

FB data collection and analysis - measuring social context from ephemeral populations

Hjalmar Bang Carlsen, SODAS, University of
Copenhagen(hc@sodas.ku.dk)

Joint work with Jonas Toubøl, Tobias Gårdhus and
Snorre Ralund

FB data collection and analysis - measuring social context from ephemeral populations

1. Bringing social context back in with social media group data.

2. Facebook data collection strategy

Hjalmar Bang Carlsen, SODAS, University of Copenhagen(hc@soc.ku.dk)

Joint work with Jonas Toubøl, Tobias Gårdhus and Snorre Ralund

1. Bringing social context back in with social media group data.

The problem

- For the last thirty years, empirical social research has been dominated by the sample survey. But as usually practiced, using random sampling of individuals, the survey is a sociological meatgrinder, tearing the individual from his social context and guaranteeing that nobody in the study interacts with anyone else in it. It is a little like a biologist putting his experimental animals through a hamburger machine and looking at every hundredth cell through a microscope; anatomy and physiology get lost, structure and function disappear, and one is left with cell biology.

Allen H. Barton. 1968. "Bringing Society Back in Survey Research and Macro-Methodology". *The American Behavioral Scientist*. 12(2): P. 1.

Survey methodologist attempt to measure interactional context:

1. *Group composition* of aggregate individual attributes within a bounded setting.

Assumptions:

Bounded setting = social network AND compositional of individuals determines interaction

Survey methodologist attempt to measure interactional context:

1. *Group composition* of aggregate individual attributes within a bounded setting.

2. *Respondent reports* of social interaction.

Assumptions:

Bounded setting = social network AND compositional of individuals determines interaction

Comparable reports AND memory of interaction AND honest report.

Survey methodologist attempt to measure interactional context:

1. *Group composition* of aggregate individual attributes within a bounded setting.

2. *Respondent reports* of social interaction.

3. *Institutional accounts* from documents.

Assumptions:

Bounded setting = social network AND compositional of individuals determines interaction

Comparable reports AND memory of interaction AND honest report.

Assumes that institutional documents or elite statements represent everyday interaction.

Survey methodologist attempt to measure interactional context:

1. *Group composition* of aggregate individual attributes within a bounded setting.

2. *Respondent reports* of social interaction.

3. *Institutional accounts* from documents.

4. *Direct observation* from in situ observer

Assumptions:

Bounded setting = social network AND compositional of individuals determines interaction

Comparable reports AND memory of interaction AND honest report.

Assumes that institutional documents or elite statements represent everyday interaction.

Assumes that one location can generalise to other locations.

Our solution

Social media infrastructures **record all interactions** and their contents and constitutes a comprehensive and **rich database of textual interaction**.

Use **social media** as a **survey sampling frame** of a **ephemeral population**.

Combine individual level survey measures with social media **data on interaction context** in order to predict individual behaviour.

Facebook data

All posts, comments, replies and likes from over 100 Facebook groups collected through the API

Resulting in over 600.000 text messages.

Facebook as sampling frame

Aim: Collect a multilevel dataset of individuals (level 1) in relevant social media contexts (level 2) like for instance Facebook groups

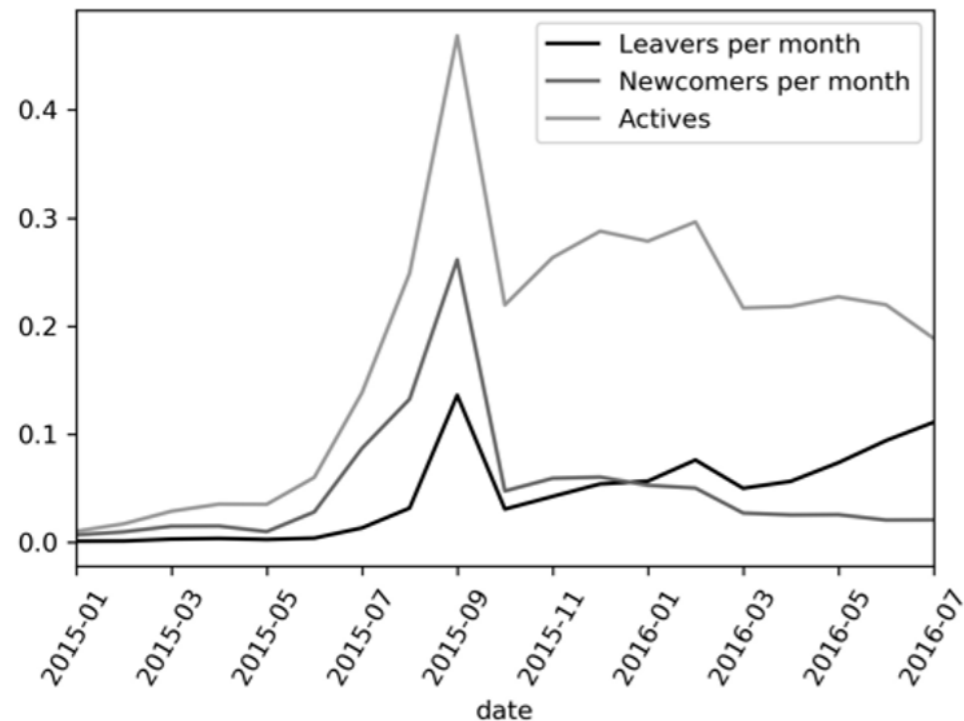
Use group specific survey links and self assignment to group.

