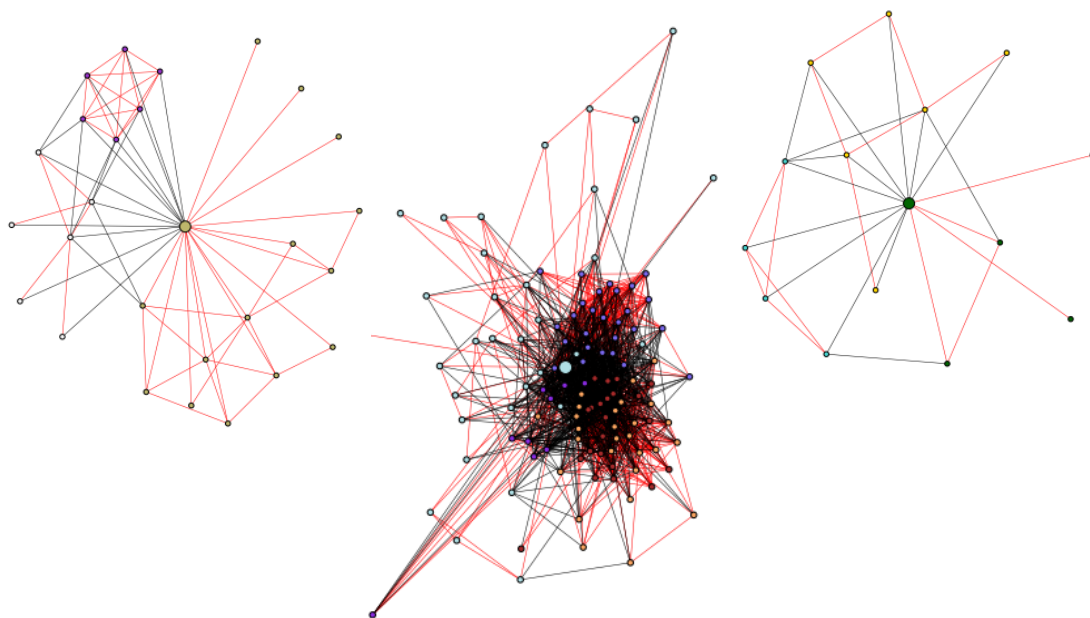




# Essays in Social Data Science

Group structure, integration and solidarity



Doctoral Thesis 2021  
Snorre Ralund

DEPARTMENT OF SOCIOLOGY  
FACULTY OF SOCIAL SCIENCES  
UNIVERSITY OF COPENHAGEN · DENMARK  
DOCTORAL THESIS 2021 · ISBN 978-87-7209-405-2

SNORRE RALUND

**Essays in Social Data Science**

Group structure, integration and solidarity

---

UNIVERSITY OF COPENHAGEN

DOCTORAL THESIS

---

**Essays in Social Data Science**  
**Group structure, integration and solidarity**

---

*Author:*  
Snorre Ralund

*Supervisor:*  
Anders Blok

*A thesis submitted in fulfillment of the requirements  
for the degree of Doctor of Philosophy  
in the*

Faculty of Social Sciences  
Department of Sociology

December 1, 2020

*"I can see the sun, but even if I cannot see the sun, I know that it exists. And to know that the sun is there - that is living."*

Fyodor Dostoevsky

## *Acknowledgements*

Det er med en sær vemodig lettelse at jeg sætter det sidste punktum. Et punktum som i lange tider, indtil det sidste, virkede usandsynligt, og som under ingen omstændigheder kunne være sket uden hjælp og støtte fra særligt en nær ven og samarbejdspartner, Hjalmar Bang Carlsen, hvis selskab, tænkning, begejstring og altid åbne og lyttende tilstedeværelse har sørget for at mit arbejde aldrig har været ensomt. Jeg har haft det store privilegie at bruge fire år af mit liv på at uddanne mig som forsker. Jeg er taknemmelig for det rum, og den tillid som jeg fået. Jeg vil gerne takke David Dreyer Lassen for denne tillid, men også for hans støtte, interesse, og vejledning. Jeg vil gerne takke min vejleder Anders Blok for at skubbe og skærpe min tænkning og for at tage enhver formulering alvorligt. Jeg vil endvidere gerne takke: Morten Axel Pedersen for diskussioner og højt humør; Robert Klemmensen ditto; Jonas Toubøl for gode stunder og teoretiske diskussioner. Sidst vil jeg takke min familie. Min kone for at holde ud, til tider lægge øre til hvad der må lyde som volapyk, og for at føde (og til tider passe alene) mine tre børn hvis aftenerne blev lange. Tak.

