

First results of the dataset analysis

Decentralised Citizens ENgagement Technologies

Specific Targeted Research Project Collective Awareness Platforms



Creative Commons
Attribution-NonCommercial-
ShareAlike 4.0 International
License



FP7 – CAPS
Project no. 610349
D-CENT
Decentralised Citizens
ENgagement Technologies

Lead beneficiary: UOC / Barcelona Media

D2.2

First results of the dataset analysis

September 2014
Version number: 3

Authors:
Pablo Aragón

Contributor(s):
David Laniado
Andreas Kaltenbrunner (Barcelona Media)
Miguel Aguilera (Universidad de Zaragoza)

Editors and reviewers:
Joonas Pekkane
Marco Sachy
Stefano Lucarelli
Francesca Bria
Kelly Armstrong



The work leading to this publication has received funding from the European Union's Seventh Framework Programme (FP7/2007 – 2013) under grant agreement n° 610349.

The content of this report reflects only the author's view and that the Commission is not responsible for any use that may be made of the information it contains.

Contents

Contents	2
1. Introduction	3
1.1 Networked citizen movements.....	3
1.2 Social Media and Data Analysis.....	4
2. Description of the datasets	6
2.1 Dataset 1: Collective Action Facebook Fan Pages.....	6
2.1.1 Data extraction	7
2.1.2 Specification of the data	8
2.1.3 Description of the fan pages	11
2.2 Dataset 2: Partido X Multilayer Organization.....	22
2.2.1 Facebook Layer	24
2.2.2 Twitter Layer.....	24
2.2.3 Forum Layer.....	25
2.2.4 Mailing list Layer.....	27
3. Preliminary data analysis results	28
3.1 Dataset 1: Collective Action Facebook Fan Pages.....	28
3.1.1 Distribution of documents over time.....	29
3.1.2 Distribution of activity per user.....	32
3.1.3 Overlap of users among fan pages.....	34
3.1.4 Social interaction networks.....	35
3.2 Dataset 2: Partido X Multilayer Organisation.....	38
3.2.1 Distribution of documents over time.....	38
3.2.2 Distribution of activity per user.....	39
4. Conclusions.....	40
5. Bibliography.....	42

1. Introduction

This deliverable describes the datasets that will be used in D-CENT Work Package 2: “Network driven data analysis, modelling and visualization”. The main goal of this Work Package is to study new organizational models from emerging citizen movements to provide inputs for the design and development of the D-CENT platform. At this stage of research, we do not intend to propose a socio-political interpretation of the data analysis results. Further steps will be aimed to characterise the behaviour of connected multitudes of citizens, and collective intelligence self-organisation dynamics, as well as the assessment of the models based on real data, and on the usage of the D-CENT platform. The aim of this document is to offer a description of the data to be used in further steps of analysis, and some preliminary results to characterize the datasets.

1.1 Networked citizen movements

In the recent years, the world has witnessed a new format of social uprisings: the Arab Spring in Tunisia and Egypt (2011), the Spanish 15M movement (2011), Occupy Wall Street in the United States (2011), the Mexican movement #YoSoy132 (2012), the Occupy Gezi protest in Turkey (2013) and the Brazilian Spring (2013). Although there are notable differences between all these protests because of their local economic, political and social contexts; all of them (1) share common claims like the criticism to the democratic deficiencies of political institutions and (2) have exhibited a great usage of social media to organize their inner structures, coordinate their online/offline actions and diffuse multimedia material (see Figure 1).

Many sociologists interested in the Information Age and the Network Society have formulated hypotheses about these recent protests. On the one hand, Castells (2013) describes the protests in Tunisia, Iceland, Egypt, Spain and the United States as the first ‘networked social movements’, characterized by their birth, diffusion and own maintenance on the Internet as well as the occupation of public space. Bennet et al. (2012) consider that the main feature of these movements is a “*change from logic of collective action, associated with high levels of organizational resources and the formation of collective identities, to a logic of connective action, based on personalized content sharing across media networks*”. On the other hand, some authors like Morozov (2011) are sceptical with the hidden costs of the digital forms of activism like “*loss of coherence, morality or even sustainability of the opposition movement*”. One year before the emergence of these uprisings, Gladwell (2010) stated that “*the revolution will not be tweeted*” and imagined a wave of digital protesters connected through weak ties (Granovetter, 1973) but unlikely able to lead to high risk activism. However, although it might be still early to evaluate the depth and real impact of these movements, notable effects are already visible in their local countries like the regime changes in Tunisia and Egypt or the emergence of new networked political parties in Spain. Consequently, social media is playing a key role in current politics and social movements and we can already state, as Lotan et al. (2011) pointed in their empirical study of the Tunisian and Egyptian protests, that “*the revolutions were tweeted*”.



Figure 1: Image from #Yosoy132 movement that praises the role played by social media.

1.2 Social Media and Data Analysis

Social media has become an extraordinary repository of digital traces to conduct quantitative data analysis in order to understand patterns and trends of human and social behaviour. For example:

- the Milgram's small world experiment (Milgram, 1967) was successfully assessed in a collaboration graph of actors extracted from the web portal IMDB¹ (Watts et al., 1998)
- the network theory of the Strength of the Weak Ties (Granovetter, 1973) was proved in the diffusion of links on Facebook (Bakshy, 2012) and Twitter (Grabowicz et al., 2012).

Therefore, an important value of social media is the possibility to validate theories on new forms of political expression. At the beginning of the 21st century, Negri et al. (2000) believed in the appearance of “*multitudes who organise resistance to globalisation through networks and substitute collaborative relationships for hierarchical authority*” and Castells (2006) forecasted that the network society would transform political processes under the conditions of the culture of real virtuality.

The emergence of the uprisings mentioned above and the large amount of traces collected from social media have opened the door for scholars to deepen the understanding of collective behaviour in networked citizen movements. The first studies, focused on the Arab Spring, examined the role played by social media (Howard et al., 2011; Lotan et al., 2011; Khondker, 2011; Anderson 2011). One of the first studies of the inner structures of networked social movements was conducted by González-Bailón et al. (2011) on a Twitter dataset of the Spanish #15M movement. The authors performed a k-core decomposition (Seidman, 1983) of the Twitter diffusion network and found interesting insights about the dynamics of recruitment and information diffusion stating that “*a small core of central users is still critical to trigger chains of messages of high orders of magnitude*”. A following comparative study of the Twitter anatomy of the 15M movement and Occupy Wall Street suggested that “*the global connectivity of social media networks is undermined by structural holes that only a minority of users bridge*” (González-Bailón et al., 2013). These authors claimed that their empirical findings, obtained through network research, contradict the theoretical assumptions of “*the logic of connective action*” (Bennett & Segerberg, 2012) and “*networked social movements*” (Castells, 2012). Gerbaudo (2012) also considers that Castells’ and Hardt&Negri’s formulations of the distributed forms of the so-called “*networked social movements*” follow

¹ <http://us.imdb.com>

the discourse among the activists themselves, but refuses to accept this theoretical framework as a fact. Although these two empirical studies provide high valuable insights about recent protests, it is questionable to establish that those empirical findings contradict the theories on networked/connected movements just through network research. For example, the study conducted by Toret et al. (2013) includes a comprehensive network, text and time series analysis and reveals that the so-called “15M network-system” is “*activated by sequences of collective synchronization and the moments of peak intensity take place in the feedback between physical and virtual space, when languages and moods are tuned*”. Consequently, a multidisciplinary data analysis of these new protests will provide a deeper understanding of these movements than just a network analysis of data snapshots from Twitter. In spite of the important role played by social media in these movements, findings from Twitter analysis can be biased by the particular features of this microblogging service. We note that Twitter is a network oriented to the diffusion of contents but new forms of political organization also rely on personal social networks (e.g. Facebook) as well as their own private forums. Moreover, we highlight the critical importance of the chosen strategies to sample and analyse data from social media. A study by González-Bailón et al. (2014) that applied three different sampling strategies to reconstruct networks on Twitter revealed “*a bias in the reconstructed networks that goes unnoticed in most research but that might have important theoretical implications for some of the questions that have been posed to the data in the past*”.

In addition, although the theoretical literature has found qualitative bounds between the cited networked social movements, not enough comparative studies have been conducted. Therefore, we consider essential to perform an empirical and broad analysis with a special emphasis on the most recent and less reviewed revolts: Mexico, Turkey and Brazil. The researchers who are responsible of the D-CENT Work Package: “*Network driven data analysis, modelling and visualization*” also participate actively in the research network DatAnalysis15M². This research network started a first approach to this analysis in the event “*#GlobalRev: 3 years of interconnected revolts*”³ that gathered scholars, journalists and activists from these countries. The event, devised as a workshop+hackathon, concluded with the agreement that, in spite of the uniqueness of each networked movement, all of them share the following features: (1) the central role played by digital communication networks, (2) their global dimension, (3) the existence of common patterns of collective action and (4) the active defence of democracy and freedom of access to information. Further steps consist of validating the hypotheses, research questions and models formulated in this event to deepen into the characterisation of networked social movements. Some of the ideas and hypotheses to be assessed were published in the previous D-CENT deliverable “*D2.1 Socio technical framework on collective intelligence*”⁴. The II DatAnalysis15M meeting⁵ has recently taken place at the Internet Interdisciplinary Institute (IN3) of the Open University of Catalonia (UOC) to address the data analysis to be performed.

In conclusion, we propose a multidisciplinary data analysis of movements of networked citizens on several online platforms in order to compile the essential features of new organizational models. The analysis, modelling and visualization of the data specified in this deliverable will provide valuable insights to drive the development of the D-CENT platform.

² <http://civilsc.net/node/43>

³ <http://civilsc.net/node/59>

⁴ Accessible at: <http://dcentproject.eu/wp-content/uploads/2014/04/D2.1-.pdf>

⁵ <http://civilsc.net/node/100>

2. Description of the datasets

In this section we describe the datasets that will be used in D-CENT Work Package: “*Data Analysis and Data Modelling*” to understand how current communication systems are being used. The datasets are composed of millions of digital traces, mostly messages, from:

- online social networks (Facebook),
- microblogging services (Twitter),
- mailing lists,
- private forums

focused on the organization and discussion of collective actions in social movements emerged in the recent years. Further analysis, modelling and visualization of the data through different quantitative and qualitative techniques will unveil patterns of collective awareness and self-organization dynamics on a large scale.

2.1 Dataset 1: Collective Action Facebook Fan Pages

Facebook is one of the most popular online social networks. It was founded in 2004 and, at the end of January 2014, more than a billion users were active monthly becoming the second most visited domain on Internet according to the ranking updated by Alexa⁶. On Facebook, users are able to publish, share, comment and “like” diverse kinds of content such as posts, external links, pictures or videos.

Commonly, users post content visible only to the circle of users with whom they have a reciprocated “friendship” connection (or with whom they are connected through a shared connection). Therefore, and although users are able to not restrict the visibility of their personal information, an important fraction of data generated on Facebook is only accessible by its author and corresponding direct or indirect friends. Nevertheless, Facebook launched in 2007 topic pages, called “Fan Pages”, which can be created by any user and where users can share and comment content related to businesses, organizations, celebrities, products or ideas that the administrator(s) establish as the topic of the page. Consequently, fan pages become public forums on Facebook.

Recent social movements have taken advantage of the massive usage of Facebook to set fan pages as communication channels to organize actions, share related content, discuss ideas and, ultimately, engage users. Taking the “Arab Spring” as one of the first uprisings of the social media era, Howard et al. (2011) found that social media played a central role in shaping political debates. Although some authors argued that the Egyptian Facebook campaigners were just the modern incarnation of Arab nationalist networks (Anderson, 2011), Lotan et al. (2011) showed that both the Egyptian and the Tunisian uprisings were reported in online social platforms. Other studies confirm the massive usage of Facebook in later uprisings in Spain (Toret et al., 2013), United States (Caren et al., 2011), Mexico (Díaz Alba, 2013), Turkey (Yildirim, 2013) and Brazil (MacKenzie, 2013).

