Slide 1

CASOS

Exploring pro and anti-government movements during the 2019 Ecuadorian protests

Ramon Villa-Cox

rvillaco@andrew.cmu.edu

School of Computer Science, Carnegie Mellon
Summer Institute 2020

isr institute for SOFTWARE RESEARCH

Carnegie Mellon

Center for Computational Analysis of Social and Organizational Systems
http://www.casos.cs.cmu.edu/

Slide 2

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

2019 Latin American Protests

Rodrigo Buendia/Agence France-Presse — Getty Images

- A series of protests that shocked the region at the end of 2019.
- They started in Haiti, followed by Ecuador, Chile, Bolivia and Colombia.

CASOS

AP Photo/Ariana Cubillos

11 June 2020   **Ramon Villa-Cox**   **2**

<Your Name>

2019 Latin American Protests

- These effectively paralyzed the countries for weeks and in some cases, months.
- They also had a massive online presence and there was reported involvement of international and regional actors that sought to influence the evolution of the different protests.

AP Photo/Ariana Cubillos

2019 Latin American Protests

- We collected Twitter data across the different countries. More than 180 hashtags and terms were used for each countries.
- A special effort was taken to collect conversations around antagonistic positions, by including hashtags that were used by different groups (for and against the different governments).
- During this hands-on session, we are going to focus on a subset of the Ecuadorian data.

<Your Name>

Ecuadorian Protests

- Protests were originated as a response of an International Monetary Fund (IMF) sponsored austerity package which involved a rise in fuel costs.
- Interested parties that fomented the protests included indigenous leaders, student organizations and followers of former president.

Agence France-Presse — Getty Images

11 June 2020   Ramon Villa-Cox   5

Ecuadorian Protests

- Protests occurred from September to October and included violent incidents. The strike caused the paralysis of the economy due to looting and closed highways.
- After two weeks of violent manifestations in several of the main cities of the country, the President agreed with indigenous leaders to cancel the austerity package proposed.

Agence France-Presse — Getty Images

Agence France-Presse — Getty Images
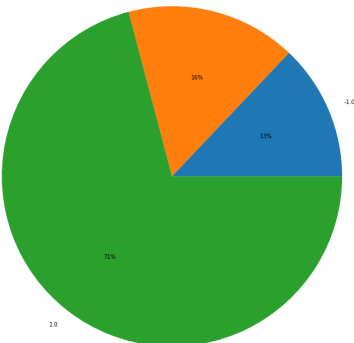
11 June 2020   Ramon Villa-Cox   6

<Your Name>

Determining Pro and Anti Protests Tweets

- 180+ Hashtags and terms were used to collect data around the Ecuadorian protests, from September 20 to October 21 of 2019.
- This resulted in over 11 million tweets from 1.4+ million users.
- Hashtags were classified into either pro, anti or neutral to the protests based on a sample of the tweets.
- This resulted in 64 pro tags and 28 anti (the remainder being either neutral or not a hashtag).

Co-Hashtag Network

<Your Name>

Assigning Stance to Users

- Noisy stance labels were assigned to users based on their usage.
- Users were assigned a label if they only used tags for one side of the argument, either on their tweets or their user descriptions.
- This resulted in a subset of 203990 users.

User Protest Stance Distribution

What we can do with identified stances?

- Contrast Bot presence.
- Consumption of official and alternative news media. This includes Venezuelan and Russian news media.
- Presence of international campaigns seeking to incentivize the riots. There are multiple accounts from Venezuelan origins that were involved in the discussion across multiple countries.
- Interactions within and between groups.
- Construct a classifier to extrapolate the results from these accounts to the rest of the data collected.

<Your Name>

Carnegie Mellon
institute for
SOFTWARE
RESEARCH

PREPROCESSING THE DATA

CASOS

11 June 2020   Ramon Villa-Cox   11

Carnegie Mellon
institute for
SOFTWARE
RESEARCH

Preparing Data for Hands On Session

- The subset of the data shown is too large to use in the present session. This provides the opportunity to review tools available in ORA to work with big data.
- The first thing is to exclude retweets and users that tweeted only one time. This was not done with ORA.
- The following slides show the steps taken in ORA to construct the data that is provided with the lecture.

CASOS

11 June 2020   Ramon Villa-Cox   12

<Your Name>

Step 1: Importing Data

First, we are going to import the raw JSON file to ORA by using the Twitter importer as shown in the figures. By clicking in the derived networks tab, we also deselect networks related to location and words (as we won't use it). Make sure Hashtag x Hashtag – Co-Ocurrence is selected.

Step 2: Selecting the Principal Component

- We are going first select users that are in the main component of the "Agent x Agent – Common Hashtags" network.

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Step 3: Selecting the Principal Component

Then select the giant component and ask ORA to extract all the relevant networks that involve the Agent nodeset.

CASOS

11 June 2020   Ramon Villa-Cox   15

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Step 4: Removing Isolates

- The newly created Meta-Network not only includes the main component of the Common Hashtag network, as it extracts it based on the networks selected at the end.
- We need to remove the remaining isolates.