Facebook as sampling frame

Aim: Collect a multilevel dataset of individuals (level 1) in relevant social media contexts (level 2) like for instance Facebook groups

Use group specific survey links and self assignment to group.

Active, new and leaving members' share of total movement population per month, January 2015-July 2016

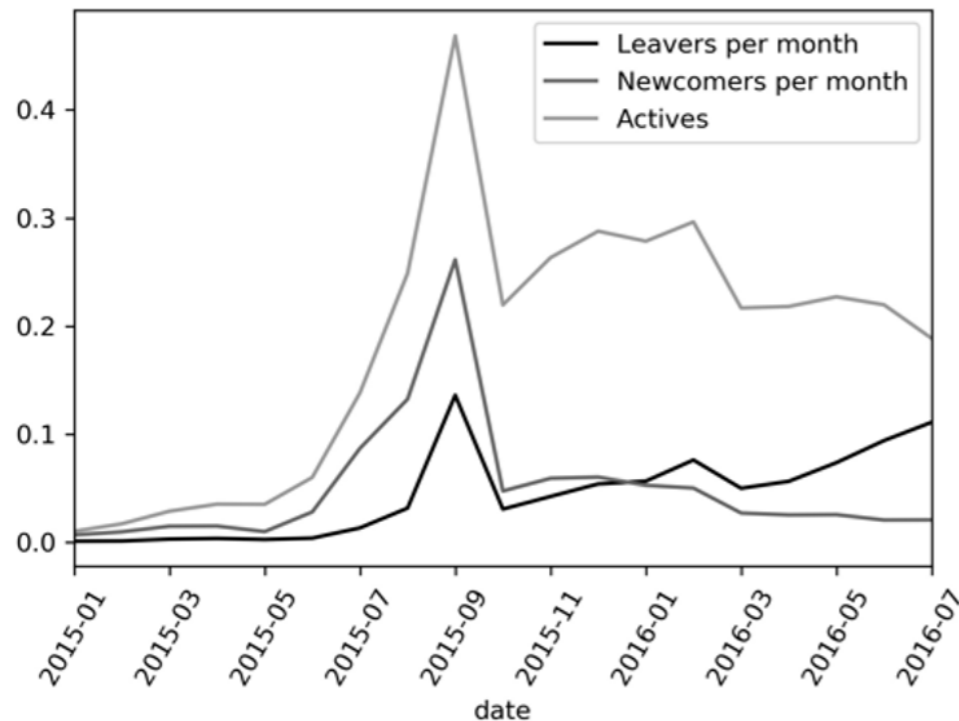


Facebook as sampling frame

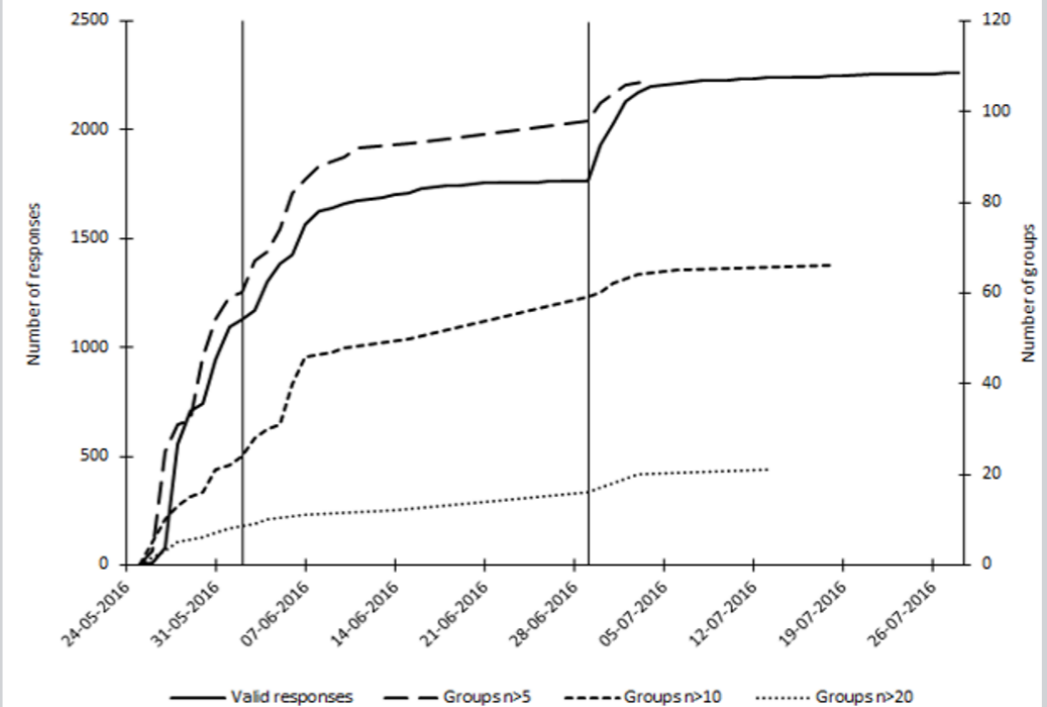
Aim: Collect a multilevel dataset of individuals (level 1) in relevant social media contexts (level 2) like for instance Facebook groups

Use group specific survey links and self assignment to group.

Active, new and leaving members' share of total movement population per month, January 2015-July 2016



Accumulation of valid individual and group level cases during data collection



Note: Data collection began on May 24 2016. Drop lines indicates posting of reminders.