⁶ <http://www.alexa.com/siteinfo/facebook.com>

2.1.1 Data extraction

Data from the fan pages have been collected through the Facebook Graph API. The Graph API is based on the idea of modelling Facebook as a large-scale graph where:

- **nodes** are basic items (users, photos, pages, comments...)
- **edges** are connections between nodes (a user uploads a photo, a page contains a photo, a user posts a comment on a page...)
- **fields** are the information related to nodes and edges.

The API is HTTP-based and, therefore, data is fetched by HTTP requests with the following structure:

- **nodes:** `GET graph.facebook.com/{node-id}`
- **edges:** `GET graph.facebook.com/{node-id}/{edge-name}`
- **fields:** by default, the two previous request formats retrieve all the fields.

As we commented above, some data on Facebook are not fully available and require secure access. The feed of a fan page, publicly available, only demands a user access token generated by a Facebook app, so that the app will read/modify/write the user's data on her/his behalf. The authorization relies on the OAuth 2.0 protocol (see Figure 2) already described in the previous D-CENT deliverable “D4.2 Technical requirements”⁷.

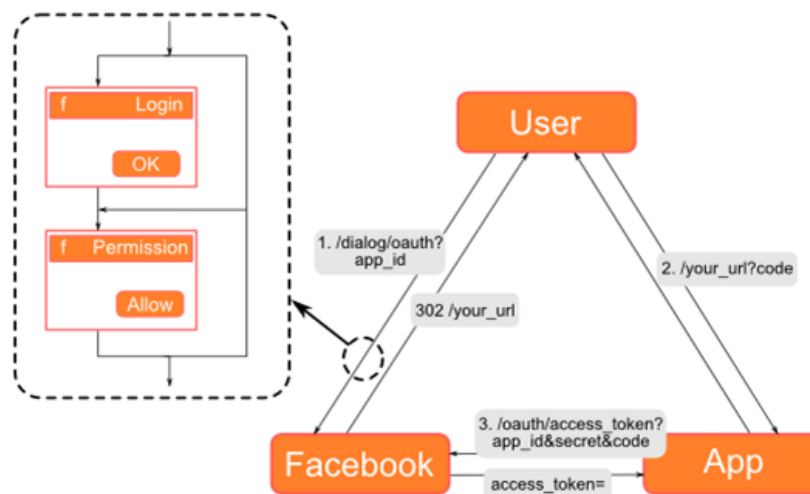


Figure 2: Facebook OAuth scheme

This authorization is performed by a login dialog that, after the validation of the user, returns an access token to be used as:

`GET graph.facebook.com/(...)?(...)&access_token=XXX`

⁷ Accessible at: <http://dcentproject.eu/wp-content/uploads/2014/04/D4.2-.pdf>

In the data collection process of the Collective Action Facebook Fan Pages, we created an app to generate several access tokens and we implemented a project on Java to fetch every post and comment published on the selected fan pages. The request to retrieve a feed of a fan page, taking “Spanish Revolution” page as an example, is the following:

`GET https://graph.facebook.com/v1.0/spanishrevolution/feed/?access_token=XXX`

Fan pages like “Spanish Revolution” contain hundreds of thousands of posts and comments. These large lists of results cannot be requested in a single response because the Facebook Graph API paginates most responses by default. Specifically, responses contain two items: the requested data and pagination metadata that inform the current, previous and next positions of the cursor. Therefore, our system retrieves the newest 25 posts from a fan page and goes forward to the next cursor to extract the 25 preceding posts. This process is performed iteratively until the cursor reaches the first posts published in the fan page. On the other hand, these responses only provide a limited number of comments and likes from each document. Consequently, the system requests the full list of comments and the full list of “likes” for each post to obtain a complete snapshot of the fan pages. During this process, retrieved data are modelled to be stored in a server that will be described in the next subsection.

2.1.2 Specification of the data

The dataset is formed by millions of posts and comments. These documents are stored in a Solr index. Solr⁸ is an open source enterprise search server which wraps the information retrieval library Apache Lucene⁹. This technology has been selected because of its advanced full-text search capabilities to retrieve and analyse documents. Additionally, Solr is linearly scalable and is optimized for high volumes of traffic, which are basic requirements for large datasets as the Collective Action Facebook Fan Pages are.

In Solr, each document (post, comment) is a record composed of several fields that contain the associated information. The fields are described in Table I.

⁸ <http://lucene.apache.org/solr/>

⁹ <http://lucene.apache.org/core/>

Field name	Type	Multivalued	Description
id	String	No	Id of the post/comment on Facebook
application_id	Number	No	Id of the application used to publish the post/comment (optional)
application_name	String	No	Name of the application used to publish the post/comment (optional)
application_namespace	String	No	Namespace of the application used to publish the post/comment (optional)
Caption	String	No	The caption of a link, appears beneath the name (optional)
comments_count	Number	No	Number of received comments
created_time	Date	No	The time the post/comment was initially published
depth	Number	No	If post 0 If comment 1
Description	Text	No	A description of a link in the post, appears beneath the caption (optional)
from_category	String	Yes	Category of the profile that posted the message
from_category_list_id	Number	Yes	Id of the categories of the profile that posted the message
from_category_list_name	String	Yes	Names of the categories of the profile that posted the message
from_id	Number	No	Id of the profile that posted the message
from_name	String	No	Name of the the profile that posted the message
Icon	String	No	A link to an icon representing the type of this post (optional)
in_reply_id	String	No	If comment, id of the post it replies to
in_reply_to_id	String	No	If comment, id of the author of the post it replies to
in_reply_to_name	String	No	If comment, name of the author of the post it replies to
likes_count	Number	No	Count of the users who liked the post/comment
likes_id	String	Yes	Id of the users who liked the post/comment
likes_name	String	Yes	Name of the users who liked the post/comment
Link	String	No	The link attached to this post (optional)
message	Text	No	The status message in the post/comment
message_tags_id	Number	Yes	Id of the profiles tagged in message.
message_tags_length	Number	Yes	Length of the profiles tagged in message.
message_tags_name	String	Yes	Name of the profiles tagged in message.
message_tags_offset	Number	Yes	Offset of the profiles tagged in message.
message_tags_type	String	Yes	Type of the profiles tagged in message.

name	Text	No	The text used in the tag (optional)
object_id	Number	No	The ID of any uploaded photo or video attached to the post (optional)
Picture	String	No	The picture scraped from any link included with the post (optional)
properties_href	String	Yes	HREF for any attached video, for example, the length of the video (optional)
properties_name	String	Yes	Name for any attached video, for example, the length of the video (optional)
properties_text	Text	Yes	Text for any attached video, for example, the length of the video (optional)
shares_count	Number	No	If post, count of shares
Source	String	No	If post, a URL to any Flash movie or video file attached to the post.
status_type	String	No	Description of the type of a status update mobile_status_update, created_note, added_photos, added_video, shared_story, created_group, created_event, wall_post, app_created_story, published_story, tagged_in_photo, approved_friend}
story	Text	No	Text from stories not intentionally generated by users, such as those generated when two people become friends, or when someone else posts on the person's wall.
thread_id	String	No	If post, Id If comment, Id of the post it replies to
fanpage	String	No	Fanpage the post/comments belongs
to_data_category	String	Yes	Category of the profiles mentioned or targeted in this post.
to_data_category_list_id	Number	Yes	Id of the categories of the profiles mentioned or targeted in this post.
to_data_category_list_name	String	Yes	Name of the categories of the profiles mentioned or targeted in this post.
to_data_id	Number	Yes	Id of the profiles mentioned or targeted in this post.
to_data_name	String	Yes	Name of the profiles mentioned or targeted in this post.
to_data_namespace	String	Yes	Namespace of the profiles mentioned or targeted in this post.
Type	String	No	A string indicating the object type if post {link, status, photo, video}
updated_time	Date	No	The time of the last change to this post, or the comments on it.
url	String	No	URL to the post/comment
Text	Text	Yes	Concatenation of the fields: description, message, name, properties_text, story

Table 1: Fields of the Dataset 1: Collective Action Facebook Fan Pages

2.1.3 Description of the fan pages

In this subsection we describe the fan pages about Collective Action. We selected 2 relevant fan pages for each country to be analysed: Spain, Mexico, Turkey and Brazil. Despite the concrete characteristics of each fan page, in all of them users discuss in threads started by messages, links or multimedia content related to information about the corresponding networked social movement (e.g. news, calls for events... etc.). The fan pages are described through the official information published by their administrators. This information is updated to September 2014 and, when needed, texts are translated to English in the right column.

Democracia Real Ya! Association (former fan page of Democracia Real Ya! Platform)



Figure 3: Facebook Cover of Democracia Real Ya! Association

- Name: Democracia real YA
- Id: 116291108447508
- Category: Political organization
- Link: <https://www.facebook.com/AsociacionDRY>
- Website: <http://www.asociaciondry.org>
- Likes: 555 308

About

¡Únete!

- SIMPATIZANTES: <http://bit.ly/IIYOzAG>
- SOCIOS: <http://bit.ly/IipUIwj>
- Colectivos: <http://bit.ly/U6B9UR>

Join us!

- SUPPORTERS: <http://bit.ly/IIYOzAG>
- PARTNERS: <http://bit.ly/IipUIwj>
- Collectives: <http://bit.ly/U6B9UR>

Description

Asociación DEMOCRACIA REAL YA.
INSCRITA EN EL REGISTRO NACIONAL DE
ASOCIACIONES: GRUPO 1º / SECCIÓN 1ª / NÚMERO
NACIONAL: 602263
CIF: G86489291

REAL DEMOCRACY NOW Association.
REGISTERED IN THE NATIONAL REGISTER
OF ASSOCIATIONS: GROUP 1 / SECTION 1 /
NATIONAL NUMBER: 602 263
CIF: G86489291

¡España para los ciudadanos, y no para los mercados!
El #I5M cambiaste tu forma de pensar... ¡Ahora cambia
tu forma de actuar!

Acusamos a los poderes políticos y económicos de
nuestra precaria situación y exigimos un cambio de
rumbo.

Porque no somos mercancía en manos de políticos y
banqueros:

Unidos, podemos.

¡Síguenos en twitter! @Asocdryeventos

MÁS INFORMACIÓN:

► ÚNETE: http://www.asociaciondry.org/?page_id=1240

► ADHESIÓN DE COLECTIVOS: <http://bit.ly/U6B9UR>

► ¿QUIERES SER SOCIO? ¡APÚNTATE!

<http://asociaciondry.info/socios.aspx>

MÁS INFORMACIÓN:

► MANIFIESTO:

http://www.asociaciondry.org/?page_id=17

► DECLARACIÓN DE INTENCIONES DE LA
ASOCIACIÓN "DEMOCRACIA REAL YA!":

http://www.asociaciondry.org/?page_id=213

► VERSIÓN SINTÉTICA DE NUESTROS OBJETIVOS:

http://www.asociaciondry.org/?page_id=74

► PROGRAMA DESARROLLADO:

<http://www.asociaciondry.org/wp-content/uploads/2012/11/Asociaci%C3%B3n-DRY-Programa-desarrollado-y-objetivos-pol%C3%ADticos.pdf>

Escríbenos: info@asociaciondry.org

Spain for citizens, not for the markets!

The # I5M changed your way of thinking ... Now
change the way you act!

We accuse the political and economic powers of
our precarious situation and we demand a
change of course.

Because we are not merchandise in the hands of
politicians and bankers:

Joined, we can.

Follow us on twitter! @Asocdryeventos

MORE INFORMATION:

► JOIN US:

http://www.asociaciondry.org/?page_id=1240

► ACCESS FOR COLLECTIVES:

<http://bit.ly/U6B9UR>

► WANT TO BE A MEMBER? SIGN-UP!