CASOS

11 June 2020   Ramon Villa-Cox   16

CASOS

<Your Name>

Step 4: Removing Isolates

- We are going to remove isolates based on the following networks:
  - Agent x Agent – Common Hashtags
  - Agent x Agent – All Communication
  - Agent x Hashtag
  - Agent x Tweet – Sender

11 June 2020   Ramon Villa-Cox   17
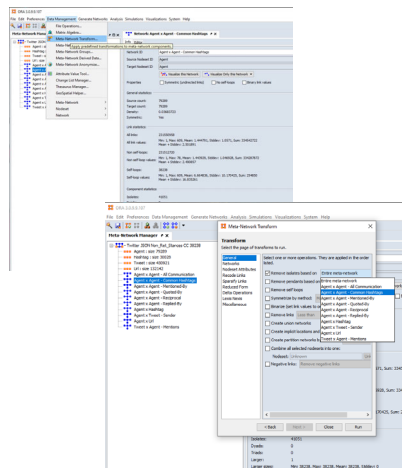
Step 5: Select Maximal K-Core

- Finally, we are going to select the Maximal K-Core of the Common Hashtag network.
- A K-Core of a network is a maximal subgraph were all nodes have at least K connections.

11 June 2020   Ramon Villa-Cox   18
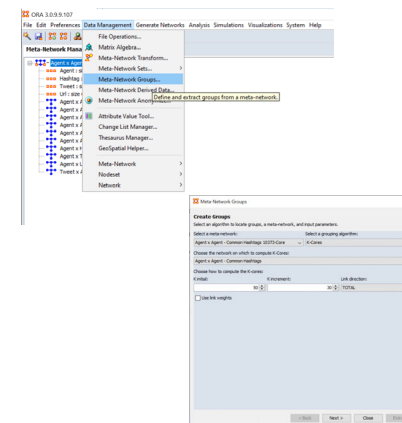
Carnegie Mellon
institute for SOFTWARE RESEARCH

Step 5: Select Maximal K-Core

- This subset is still too big, so we are going to select the Maximal K-Core of the All Communication networks.

CASOS

11 June 2020   Ramon Villa-Cox   19

Carnegie Mellon
institute for SOFTWARE RESEARCH

Step 5: Select Maximal K-Core

- This still includes isolates (as we specified the extraction of all other networks). So we are going to remove the remaining isolates (based on the same networks specified before), resulting in 2000+ agents.

CASOS

11 June 2020   Ramon Villa-Cox   20

CASOS

Carnegie Mellon
institute for
SOFTWARE
RESEARCH

FINDING COMMUNITIES IN THE DATA

CASOS

Carnegie Mellon
institute for
SOFTWARE
RESEARCH

Steps to take

- Open the file StancesEcuador.json. This is a de-identified version of the one constructed in the previous slides. This is done to adhere to Twitter's regulations for sharing collected tweets.
- We are going to import the data and identify the different communities present in the data.
- Then we will contrast them to the observed stances derived for the users.

CASOS

CASOS

<Your Name>

Import Data

- We are going to import the JSON data, making sure that we include the custom attributes included in the JSON specifying the stance of the users.

11 June 2020   Ramon Villa-Cox   23

Import Data

- This imports the extra attribute as shown in the figure. We could also import the data if we have a separate file with the value for the different users.

11 June 2020   Ramon Villa-Cox   24

<Your Name>

Remove Extra-Tweets

- ORA parses the JSON strings taking the ids to users and tweets that were not part of our original sample.
- To maintain the small subset relevant to us, we are going to take again the K-Core of the relevant network.

Remove Extra-Tweets

- Again, given we extract all other networks, the extracted K-core includes a lot of isolates that we need to remove.

Carnegie Mellon
institute for SOFTWARE RESEARCH

Determine User Communities

- There are several ways we can find the communities in the data. First, we can do it by using the visualizer.

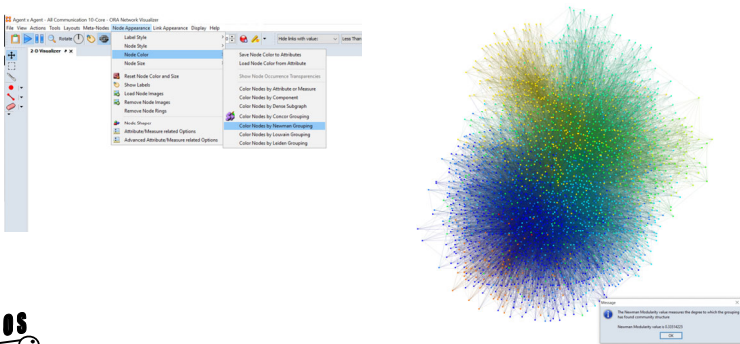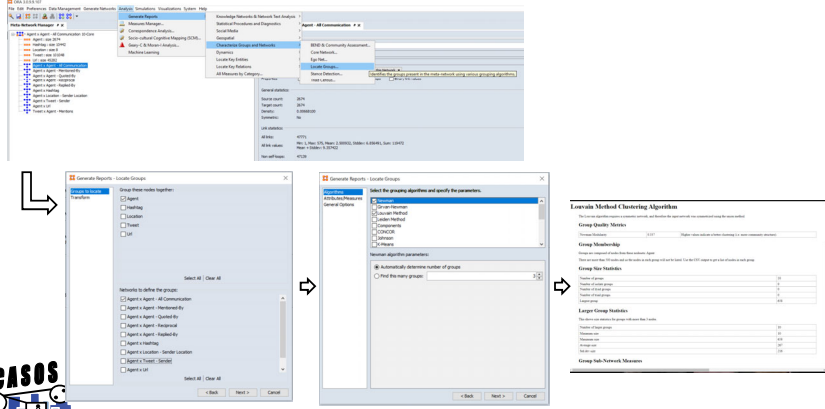CASOS