How to find groups/pages on Facebook

Snowballing

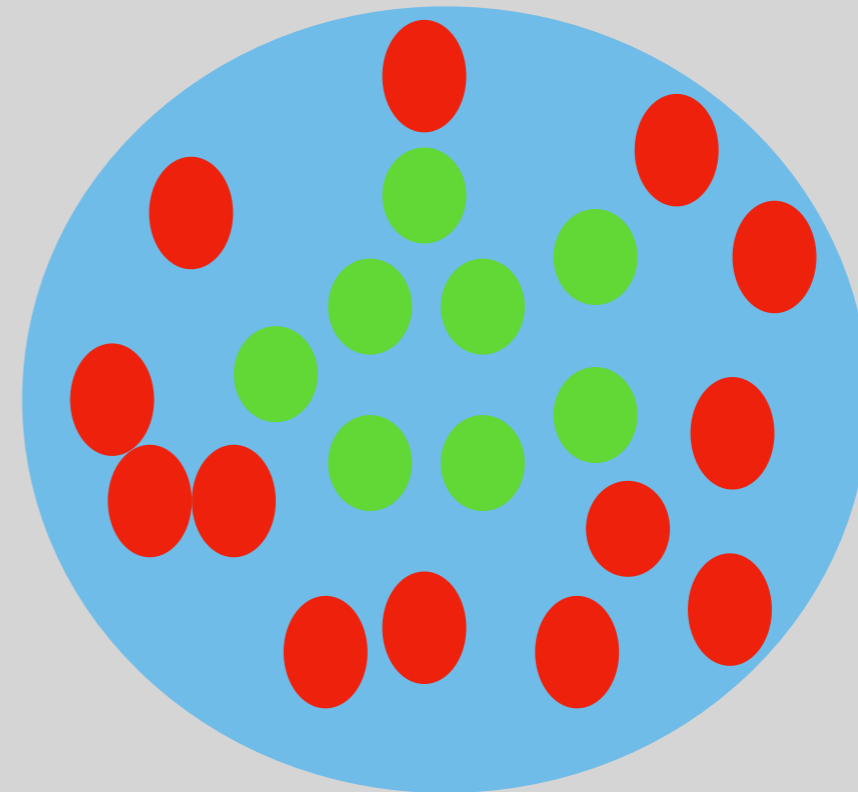
Large scale search

Verification

2. Facebook data collection

Table 1. Procedure for group sampling and basic description

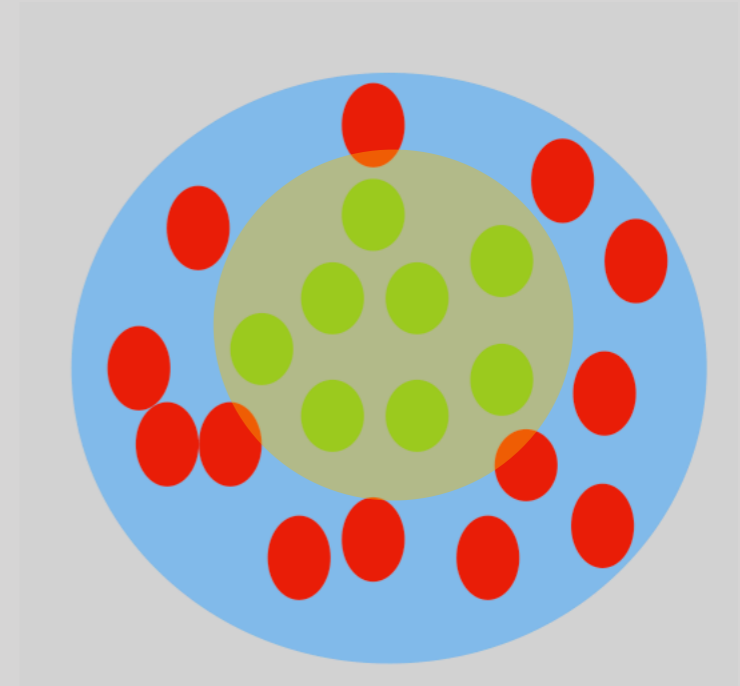
<u>Steps</u>	Step 1: Extensive group search
<u>General description of procedure</u>	Use social media search for both a general query targeting national groups and a location specific query targeting local groups.
<u>Goal</u>	High recall
<u>Specific procedure</u>	Our query was “ukraine hjælp”(ukraine help), our locational query was “[location]” (going through a full list of names of Danish
<u>Result</u>	A list of over 800 unique groups from the general and locational search.



2. Facebook data collection

Table 1. Procedure for group sampling and basic description

<u>Steps</u>	Step 1: Extensive group search	Step 2: Group selection
<u>General description of procedure</u>	Use social media search for both a general query targeting national groups and a location specific query targeting local groups.	The full list of groups returned by the extensive groups search is manually checked to ensure that the groups fall within the sample frame .
<u>Goal</u>	High recall	High precision
<u>Specific procedure</u>	Our query was “ukraine hjælp”(ukraine help), our locational query was “[location]” (going through a full list of names of Danish	We read the group description and selected those which aimed to provide practical support or expressed solidarity with ukrainians.
<u>Result</u>	A list of over 800 unique groups from the general and locational search.	128 groups were classified as Ukraine solidarity groups. 60 of these are tied to a location .



2. Facebook data collection

Table 1. Procedure for group sampling and basic description

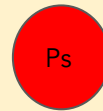