<http://asociaciondry.info/socios.aspx>

MORE INFORMATION:

► MANIFESTO:

http://www.asociaciondry.org/?page_id=17

► MISSION STATEMENT OF THE
ASSOCIATION "REAL DEMOCRACY NOW":

http://www.asociaciondry.org/?page_id=213

► SYNTHETIC VERSION OF OUR GOALS:

http://www.asociaciondry.org/?page_id=74

► PROGRAM DEVELOPED:

<http://www.asociaciondry.org/wp-content/uploads/2012/11/Asociaci%C3%B3n-DRY-Programa-desarrollado-y-objetivos-pol%C3%ADticos.pdf>

Write to: info@asociaciondry.org

Spanish Revolution



Figure 4: Facebook Cover of Spanish Revolution

- Name: Spanish Revolution
- Id: 110177082404435
- Category: Community
- Link: <https://www.facebook.com/SpanishRevolution>
- Website: <http://tomalaplaza.net/>
- Likes: 267 618

About

<http://tomalaplaza.net/> también en: <https://www.rebelmouse.com/spanishrevolution/>

Description

WHO ARE WE?

We are individuals who have come together freely and voluntarily. Each of us has decided, after the concentrations on Sunday, May 15, that we are determined to continue fighting for dignity and political and social awareness.

We do not represent any political party or association.

We are joined by the singular cause of change.

We are brought together by integrity and solidarity with those who are unable to join us.

WHY ARE WE HERE?

We are here because we desire a new society that puts lives above political and economic interests. We demand a change in society and an increase in social awareness. We are here to make it known that the people have not fallen asleep, and we will continue fighting...peacefully.

We send our support to the friends that have been detained for participation in these concentrations, and we demand their immediate release with no criminal charges.

We want all of this, and we want it now. If you are with us, come join us.

Más de 131



Figure 5: Facebook Cover of Más de 131

- Name: Más de 131
- Id: 238795786225832
- Category: Media/news/publishing
- Link: <https://www.facebook.com/masde131>
- Website: <http://www.masde131.com>
- Likes: 27 578

About

Twitter: [www.twitter.com/masde131](https://twitter.com/masde131)

Description

En el primer momento fuimos un grito, fuimos rabia, fuimos indignación. Después, fuimos el color de la sangre en nuestros rostros, en nuestra ropa, en la fuente de nuestra universidad. Y fuimos también una persecución, los gritos fuera de un baño, cada una de las mujeres asesinadas en el Estado de México, la memoria contra el salinato y, finalmente, fuimos Atenco. Después nos señalaron, y fuimos porros, y fuimos pagados, y fuimos pseudoestudiantes. Y entonces, como respuesta, fuimos un video de másde131 estudiantes no entrenados por nadie para decir lo que somos y lo que opinamos.

At first we were a yell, we were angry, we were indignant. Then went the color of the blood in our faces, our clothes, in the fountain of our university. And we also were a chase, yells out of a bath, each of the women murdered in the State of Mexico, memory against salinato and finally we were Atenco. After we pointed out, and we were porros, paid, and pseudo-students. And then, in response, we were a video by másde131 students, not trained by anyone, to say who we are and what we believe.

Esto no es nuevo. Que nadie se sorprenda, que a nadie le extrañe nuestro despertar. De entre las sombras de la historia, la otra historia, ya latía esta rebelión. En nosotros está la conciencia crítica de quienes luchan en este país desde hace siglos por la educación, por conservar nuestra raíz, por la fe en el ser humano.

Defendimos los derechos humanos, caminamos con toda la humildad del corazón y el pensamiento, todos aquellos lugares que con dignidad caminan para transformar nuestro país.

Como colectivo131, actualmente desarrollamos proyectos enfocados en fomentar el reconocimiento de voz, de persona a persona.

1.- Queremos un México que reniegue la indiferencia, buscamos quejas acompañadas de acciones conscientes, de participaciones propositivas por parte de todos los mexicanos.

2.- Mexicanos informados por una verdad plural, no sólo una que reproduzca las voces tradicionales y de unos pocos. "La verdad nos hará libres"

3.- Queremos, y vamos por una democracia auténtica.

This is not new. Let no one be surprised that anyone would miss our awakening. From the shadows of history, the other story, and beat this rebellion. We keep the critical conscience of those who struggle in this country, during centuries, in order to preserve the education, our roots and faith in humans.

We defended human rights, walked with all humility, all those places to walk with dignity and to transform our country.

As colectivo131, we currently develop projects aimed to promote speech recognition, person to person.

1.- We want Mexico to deny the indifference, we seek complaints followed by conscious actions of purposeful holdings by all Mexicans.

2. Informed mexicans for a plural truth, not one that plays traditional voices of a just a few. "Truth will set us free"

3. We want and go for a true democracy.

#YoSoy132

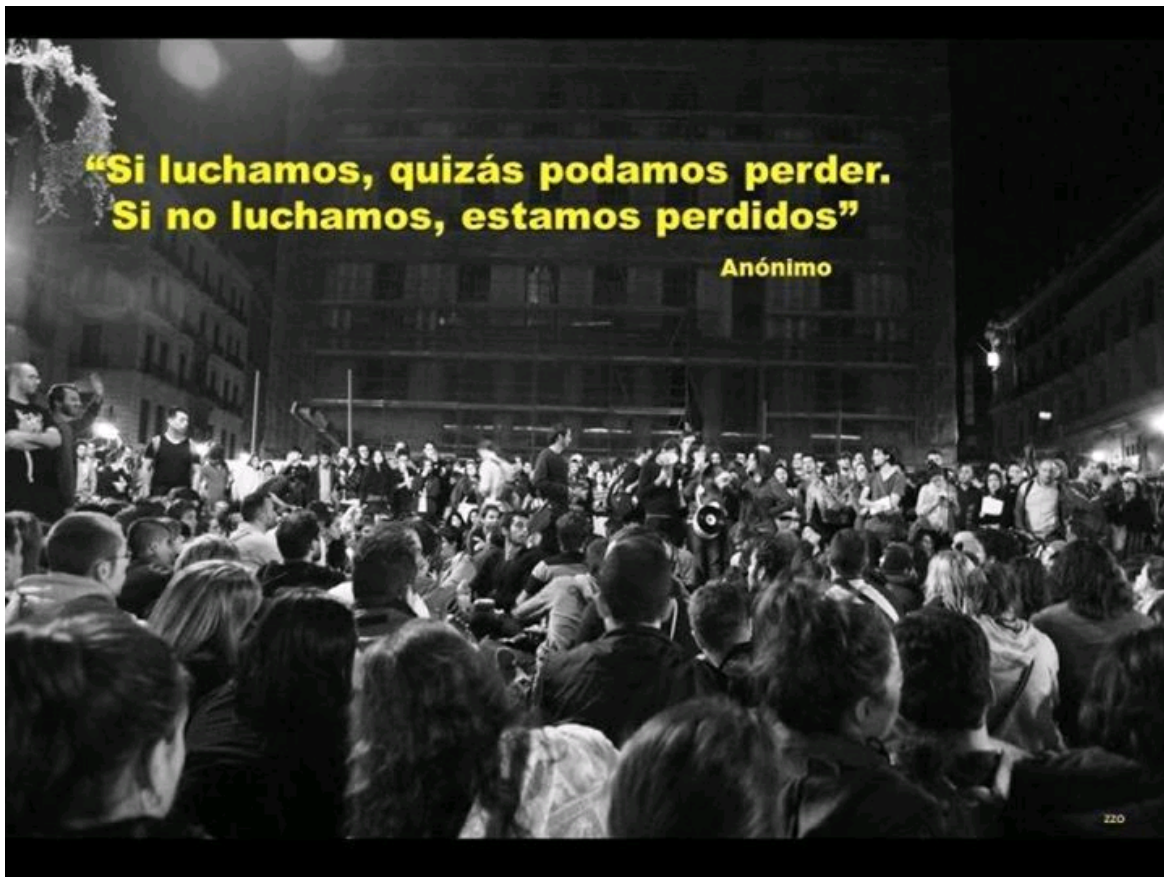


Figure 6: Facebook Cover of #YoSoy132

- Name: #YoSoy132
- Id: 436916642986136
- Category: Community
- Link: <https://www.facebook.com/yosoy132>
- Website: <http://www.yosoy132media.org>
- Likes: 133 388

About

Fan Page del movimiento social mexicano de origen estudiantil #YoSoy132

Fan Page of student Mexican social movement #YoSoy132

Diren Gezi Park

Figure 7: Facebook Cover of Diren Gezi Parkı

- Name: Diren Gezi Parkı
- Id: 390109394440808
- Category: Community
- Link: <https://www.facebook.com/geziparkidirenisi>
- Likes: 640 525

About

Gezi Parkı protestoları ve sonrasındaki süreçte ana akım medyada yer verilmeyen konuları paylaşarak bilgilendirir, kamuoyu oluşturur.

It shares subjects related to the protests of Gezi park that are not presented in the mainstream media during and after the uprising and creates public opinion.

Description

Bu sayfanın hiçbir siyasi, dini, ekonomik ya da etnik grupla bağlantısı yoktur.

This page has no connections to any political, religious, economic or ethnic group.

Occupy Gezi



Figure 8: Facebook Cover of Occupy Gezi

- Name: Occupy Gezi
- Id: 666357646713067
- Category: Community
- Link: <https://www.facebook.com/OccupyGezi>
- Likes: 64 968

About

This page aims to translate the social media and media coverage about Gezi Parki protests. It's not associated with any organizations or political parties.

Description

This page aims to translate the social media and media coverage about Gezi Parki protests in Turkey since May, 2013. The park became a symbol of free speech, violation of human rights and corruption in Turkey. The page is a volunteer work of independent translators and not associated with any organizations, political parties or unions.

Midia NINJA

Figure 9: Facebook Cover of Midia NINJA

- Name: NINJA
- Id: 164188247072662
- Category: Media/news/publishing
- Link: <https://www.facebook.com/midiaNINJA>
- Website: <http://www.midianinja.com>
- Likes: 302 764

About

WWW.MIDIANINJA.ORG

contato: midianinja@gmail.com

Passe Livre São Paulo



Figure 10: Facebook Cover of Passe Livre São Paulo

- Name: Passe Livre São Paulo
- Id: 176309402425319
- Category: Political organization
- Link: <https://www.facebook.com/passelivresp>
- Website: <http://saopaulo.mpl.org.br>
- Likes: 301 100

About

O Movimento Passe Livre (MPL) é um movimento social autônomo que luta por um transporte verdadeiramente público, sob controle popular e sem catracas.

Movimento Passe Livre (MPL) is an autonomous social movement to fight for a truly public transport under popular control and without turnstiles.

