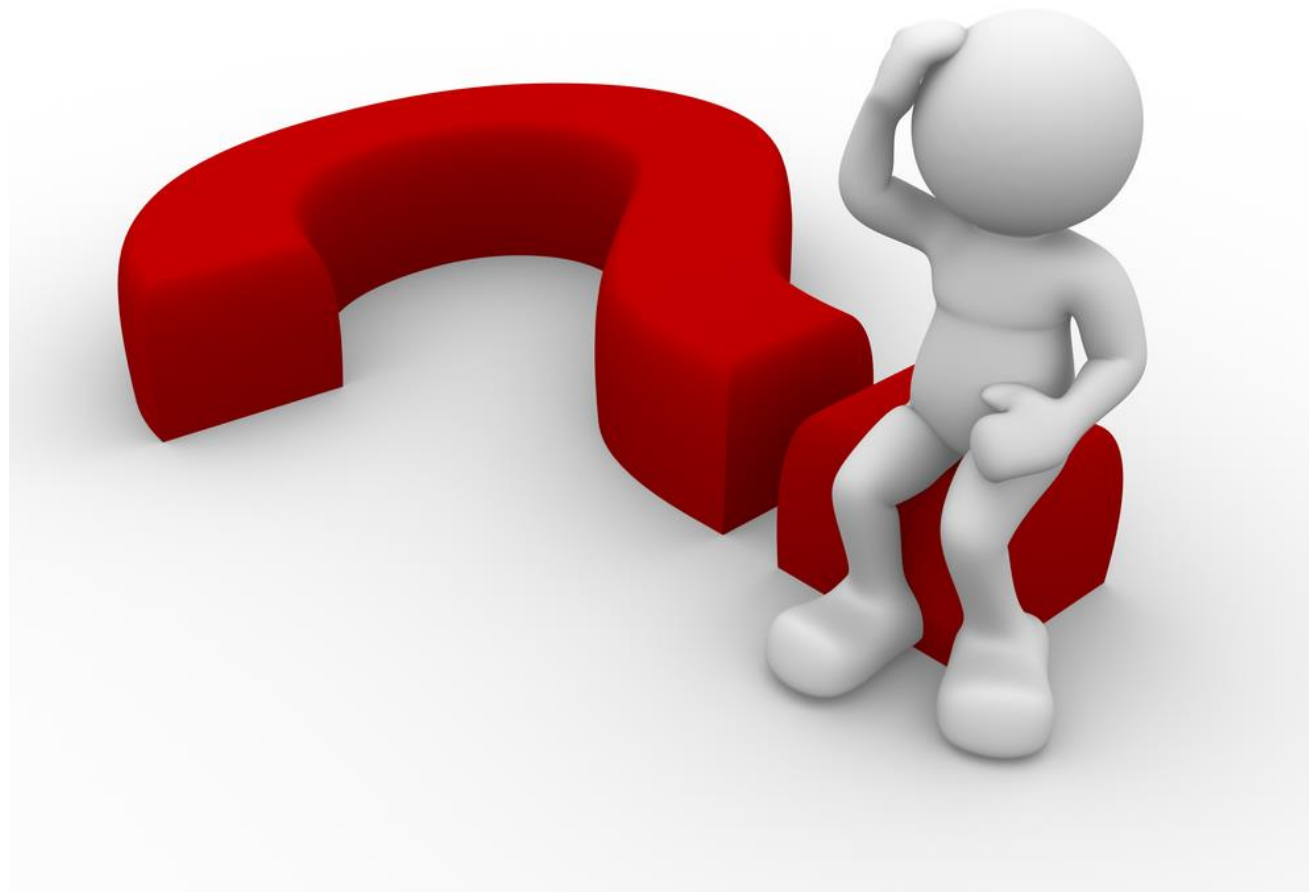 - Identify noisy groups of users (with no clear viewpoint)

➢ **Iterative Rank Difference** (**IRD**) for understanding the viewpoints
  - 1st iteration: what is the viewpoint about
  - Next iterations: what is a particular term about?

➢ Evaluation
  - Proposed viewpoint discovery method outperforms more complex topic models
  - IRD allows getting a deep understanding of a viewpoint and of the related terms
    - *Why did Trump's supporters use the hashtag **#draintheswamp**?*
    - *Why did Clinton's supporters use the hashtag **#trumpleaks?***

Viewpoint Discovery & Understanding in Social networks, WebSci'18, Amsterdam, The Netherlands, May 27-30, 2018.

18/19

Conclusion and Future Works

Future Works:

➤ **Timeline Summarization** of topics and viewpoints
- that will allow understanding how a controversial topic evolves over time and with respect to the involved entities, events and subtopics

➤ **Knowledge graphs** about controversial topics
- How to semantically represent topics and viewpoints to enable advanced query answering?

Viewpoint Discovery & Understanding in Social networks, WebSci'18, Amsterdam, The Netherlands, May 27-30, 2018.

19/19

Thank You

Viewpoint Discovery & Understanding in Social networks, WebSci'18, Amsterdam, The Netherlands, May 27-30, 2018.