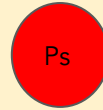
<u>Steps</u>	Step 1: Extensive group search	Step 2: Group selection	Step 3: Retrieve metadata	Step 4: Code groups
<u>General description of procedure</u>	Use social media search for both a general query targeting national groups and a location specific query targeting local groups.	The full list of groups returned by the extensive groups search is manually checked to ensure that the groups fall within the sample frame .	Facebook groups contain metadata which can be used to construct variables.	Location of group, group purpose, group rules and more.
<u>Goal</u>	High recall	High precision	Information for variable construction	Theoretical relevant variables
<u>Specific procedure</u>	Our query was “ukraine hjælp”(ukraine help), our locational query was “[location]” (going through a full list of names of Danish	We read the group description and selected those which aimed to provide practical support or expressed solidarity with ukrainians.	Retrieved group's number of members, amount of recent activity, group name, groups description and public/private status.	In our case we inferred the location of the group from the title of groups, the group purpose and rules (last two not reported).
<u>Result</u>	A list of over 800 unique groups from the general and locational search.	128 groups were classified as Ukraine solidarity groups. 60 of these are tied to a location .	A dataset with the variables specified above. We removed all groups not active within the last	A dataset with the variables specified above + those identified in step 3.

Snowball sampling the Danish Facebook Public

1. All activity from over 200.000 Danish Public Facebook pages between 2008-2018
2. Individual level activity from 3.000.000 users
3. 75 million posts with comments, 4.3 billion likes.