Description

<http://saopaulo.mpl.org>

+ entre em contato: passelivresp@gmail.com

Sobre o MPL

o MPL é um grupo de pessoas comuns que se juntam para discutir e lutar por outro projeto de transporte para a cidade. não somos filiados a nenhum partido ou instituição. o mpl é um movimento social independente e horizontal, o que significa que não temos presidentes, dirigentes, chefes ou secretários, todos têm a mesma voz

<http://saopaulo.mpl.org>

+ Contact: passelivresp@gmail.com

About MPL

MPL is a group of ordinary people who come together to discuss and fight for another transportation project for the city. we are not affiliated to any political party or institution. mpl is an independent and horizontal social movement, which means that we do not have

e poder de decisão dentro dos nossos espaços. nós acreditamos que não devemos esperar por iniciativas e ações de políticos e empresários, e que somente a organização e a iniciativa popular pode conquistar mudanças realmente significativas na sociedade. é o povo, somente ele, que tem o poder e a vontade necessária para mudar as coisas e construir um transporte, uma cidade e mesmo um mundo diferente. isso ficou claro nas revoltas da catraca de 2004 e 2005, quando a população de florianópolis ocupou as ruas desta cidade por semanas, até que o aumento absurdo das tarifas fosse cancelado. pensamos na mudança da sociedade através da mudança na lógica da mobilidade urbana. é por isso que não queremos que os ônibus tenham catracas, que impedem tanta gente de ir e vir em todas as grandes cidades do Brasil. mas sabemos que só isso não basta. além da exclusão pelo transporte, há desigualdades entre brancos e negros, homens e mulheres, ricos e pobres. temos um mundo inteiro para reconstruir! a catraca que o mpl repudia é também simbólica. existem catracas invisíveis por todas as partes, impedindo o acesso pleno aos espaços e serviços. precisamos juntos destruir todas elas. pela luta queremos construir um mundo em que não haja nenhuma catraca!

comece já a agir! discuta essas questões com seus amigos e suas amigas, na sua escola, no trabalho e com a família. leve a questão do transporte para o conselho comunitário do seu bairro, para as assembléias do plano diretor estratégico. você tem todo direito de discordar de um sistema que exclui e discrimina! entre em contato com o mpl, participe e traga suas idéias!

<http://saopaulo.mpl.org.br>

<http://tarifazero.org>

founded

Fórum Social Mundial, 2005

presidents, rulers, heads or secretaries, all we have the same voice and decision-making power within our spaces.

we believe that we should not wait for initiatives and actions from politicians and businessmen, and that only popular organization and initiatives can actually achieve significant changes in the society. the society itself has the power and the will to change things and build a transport, a city and even a different world. it became clear in the turnstiles uprisings in 2004 and 2005, when the population of Florianopolis occupied the streets of this city for weeks, until the absurd increase in tariffs was canceled.

we believe in changing society through the change in the logic of urban mobility. that's why we do not want buses with turnstiles that prevent so many people move in all the major cities of Brazil. but we know that this is not enough. beyond the exclusion for transportation, there are inequalities between blacks and whites, men and women, rich and poor. we have a world to rebuild! the turnstile repudiates by mpl is also symbolic. there are invisible turnstiles by all parties, preventing full access to spaces and services. we need to destroy them all together. we want to build a world in which there is no turnstiles!

start now to act! discuss these issues with your friends, in your school, at work and with family. light the issue of transport for the community council of their district assemblies to plan strategic director. you have every right to disagree with a system that excludes and discriminates! contact the mpl, participate and bring your ideas!

<http://saopaulo.mpl.org.br>

<http://tarifazero.org>

founded

World Social Forum 2005

2.2 Dataset 2: Partido X Multilayer Organization

Partido X, also called *Red Ciudadana Partido X* or *Partido del Futuro*, is a political party launched in January 2013. The party was started by people involved in projects related to the 15M movement and social movements for free culture and free software. Some of these projects, grouped by topic, are the following:

- Justice:
 - 15MpaRato¹⁰: Campaign to call for Rodrigo Rato (former Spanish Minister of the Economy, former Managing Director of the International Monetary Fund (IMF) and former president of Bankia until bankruptcy) to be brought to justice over several charges.
- Energy
 - Estafaluz¹¹: Platform that allowed citizens to save 500 million euros in energy bills.
- Economy:
 - OpEuribor¹²: Initiative against the management of the Eurozone interest rate because of its potential responsibility in the sub-prime mortgage crisis.
 - Plataforma por una Auditoría Ciudadana de la Deuda¹³: Platform for auditing the origin and legal responsibility for the current debt of the State of Spain.
- Democracy Models:
 - Democracia 4.0¹⁴: Initiative to promote participatory democracy through virtually seats held by the citizenship in order to vote via the Internet in the Spanish Congress of Deputies.
- Culture:
 - Xnet¹⁵: Collective for free culture and net neutrality.

The major initiative elaborated by Partido X, inspired by the previous projects is the program “Democracia y punto” (Democracy full stop). This program aims to set the following public policies:

- Citizen Legislative Power (Wikilegislation) and government under citizen control (Wikigovernment)
- Binding referendums
- Real and permanent right to vote
- Transparency in public management

Partido X obtained 100.115 votes (0,64%) in the Spanish elections to the European Parliament held the 25 May 2014. Although other emerging political parties obtained a greater number of votes (PODEMOS obtained 1.245.948 votes, i.e. the 7,97%, and 5 members of the European Parliament), the interest of

¹⁰ <http://15mparato.wordpress.com/>

¹¹ <http://www.estafaluz.com/>

¹² <http://opeuribor.es/>

¹³ <http://auditoriaciudadana.net/>

¹⁴ <http://demo4punto0.net/es/home>

¹⁵ <http://whois--x.net/>

Partido X remains in its innovative internal organization and the different communication channels set to discuss ideas and schedule actions. The website of Partido X describes a multiplex network where each organizational level is focused on different tasks. People can be involved simultaneously in more than a single level (see Figure 11). They key levels are:

- Agenda X: Public newsletter that reports the current activities.
- Support groups: Citizens invited to collaborate in a timely manner.
- Matriz: Main workspace for participants moderately committed to Partido X.
- Kernel: Coordinators highly committed to Partido X.
- Developers: Senior coordinators highly committed to Partido X.

The members designed as coordinators are responsible of general purpose groups and/or local geographical groups.

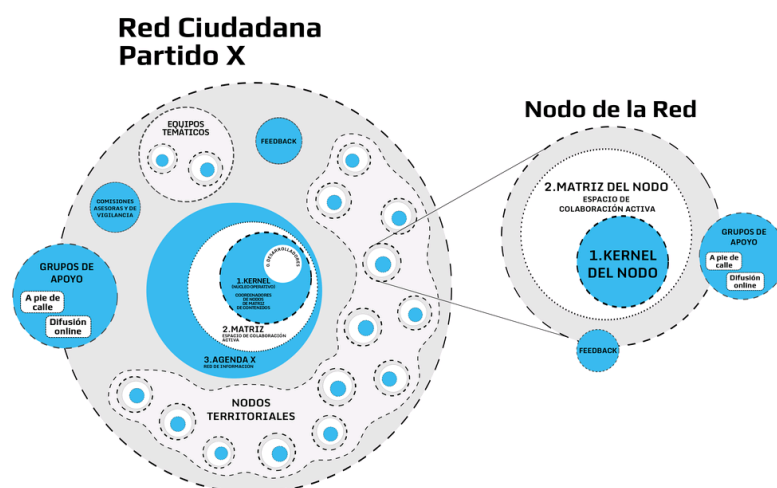


Figure 11: Multilayer Organization of Partido X.

Source: <http://partidox.org/how-does-it-work/>

Partido X is an active actor in the Spanish online politics. Part of the strategies to diffuse contents in the social networks follows the methodologies that some members of the project elaborated in workshops called “Radical Community Manager”. An important number of participants and supporters of Partido X actively discuss ideas and share material in social media. In fact, a comparative study of 5 emerging/new parties of the 2014 European Elections in Spain established Partido X as the one with the highest level of user engagement on Facebook (Bedía et al., 2014).

In comparison to the previous dataset, composed of different networked social movements on the same platform (Facebook), this second dataset comprises digital traces from individuals of the same collective on different platforms. Consequently, the interest of the second dataset is to understand the human dynamics of a group of politically engaged citizens across several communication and organization channels.

2.2.1 Facebook Layer

The Facebook layer consists of a fan page devised to diffuse the contents published in Agenda X, as well as related information. Facebook users subscribed to the fan page are able to post and comment messages. The message of the fan pages have been extracted with the same software tool employed to obtain the first dataset and, therefore, the fields of the each message are the same.



Figure 12: Facebook Cover of Partido X

- Name: Red Ciudadana, Partido X
- Id: 473881175964178
- Category: Political Party
- Link: <https://www.facebook.com/PartidoXPartidodelFuturo>
- Likes: 11 346

About

Para acabar con la corrupción y conseguir #DemocraciayPunto.
Si quieres un resultado diferente, haz algo diferente <http://partidox.org>

To end corruption and get #DemocraciayPunto.
If you want a different result, do something different <http://partidox.org>

Mission

Tenemos un método para cambiar radicalmente los cauces de la democracia y la función de lo que hasta ahora se ha dado llamar "partidos".

We have a method to dramatically change the channels of democracy and the role of what until now has been called "parties."

2.2.2 Twitter Layer

The Twitter layer shares the objectives of the Facebook layer but is characterized by the own features of this microblogging service, essentially, the viral diffusion of ideas. The messages from this layer were collected through the Twitter Streaming API¹⁶. This API provides low latency access to the global stream of tweet data. We implemented a project that connects to the public Streaming API through the endpoint POST statuses/filter. This method is designed to collect tweets in real-time that match, at least, one item from a list of keywords and/or a list of user IDs. During 6 May – 14 June 2014 (interval with a

¹⁶ <https://dev.twitter.com/streaming/overview>

high level of activity because of the European Elections), the system collected messages with the keywords *partido_x*, *conelpartidox*, *yovotopartidox* and the user ID of Partido X.

Each collected tweet is a JSON record composed of the fields defined by the official documentation of the Twitter API¹⁷.



Figure 13: Twitter Cover of Partido X

- Name: Partido X
- Id: 740841738
- Link: https://twitter.com/Partido_X
- Followers: 40 200

Description

Una Red Ciudadana para acabar con la corrupción y conseguir #DemocraciayPunto.
Si quieres un resultado diferente, haz algo diferente <http://www.PartidoX.org/>

A Citizen Network to end corruption and get #DemocraciayPunto.
If you want a different result, do something different <http://www.PartidoX.org/>

2.2.3 Forum Layer

The forum layer corresponds to the forum launched by Partido X to enable a workspace (called Nexa) to the members of Matriz, support groups and local nodes. The forum is an instance of the lightweight Internet forum package Vanilla¹⁸ where users discuss along typical web-forum threads.

¹⁷ <https://dev.twitter.com/overview/api/tweets>

¹⁸ <http://vanillaforums.org/>



Figure 14: Nexo banner of Partido X

Data access will be provided by Partido X in a CSV file where each message is a row composed of the fields defined in Table 2.

Field name	Type	Description
timestamp	String	Date of the message
id	String	Auto-generated id of the message
replied_id	String	If reply, id of the message it replies to
from_id	String	Id of the user that posted the message
from_name	String	Name of the user that posted the message
thread_id	String	Id of the thread that the message belongs
category_id	String	Id of the category that the message belongs
category_name	String	Name of the category that the message belongs

Table 2: Fields of the forum layer from dataset 2: Partido X

2.2.4 Mailing list Layer

The last layer consists of several internal mailing lists used by the following groups:

- Organizational groups:
 - admin-nexo
 - agendax
 - feedbacks
 - kernell
- Topic-based groups
 - analisis-electoral
 - anticorrupcion
 - asuntos
 - Comunicado
 - prensa
 - tech-list
- Geographical-based groups:
 - alicante
 - barcelona
 - cadiz
 - campodegibraltar
 - castellon
 - granada
 - madrid
 - malaga
 - mallorca
 - PartidoX-Zgz
 - sevilla
 - terrassa
 - valencia

Data access will be provided by Partido X in a CSV file where each message is a row composed of the fields defined in Table 3.

Field name	Type	Description
timestamp	String	Date of the message
user	String	E-mail address of the user who sent the mail
group	String	Mailing list that the message comes from (groups listed above)
subject	String	Subject of the message
id	String	Auto-generated id of the message
replied_id	String	If reply, id of the message it replies to

Table 3: Fields of the mailing list layer from dataset 2: Partido X

3. Preliminary data analysis results

In this section we present the results of the preliminary data analysis conducted on the datasets. The results describe temporal, behavioural and social patterns of the data. Essentially, the findings offer an overview on the datasets and provide a clear picture of the collected data. These findings will be taken into account in further analysis conducted in Work Package 2 devised to generate metrics, visualizations and network models to be integrated in the interface of the D-CENT platform.