11 June 2020   Ramon Villa-Cox   27

Carnegie Mellon
institute for SOFTWARE RESEARCH

Determine User Communities

- The previous methods does not create attributes in the nodeset. We can do this by using ORA reports.
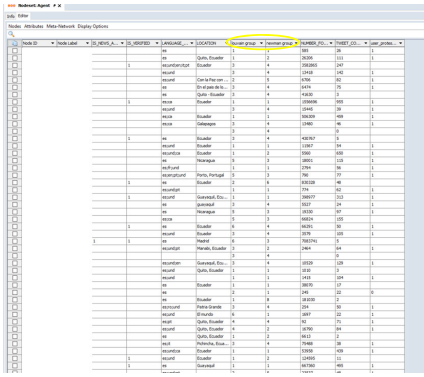
CASOS

11 June 2020   Ramon Villa-Cox   28

<Your Name>

Determine User Communities
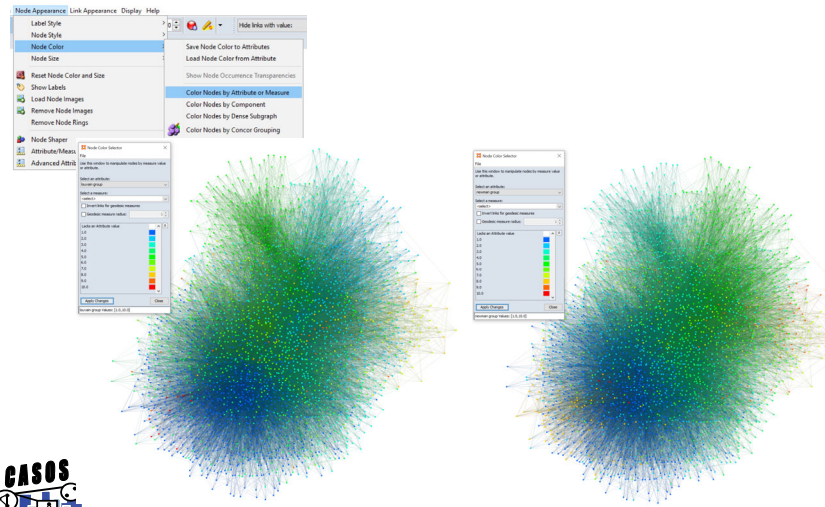
- This creates additional attributes in the agent-set specifying the group membership of an agent.

CASOS

11 June 2020   Ramon Villa-Cox   29

Color Nodes by Group
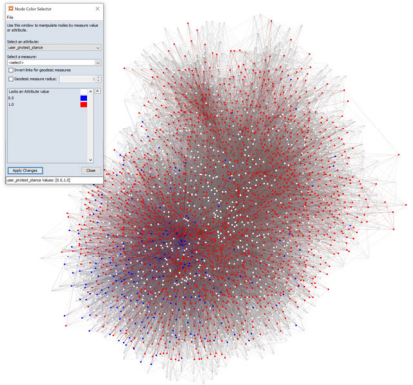
CASOS

11 June 2020   Ramon Villa-Cox   30

<Your Name>

Carnegie Mellon
institute for SOFTWARE RESEARCH

Color nodes by stance

- We see that the users against the protests are also concentrated in one of the groups identified by either algorithm.

CASOS

11 June 2020   Ramon Villa-Cox   31

Carnegie Mellon
institute for SOFTWARE RESEARCH

Discussion

- There is a clear pattern of communication within the anti-protest users. They are grouped together by both community detection algorithms.
- However, they are also grouped with several other pro-protest users.
- This is to be expected as we are not considering the nature of the interactions between the users.

CASOS

11 June 2020   Ramon Villa-Cox   32

<Your Name>

Carnegie Mellon

isr institute for
SOFTWARE
RESEARCH

Discussion

- Part of my research focuses on identifying the stances of those interactions.

- These stances can not only inform community detection algorithms, but they can be predictors of how tweets diffuse within the different communities.

CASOS

CASOS