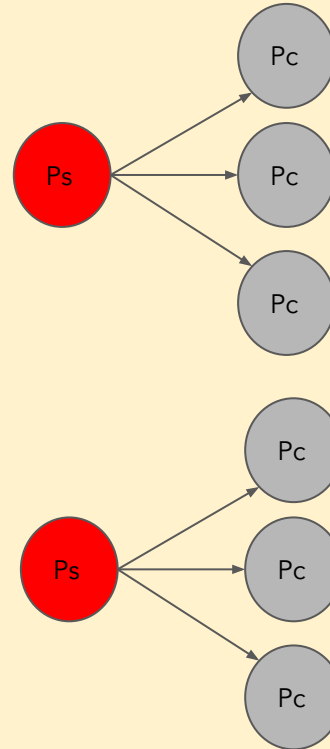
Public Facebook Data

1. Gathered an extensive seed of local and national Danish Facebook pages.



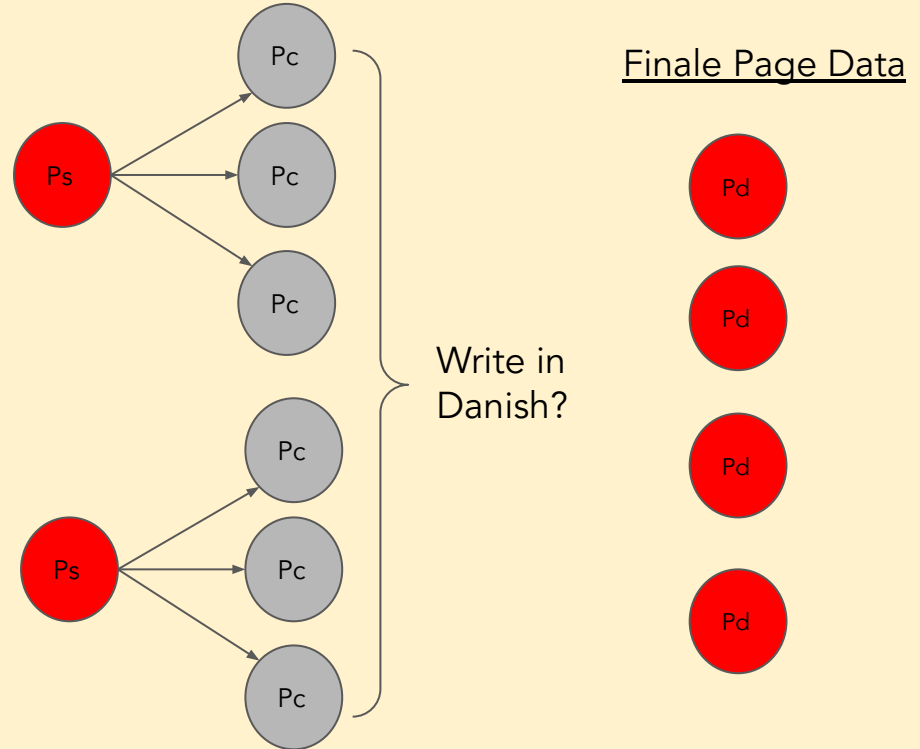
Public Facebook Data

1. Gathered an extensive seed of local and national Danish Facebook pages.
2. Snowball sampled millions of Facebook pages around the world.



Public Facebook Data

1. Gathered an extensive seed of local and national Danish Facebook pages.
2. Snowball sampled millions of Facebook pages around the world.
3. Selected the Page that wrote in Danish and collected activity data using the API.



Tak for jeres opmærksomhed

hc@sodas.ku.dk

OLS models of interaction context measures influence on participation in contentious participation (0-4)

Covariate	Model 2		Model 1		Model 3	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
<u>Group level</u>						
Cont. interactional context	5.301***	0.841	-	-	4.886***	0.806
Political composition	-0.764	0.446	-0.618	0.490	-	-
Contentious framing	0.393*	0.190	-	-	0.467*	0.185
<u>Individual level</u>						
Political attitude	-0.140***	0.028	-	-	-0.140***	0.028
<i>Additional covariates</i>		+		-		+
Constant	-0.546	0.291	1.007***	0.116	-0.796**	0.251
R ²	27.63		0.13		27.46	
Log likelihood	-1516.549		-1719.291		-1518.03	
degrees of freedom	16		2		15	

Notes: *= $p < 0.05$; **= $p < 0.01$; ***= $p < 0.001$. Coefficients are unstandardized. In model 1 and 3 additional controls are not reported. They can be obtained from the authors.