3.1 Dataset 1: Collective Action Facebook Fan Pages

First, we analyse the dataset that is composed by the selected Facebook fan pages about collective action of networked social movements in Spain, Mexico, Turkey and Brazil (2 fan pages per country). Table 4 shows the number of collected documents for each fan page. The percentages of the type of posts for each fan page are presented in Figure 15. We observe different patterns from the percentages. Some fan pages created after the uprising occurred (“Spanish Revolution”, “Yo Soy 132”, “Occupy Gezi”) tend to use multimedia content (mostly photos) to start threads. In contrast, fan pages of collectives who were actors involved in the birth of the revolts (“Democracia Real Ya!”, “Midia NINJA”, “Passe Livre Sao Paulo”) exhibit a greater usage of statuses and links. The following analysis consists of distributions of content over time, distributions of users over types of activity, the overlap of users among the fan pages and basic metrics of diverse social interaction networks.

fan page	docs	comms.	total posts	status posts	link posts	photo posts	video posts	music posts	question posts	swf posts
asociaciondry	1 258 244	1 030 988	227 256	87 768	76 207	32 010	31 143	119	5	4
spanishrevolution	394 813	264 414	130 399	42 573	46 014	19 169	22 581	48	2	12
mas131	18 370	12 342	6 028	1 796	1 451	1 711	1 064	0	6	0
yosoy132	89 513	59 291	30 222	14 191	6 420	3 697	5 900	7	6	1
geziparkidirenisi	89 673	87 710	1 963	105	690	835	333	0	0	0
occupygezi	18 127	12 692	5 435	1 745	1 729	1 314	647	0	0	0
midianinja	137 921	119 998	17 923	5 020	3 700	7 486	1 717	0	0	0
passelivresp	59 929	57 490	2 439	254	689	1 280	215	1	0	0

Table 4: Number of collected documents per type for each fan page

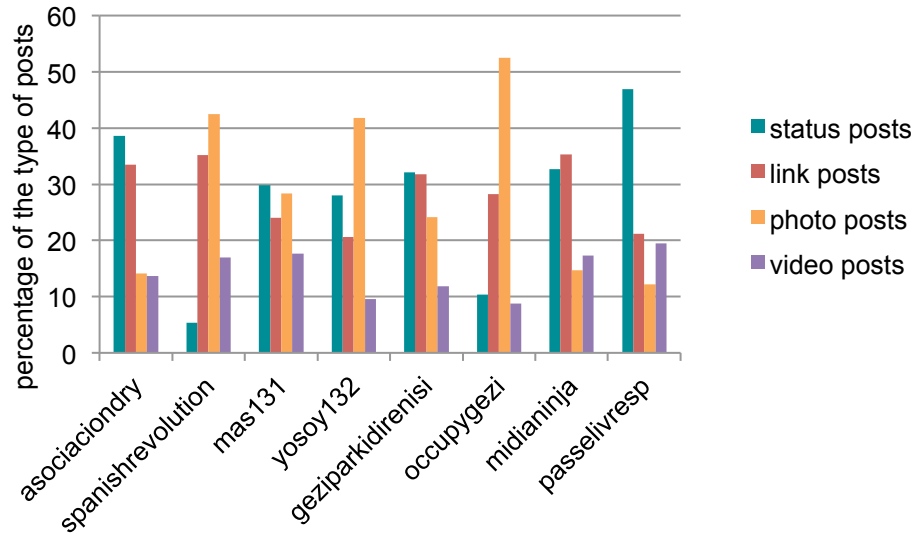


Figure 15: Percentage of the type of post for each fan page (music, questions and swf data are not shown because of their negligible values).

3.1.1 Distribution of documents over time

One of the first steps to understand a dataset formed by human traces is to plot the distribution of those traces along a temporal axis. To this purpose, for each fan page we plot the number of Solr documents (posts and comments) on each day. The starting point is fixed when the first post was published on each fan page and the ending point is the date when the data extraction process ended (31 July 2014). We also plot a similar distribution that distinguishes the documents between posts and comments in two different data series. The resulting figures are contained in Table 5.

At a first glance, we observe that the distributions of documents over time depict different kinds of fan pages according to the life cycle of their inner activity. “Democracia Real Ya! Association”’s fan page was launched in March 2011 to promote the demonstration that took place on May 15, which marked the beginning of the 15M (or *indignados*) movement. The distribution shows a low level of activity before the event, then a peak in the first weeks of the movement, and a progressive decrement of the activity until reaching a stable level. In contrast, the “Spanish Revolution” fan page was launched two days after the demonstration. Consequently, the highest peak occurs during the first days and is followed by a similar decrement until the activity stabilizes. The distributions of posts and comments of these fan pages reveal an interesting finding: although the number of daily posts falls off, the number of daily comments remains stable.

The fan pages “Más 131” and “#YoSoy132” were created after the event considered as the starting point of the Mexican networked social movement: the release of a video¹⁹ by 131 students at the Ibero-American University in response to the incidents with the Mexican Presidential Candidate (currently President) Enrique Peña Nieto. The distributions of documents in both fan pages are similar to the distribution observed in “Spanish Revolution” fan page. However, while the number of daily comments

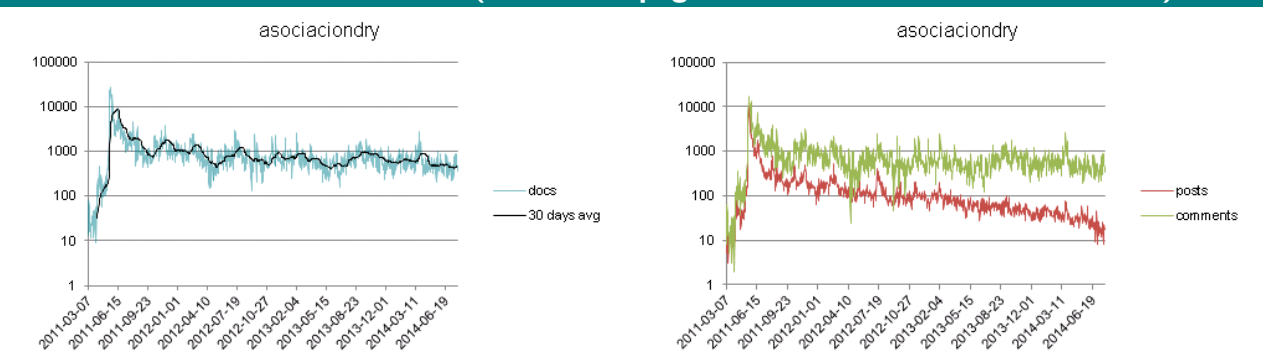
¹⁹ <http://www.youtube.com/watch?v=P7XbocXsFkI>

in “Spanish revolution” remains stable in spite of the decrement of posts per day, the number of daily comments in the Mexican fan pages seems to be more dependent on the number of daily posts.

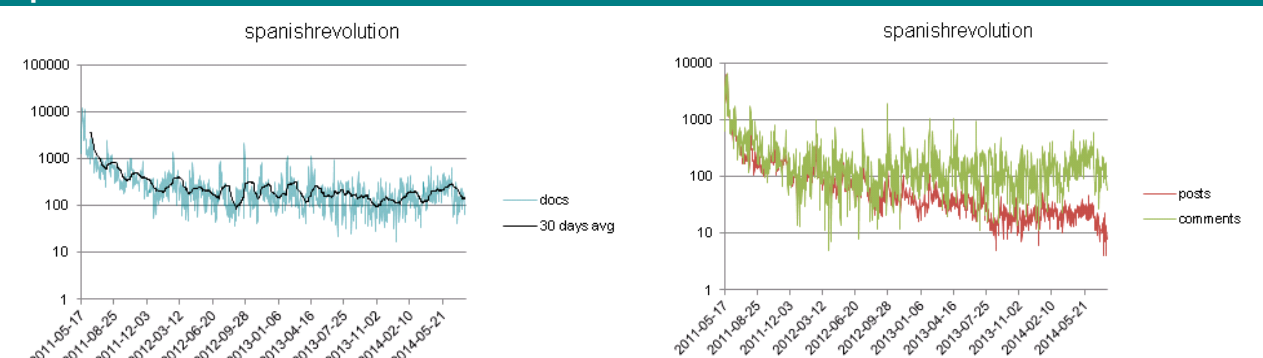
The Turkish fan pages were also created after the birth of the movement: the protests against the urban development plan for Istanbul's Taksim Gezi Park. The distributions are comparable to the ones of the Mexican fan pages but the current activity is residual in most days except for the days related to violent repression occurred in June 2013.

Finally, the selected Brazilian fan pages “midia NINJA” and “Passe Livre Sao Paulo” correspond to collectives born before the Brazilian uprising which acquired high notoriety when the protests aroused. Therefore, the activity is low in the first period and exhibits a remarkable increase; it does not seem to have reached a stable level yet.

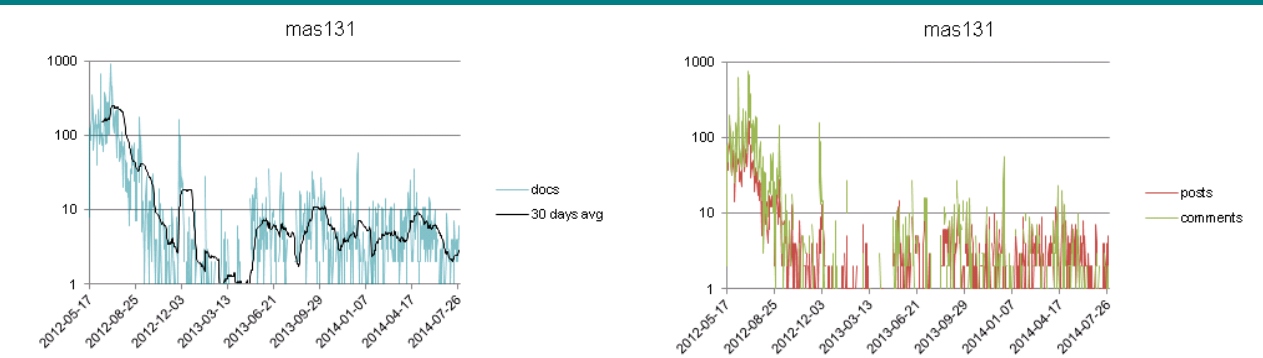
Democracia Real Ya! Association (former fan page of Democracia Real Ya! Platform)



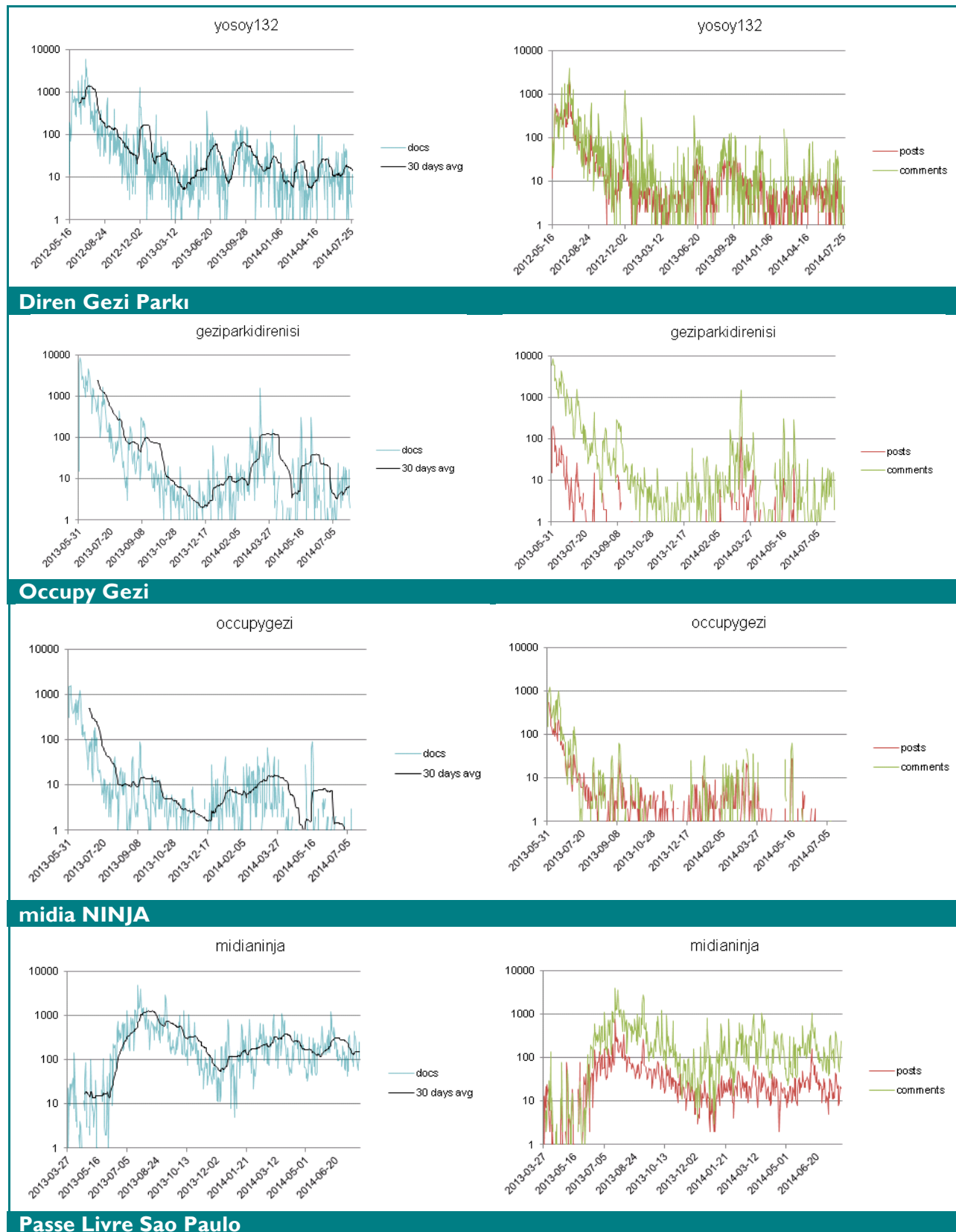
Spanish Revolution



Más 131



#YoSoy132



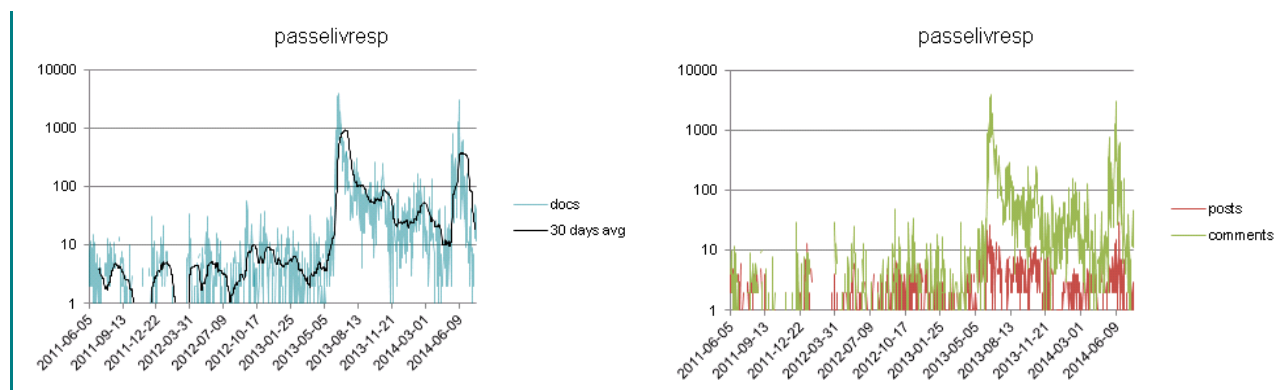


Table 5: Distribution of documents over time (Dataset 1).

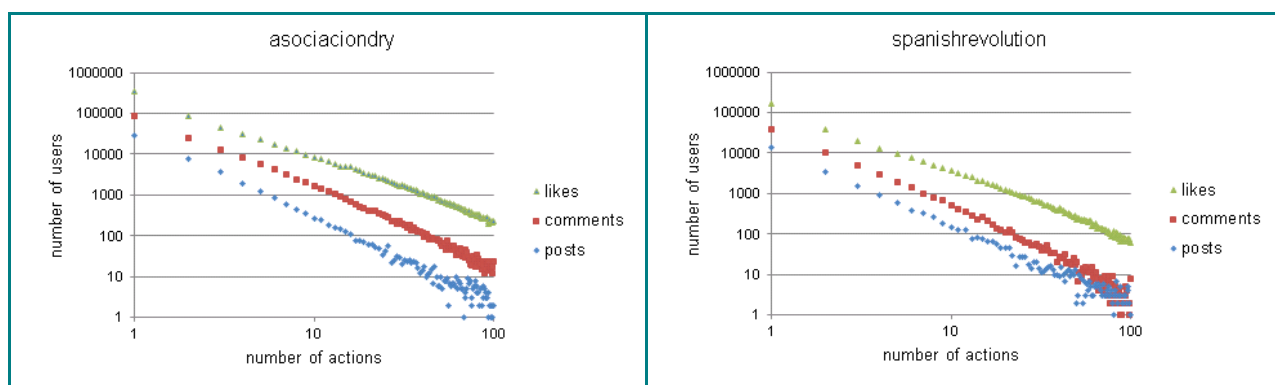
Left figures include a solid black line that corresponds to a moving average trendline (period = 30 days) in order to smoothes out fluctuations in data.

3.1.2 Distribution of activity per user

On Facebook fan pages, users are able to participate through different types of activity. The most common ones are: publishing posts (when allowed), commenting posts and liking posts/comments. To understand the involvement of users in these 3 kinds of activity we plot, for each fan page, the distribution of the number of posts, comments and likes per user. Posting in “Diren Gezi Parkı” and “Passe Livre Sao Paulo” fan pages is forbidden to common users, therefore we omit the time series related to posts. The resulting figures are contained in Table 6

We observe a clear common pattern among the 8 fan pages:

- users are more involved in liking posts/comments than commenting posts
- users are more involved in commenting posts than publishing new posts



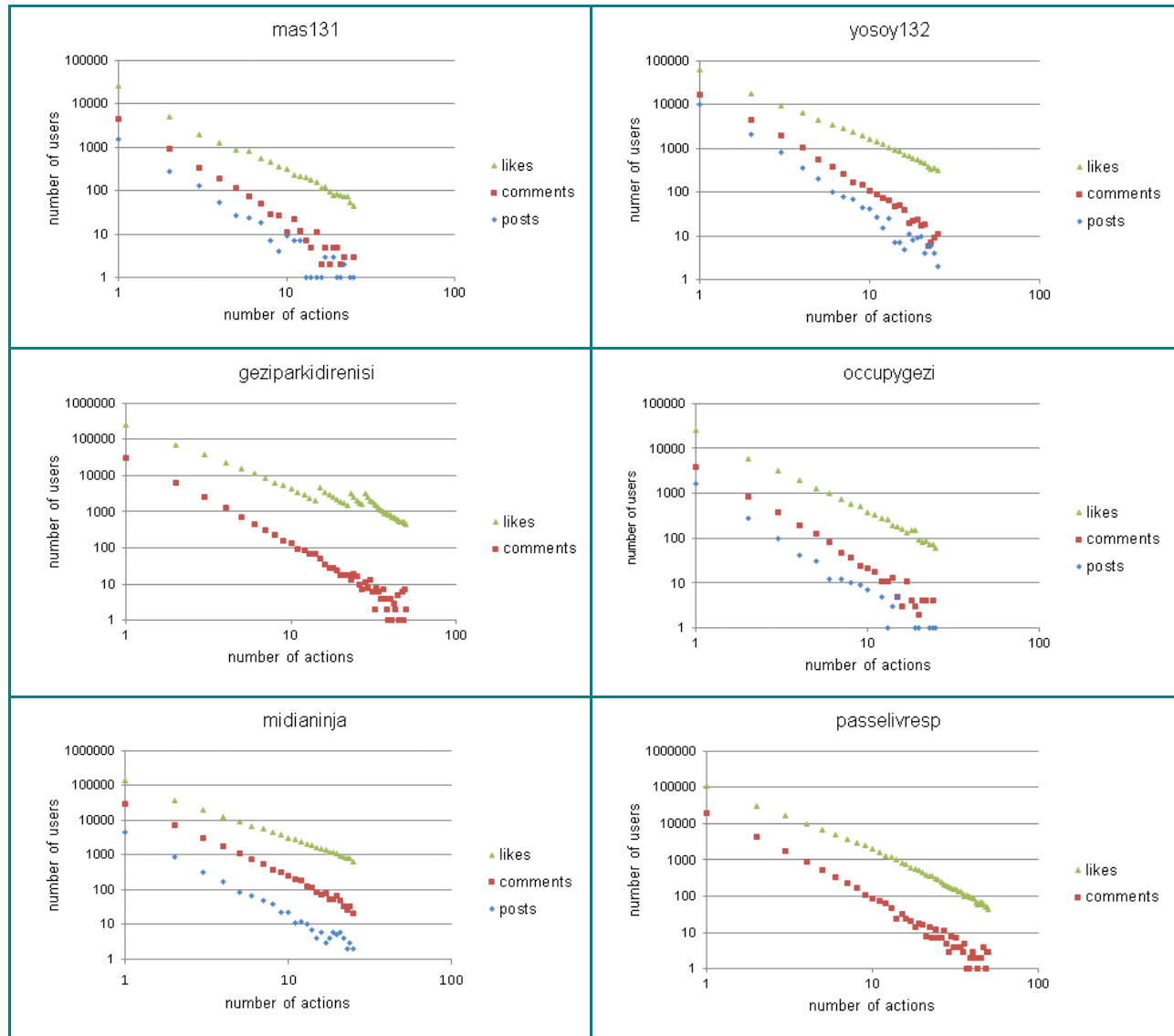


Table 6: Distribution of users over type of activity (Dataset 1)

3.1.3 Overlap of users among fan pages

As we described above, the first dataset is formed by significant fan pages of the networked social movements occurred in Spain, Mexico, Turkey and Brazil. Although the literature establishes qualitative bounds among these movements, we are interested in quantifying the users' overlap as a metric of the new wave of protest.

For this purpose, first we identify the number of fan pages each user in the overall dataset has participated in (through posting, commenting or liking). Table 7 presents the results showing that a great majority of users only participated in one of the selected fan pages, a still remarkable number of users participated in two and just a small amount of users were involved in more than two fan pages. Figure 16 represents a plot of data from Table 7 and reveals an exponential decay of the number of users along the number of fan pages they participated in.

Number of fan pages	Number of users
1	1 946 599
2	208 205
3	2 908
4	403
5	62
6	12
7	1
8	0

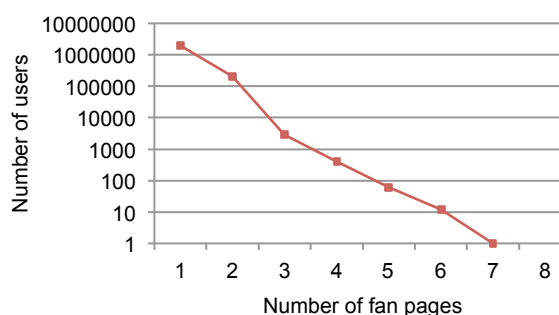


Table 7: Distribution of the number of users over the participation in their selected fan pages

Figure 16: Distribution of the number of users over their participation in the selected fan pages

In Table 7 we present the number of overlapping users between every pair of fan pages. We observe that, as one could expect, the pairs of fan pages with higher overlap are the pairs of fan pages from the same country (values underlined in the table). It is interesting to observe a non negligible overlap also between pages from different countries, such as between the Spanish pages and OccupyGezi or MidiaNinja, indicating that some users were participating in the conversations of more than one movement.

DRY Assoc.								
Sp. Revolution	<u>146221</u>							
Más131	688	418						
YoSoy132	1438	967	<u>12547</u>					
GPDirenisi	749	840	43	83				
OccupyGezi	1169	1293	58	129	<u>12044</u>			
midiaNINJA	1210	1280	81	211	404	661		
PasseLivreSP	463	544	49	94	189	324	<u>35971</u>	
	DRY Assoc.	Sp. Revolution	Más131	YoSoy132	GPDirenisi	OccupyGezi	midianinja	PasseLivreSP

Table 8: Overlap of users between pairs of fan pages (pairs of fan pages from same country are bolded).

The previous result is also visualized in a graph (see Figure 17) in which:

- **nodes** represent fan pages (sized by the number of participants and coloured through a clustering algorithm (Blondel et al., 2008))
- **edges** represent overlapping participants (sized by the number of users who have participated in the corresponding adjacent nodes)

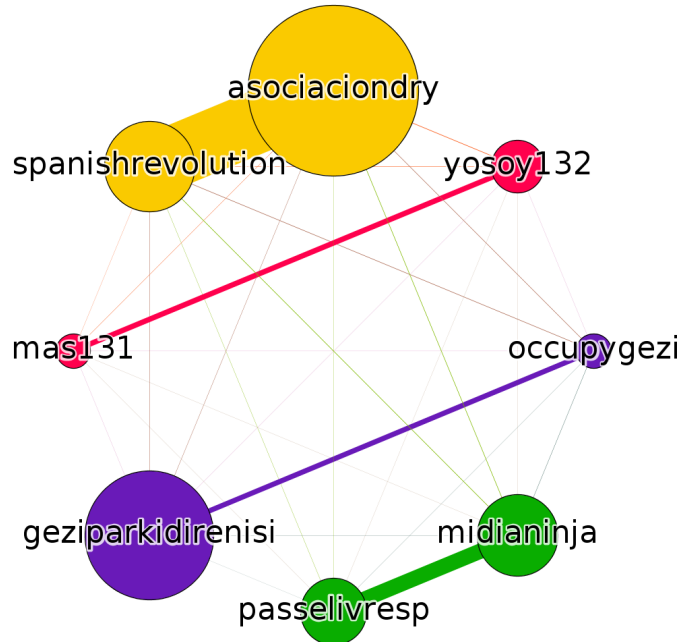


Figure 17: Network of the overlap of users between fan pages.

3.1.4 Social interaction networks

In this subsection we model social behaviour of users by building two kinds of graphs that describe the datasets: (1) the *reply social graph* and (2) the *like social graph*. In short, a graph is an abstract representation of a set of objects (vertices) where some pairs of the objects are connected by links. In our case the objects are Facebook users and user B is linked to user A if B has performed an action (replied to, liked) in response to a content from user A.

Constructions of graphs

In order to analyze discussion behavior in the fan pages of the dataset we generate the *reply social graph*. We denote as $V^{rep} = \{v^{rep}_1, \dots, v^{rep}_n\}$ all users that have replied to or received a reply from another user, at least once, and as E^{rep} the set of edges connecting them. The directed edge e^{rep}_{ij} indicates that user v^{rep}_i has replied to v^{rep}_j . Finally, we assign a weight w^{rep}_{ij} representing the number of times user v^{rep}_i has replied to user v^{rep}_j . The *like social graph* is analogous to the previous one. More formally, we define a graph $G^l = G^l(V^l, E^l)$ comprising a set V^l of vertices and a set E^l of edges. There is an edge e^l_{ij} that connects user v^l_i with v^l_j if user v^l_i has liked some content by user v^l_j . Finally, we assign weight w^l_{ij} to every edge e^l_{ij} , indicating the number of times user v^l_i has liked content by user v^l_j .

Metrics

In this preliminary study of the interaction networks we will analyse the structure of the graphs through the following metrics:

- **Clustering coefficient:** It measures the level of cohesiveness of a graph. The clustering coefficient is defined as the probability that two nodes with a common neighbour are connected. It is equal to the proportion of triangles in the network, divided by the total number of possible triangles.
- **Density and Link reciprocity:** Instead of using the simple ratio between bi- and uni-directional edges which would depend on network size and link density (Garlaschelli et al., 2004) we calculate link reciprocity ρ^* to study the proportion of bidirectional edges in the network. The definition uses the adjacency matrix ($a_{ij}^* = 1$ if there exists a connection from user i to user j , and 0 otherwise) and states:

$$\rho^* = \frac{\sum_{i \neq j} (a_{ij}^* - \hat{a}^*)(a_{ji}^* - \hat{a}^*)}{\sum_{i \neq j} (a_{ij}^* - \hat{a}^*)^2},$$

where \hat{a}^* is the graph density ($\hat{a}^* = \frac{\sum_{i \neq j} a_{ij}^*}{N(N-1)}$).

Results

Table 9 presents the number of nodes and edges and the values of the metrics for the reply social graph. As we expected, the smallest graphs like "Más 131" and "Occupy Gezi" are the densest ones because of the non-trivial dependency of the metric on the network size. However, we note that "Spanish Revolution", the second largest graph, exhibits the most clustered structure. Finally, we also should mention that replies in the Spanish fan pages, the largest ones, are more reciprocal than replies in most of the other smaller fan pages.

fan page	nodes	edges	clustering c.	density	reciprocity
asociaciondry	168 350	206 725	0.000018	0.00001	0.005
spanishrevolution	68 924	97 533	0.000074	0.00002	0.006
mas131	6 412	6 639	0.000032	0.00016	0.014
yosoy132	26 963	28 218	0.000015	0.00004	0.003
geziparkidirenisi	43 981	43 980	0	0.00002	0.000
occupygezi	5 859	5 997	0.000035	0.00017	0.004
midianinja	46 558	47 626	0.000003	0.00002	0.000
passelivresp	27 921	27 920	0	0.00004	0.000

Table 9: Graph metrics of the reply social graphs

Commonly, the posts published by the own fan page are the ones that receive the greatest number of replies. Therefore, we remove that node (and his links) for each fan page and, then, we re-calculate the metrics. There are no values in the reply social graph of "Diren Gezi Parkı" and "Passe Livre Sao Paulo" because of the posting restriction mentioned above. The results are showed in Table 10 and reveal that the communities of Spanish fan pages, in comparison with the rest, are more cohesive and more reciprocal, indicating a more decentralized interaction pattern. However, low values of clustering coefficient are partly due to the way we build the reply network: as we are not able to detect when a

comment is replying to another previous comment, we consider each comment as a reply to the post under which it is situated. This makes that users who publish posts on a fan page acquire a very central position as “stars” in the network, while most users are not connected to one another but just to the stars. Being able to detect relationships between comments might lead to networks with a higher clustering coefficient.

fan page	Nodes	edges	clustering c.	density	reciprocity
asociaciondry	27 634	48 729	0.02809	0.00006	0.015
spanishrevolution	19 207	39 458	0.01758	0.00011	0.012
mas131	1 052	834	0.00044	0.00075	0.001
yosoy132	5 916	5 187	0.00818	0.00015	0.001
geziparkidirensi	0	0	-	-	-
occupygezi	723	588	0.00438	0.00113	0.003
midianinja	4 417	3 680	0.00071	0.00019	0.002
passelivresp	0	0	-	-	-

Table 10: Graph metrics of the reply social graphs after filtering the fan pages nodes

Table 11 presents the number of nodes and edges and the values of the metrics for each *likes social graph*. Again, the Spanish fan pages are the most reciprocal in this graph. However, “Occupy Gezi” is considerably more clustered than the rest.

fan page	nodes	edges	clustering c.	density	reciprocity
asociaciondry	744 203	1 989 227	0,00005	0,00000	0,016
spanishrevolution	328 550	707 815	0,00004	0,00001	0,013
mas131	40 797	52 144	0,00005	0,00003	0,012
yosoy132	133 420	193 441	0,00003	0,00001	0,007
geziparkidirensi	532 791	688 884	0,00000	0,00000	0,001
occupygezi	45 356	69 618	0,00013	0,00003	0,006
midianinja	283 906	528 506	0,00003	0,00001	0,005
passelivresp	208 928	280 954	0,00001	0,00001	0,003

Table 11: Graph metrics of the like social graphs

As we did above, we filter the node (and links) that correspond to the own fan page and re-calculate the metrics (see Table 12). The Spanish fan pages are the ones that obtain the highest level of reciprocity and, after “Passe Livre Sao Paulo”, the most clustered ones. The results also reveal the important role played by the administrator of “Occupy Gezi” fan page in the inner structure.

fan page	nodes	Edges	clustering c.	density	reciprocity
asociaciondry	218 181	1 288 504	0,01631	0,00003	0,022
spanishrevolution	112 187	417 558	0,01286	0,00000	0,020
mas131	12 717	18 337	0,00141	0,00011	0,007
yosoy132	40 304	76 709	0,00639	0,00005	0,011
geziparkidirensi	80 725	182 020	0,00423	0,00003	0,003
occupygezi	14 154	31 225	0,00791	0,00016	0,010
midianinja	102 987	284 705	0,00589	0,00003	0,009
passelivresp	43 565	88 166	0,01925	0,00005	0,007

Table 12: Graph metrics of the like social graphs after filtering the fan pages nodes

3.2 Dataset 2: Partido X Multilayer Organisation

In this subsection, we explore the second dataset about the Multilayer Organisation of Partido X. This deliverable only reports results on Facebook and Twitter because access to data from Forum and Mailing List will be provided after the submission of this document.

layer	docs	comms	total posts	status posts	link posts	photo posts	video posts	question posts	tweets	replies	retweets
facebook	35 517	29 439	6 078	2 718	1 858	1 000	500	2	-	-	-
twitter	80 045	-	-	-	-	-	-	-	12 587	6 696	60 762

Table 13: Table 14: Number of collected documents per type for layer of Partido X

The analysis consists of the previous techniques: the distribution of documents over time and the distribution of activity per user. Matching of users' profiles among the different layers will be performed by actors involved in Partido X who will identify the corresponding user names of each participant across different layers. Full data access will be provided by Partido X through an anonymization process to preserve private and personal information. The analysis of the social interactions networks will be performed, once the data have been provided, through modelling the interactions in different layers as a multilayer network (De Domenico et al., 2013).

3.2.1 Distribution of documents over time

Table 15 shows the distribution of documents, posts and comments over time in the first two layers. On Facebook, the highest peak occurs when the party was launched and, then, the activity decays until the end of the 2013. The number of documents in 2014 increases irregularly until the 2014 European elections and then activity drops drastically after them. On Twitter, data was collected during an interval of time that comprises the electoral campaign. The distribution reveals that there were thousands of messages during the campaign: the highest peak occurs on the Election Day and the activity drops moderately after the elections.

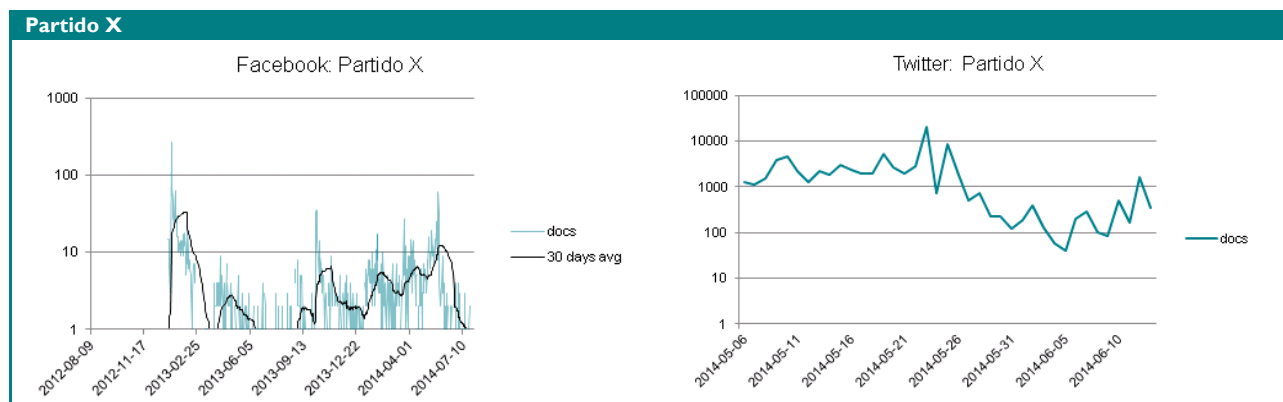


Table 15: Distribution of documents over time (Dataset 2).

The left figure includes a solid black line that corresponds to a moving average trendline (period=30 days) in order to smooths out fluctuations in data.

3.2.2 Distribution of activity per user

On Facebook, we identified as three key activities: publishing posts (when allowed), commenting posts and liking posts/comments. Twitter also establishes 3 main mechanisms to publish messages: posting a tweet, posting a reply to other tweet and retweeting (re-diffusing) another tweet. We plot the distribution of activity per user over the three kinds of activity on Facebook and Twitter analogously to the first dataset (see Table 16). The results on Facebook are similar:

- there are more users involved in liking posts/comments than users commenting posts
- there are more users involved in commenting posts than users publishing new posts

On Twitter:

- there are more users involved in retweeting than publishing tweets/replies
- there is not a clear barrier between publishing tweets or replies to other tweets

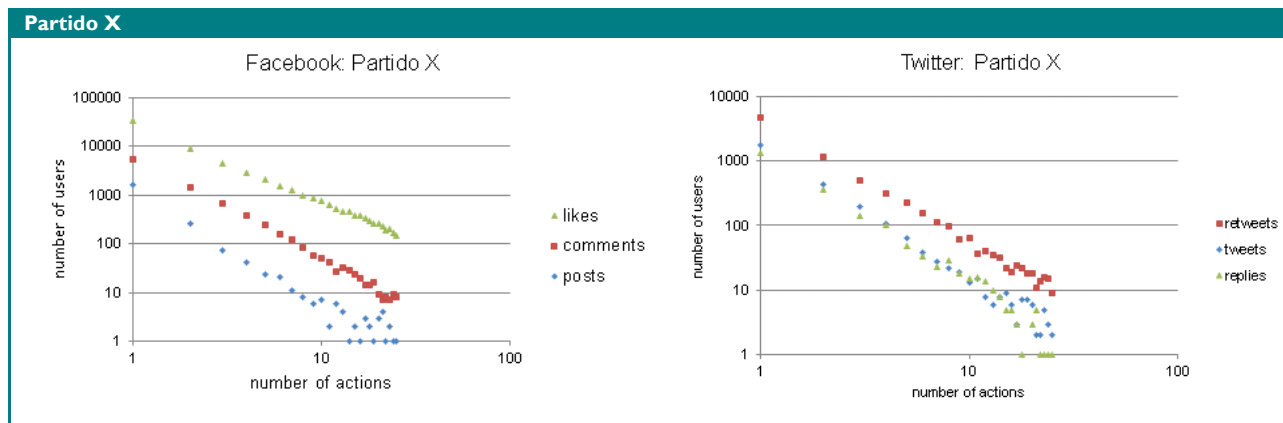


Table 16: Distribution of users over type of activity (Dataset 2)

4. Conclusions

In this document, we have presented:

- a brief review of the state of the art at the intersection between the study of networked social movements and data analysis
- the motivation to deepen into these fields in order to generate new models and visualization that will provide feedback to the D-CENT platform
- a detailed description of two datasets that have been collected to achieve the previous goal
- a preliminary data analysis of the datasets that provides insights which will be discussed below.

First, we have shown the temporal series of posts and comments from Facebook fan pages about communities focused on social collective action. The results describe different type of communities whose peculiarities should be taken into account in further analysis. Some fan pages (e.g. “Democracia Real Ya! Association”, “Passe Livre Sao Paulo”) were generated by collectives that organised an action that evolved into a social revolt and, finally, a networked social movement. This evolution follows the temporal framework: “gestation-explosion-latency” described by Toret et al. (2013). In contrast, other fan pages (e.g. “Spanish Revolution”, “#YoSoy132”, “Occupy Gezi”) were created to establish a communicational channel on Facebook once the uprising had begun. In addition, we have noted the stability of daily comments on Spanish fan pages in spite of the progressive decrement of daily posts. We consider this finding as an interesting metric of stability of a community in comparison to other fan pages whose number of daily comments is highly dependent on the amount of daily posts. Moreover, Spanish likes/reply social graphs present high values of clustering and reciprocity, suggesting a more cohesive and decentralized pattern of interaction.

Secondly, the distribution of activity per user for different kinds of activities reveals that Facebook users tend to participate more actively through likes than replies, and much less publishing new posts. A similar effect is also visible on Twitter where users produce *more* retweets than new messages (tweets or replies to other tweets). Therefore, we note the value of one-click actions (e.g. likes, retweets) in order to improve user engagement. This insight might be taken into account during the design phase of the D-CENT platform. The activities data series, plotted in a log-log diagram, follow a Zipf curve. This result can be easily associated to the “90-9-1 principle” formulated by Nielsen:

“When you plot the amount of activity for each user, the result is a Zipf curve, which shows as a straight line in a log-log diagram. User participation often more or less follows a 90-9-1 rule:

- *90% of users are lurkers (i.e., read or observe, but don't contribute).*
- *9% of users contribute from time to time, but other priorities dominate their time.*
- *1% of users participate a lot and account for most contributions: it can seem as if they don't have lives because they often post just minutes after whatever event they're commenting on occurs.”*

This principle has been validated in other online social platforms (e.g. Wikipedia, YouTube and Flickr) and indicates an evident inequality of participation in online platforms. From this finding, some research questions emerge:

- Are networked social movements ruled by minorities or is this just a reflection of their activity in social media?
- If there is a minority of users responsible for driving a community:
 - Which are the distinguishing characteristics of these users?
 - Which role is played by these users?
 - Is it a stable minority? Does it mute?

Third, we found that (1) most users only participate in one fan page, (2) a still remarkable amount of users participate in two fan pages (mostly from the same country) and (3) a few of users are active in more than two pages, from different countries. However, the overlap of users does not fully describe the mutation and global perspective of networked social movements, as witnessed by Peña-López et al. (2014) in a study of three events of the Spanish 15M movement (explosion, anniversary and the protests in 25 September 2012) on Twitter, which revealed that only the 3.22% of the users from the datasets participated in the three events. We are interested in the exploration of new quantitative approaches (e.g. content analysis) to assess the qualitative bounds, found in the theoretical literature, between the revolts.

In conclusion, we have presented some valuable preliminary results. Further tasks will include analysis, modelling and visualization of the data presented in this document to generate knowledge for the design of the D-CENT platform.

5. Bibliography

- Anderson, L. (2011). Demystifying the Arab Spring: parsing the differences between Tunisia, Egypt, and Libya. *Foreign Aff.*, 90, 2.
- Bakshy, E. (2012) Rethinking information diversity in networks. URL www.facebook.com/notes/facebook-data-team/rethinking-information-diversity-in-networks/
- Bedia, M., & Gómez, T. (2014). Informe sobre el impacto de partidos politicos emergentes EP2014 en redes sociales.
- Bennett, W. L., & Segerberg, A. (2012). The logic of connective action: Digital media and the personalization of contentious politics. *Information, Communication & Society*, 15(5), 739-768.
- Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008.
- Caren, N., & Gaby, S. (2011). Occupy online: Facebook and the spread of Occupy Wall Street.
- Castells, M., & Cardoso, G. (Eds.). (2006). The network society: From knowledge to policy (pp. 3-23). Center for Transatlantic Relations, Paul H. Nitze School of Advanced International Studies, Johns Hopkins University.
- Castells, M. (2013). Networks of outrage and hope: Social movements in the internet age. John Wiley & Sons.
- De Domenico, M., Solé-Ribalta, A., Cozzo, E., Kivelä, M., Moreno, Y., Porter, M. A., ... & Arenas, A. (2013). Mathematical formulation of multilayer networks. *Physical Review X*, 3(4), 041022.
- Díaz Alba, C. (2013). Tres miradas desde el interior de #YoSoy132. *Desacatos*, (42), 233-243.
- Garlaschelli D. & Loffredo M.I. (2004). "Patterns of link reciprocity in directed networks", *Physical Review Letter*, 93(26), p. 268701.
- Gladwell, M. (2010). Small change. *The New Yorker*, 4(2010), 42-49. ISO 690
- González-Bailón, S., Borge-Holthoefer, J., Rivero, A., & Moreno, Y. (2011). The dynamics of protest recruitment through an online network. *Scientific reports*, 1.
- González-Bailón, S., Wang, N., Rivero, A., Borge-Holthoefer, J., & Moreno, Y. (2014). Assessing the bias in samples of large online networks. *Social Networks*, 38, 16-27. ISO 690
- Grabowicz, P. A., Ramasco, J. J., Moro, E., Pujol, J. M., & Eguiluz, V. M. (2012). Social features of online networks: The strength of intermediary ties in online social media. *PloS one*, 7(1), e29358.
- Granovetter, M. S. (1973). The strength of weak ties. *American journal of sociology*, 1360-1380.

Howard, P.N., Duffy, A., Freelon, D., Hussain, M., Mari, W. & Mazaid, M. (2011). *Opening Closed Regimes: What Was the Role of Social Media During the Arab Spring?*. Seattle: PIPTI.

Lotan, G., Graeff, E., Ananny, M., Gaffney, D., & Pearce, I. (2011). The Arab Spring| the revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions. *International Journal of Communication*, 5, 31.

Khondker, H. H. (2011). Role of the new media in the Arab Spring. *Globalizations*, 8(5), 675-679.

MacKenzie, D. (2013). Brazil's uprising points to rise of leaderless networks. *New Scientist*, 218(2923), 9.

Milgram, S. The small world problem. *Psychol. Today* 2, 60–67 (1967).

Morozov, E. (2011). *The Net Delusion: How not to liberate the world*. Penguin UK.

Negri, A., & Hardt, M. (2000). *Empire*. Exils Éditeurs, Paris, 1.

Nielsen J (2006). Participation Inequality: Lurkers vs. Contributors in Internet Communities.
<http://www.nngroup.com/articles/participation-inequality/>

Peña-López, I., Congosto, M., & Aragón, P. (2014). Spanish Indignados and the evolution of the 15M movement on Twitter: towards networked para-institutions.

Seidman, Stephen B. (1983), "Network structure and minimum degree", *Social Networks* 5 (3): 269–287, doi:10.1016/0378-8733(83)90028-X.

Toret, J. (2013). *Tecnopolítica: la potencia de las multitudes conectadas. El sistema red 15M, un nuevo paradigma de la política distribuida*. IN3 Working Paper Series.

Vega, V., & Bertha, C. (2013). Investigación etnográfica virtual de la construcción del ciberactivismo internacional: análisis del movimiento #YoSoy132 a través de Facebook.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *nature*, 393(6684), 440-442.