It is with strange melancholy and relief that I end the last paragraph. A paragraph, which for a long time, all the way to the end, seemed unlikely, and which under no circumstances could have happened without the help and support of especially one close friend and collaborator - Hjalmar Bang Carlsen - who's company, thinking, enthusiasm and an always open and listening presence have made sure that my work has never been lonely. I have had the great privilege of spending four years of my life educating myself as a researcher. I am grateful for this time and the trust which was given me. I would like to thank David Dreyer Lassen for this trust, but also for his support, interest, and guidance. I would like to thank my supervisor Anders Blok for always pushing and sharpening my thinking and for taking any wording seriously. I would also like to thank: Morten Axel Pedersen for discussions and high spirits; Robert Klemmensen ditto; Jonas Toubøl for good times and theoretical discussions. Finally, I want to thank my family. My wife for enduring, sometimes listening to what must have sounded like gibberish, and for giving birth to my three children. Thank you.

Doctoral Thesis 2021 © Snorre Ralund  
ISBN 978-87-7209-405-2 (Printed book)  
ISBN 978-87-7206-409-0 (E-book)  
Printed by SL grafik, Frederiksberg, Denmark (slgrafik.dk)

# Contents

Acknowledgementsiii

<b>1</b>	<b>Thesis introduction</b>	<b>1</b>
1.1	Introduction . . . . .	1
1.2	Facebook as data source . . . . .	3
1.3	Collecting public activity on Danish Facebook . . . . .	4
1.4	Research Ethics . . . . .	7
<b>2</b>	<b>Network structures and social integration</b>	<b>15</b>
<b>3</b>	<b>Foci stability and group formation</b>	<b>65</b>
<b>4</b>	<b>Measurement error and model instability in automated text analysis</b>	<b>101</b>
<b>5</b>	<b>Consequences of Group Style for Differential Participation</b>	<b>145</b>
<b>6</b>	<b>The Solidary Relationship’s Consequences for the Ebb and Flow of Activism</b>	<b>187</b>





*A true friend: Hjalmar Bang Carlsen*



## Chapter 1

# Thesis introduction

### 1.1 Introduction

This dissertation and dissertation work was done within a new field in the social sciences, social data science, with new methods and new data. I include here 5 individual articles each contributing to their respective research fields, and agendas. Common to all articles is a focus on group interaction, it's causes and consequences, as well as the ambition to leverage large scale online data to do comparative analysis across social groups and collectives.

Taking its departure in an extensive online social media data set, this PhD dissertation consists of 5 papers each developing research methods in computational social science and in particular social network analysis, computational content analysis and social media analysis. Substantially the dissertation investigate questions related to scientific fields around educational sociology, social networks and group formation, and social movements.

The research is done in two empirical settings: The social life and social networks of college students, and the social and political mobilization of activists in Denmark during the so called "refugee crisis" of 2015.

The first two papers investigate the social life of students. Under the heading "Network structures and social integration", the first paper compares the global networks structures of different study programs (N=44) to a range of social indicators: retention rates, survey based social satisfaction and social media based indicators of social activity. The paper shows how high retention rates, high social satisfaction, high social activity as well as more equally distributed social activity, is related to student cohorts with network structures that form well delimited groups instead of being widely distributed across the network. This contributes to an old theoretical discussion around the potential benefits or downsides of individualized network structures versus more locally delimited community based network structures. The second paper, "Foci stability and group formation", investigates how organizational infrastructures, and more specifically differences in how students are allocated into classes (foci) at each study program, affect the network structures as well as small group activity. The paper provide a possible mechanism for understanding the famously described "loss of the group" in contemporary society. The research focus

---

on both the mesolevel network groups ranging from 20-40 people, and on the activity of small interactional groups from 3-10. Data was obtained about how students are allocated into classes at 26 study programs, and a measure of "foci stability" was proposed. Foci Stability was then compared to a measure of network groupness, as well as social media derived aggregates of small group activity. The article shows that foci stability is highly predictive of network groupness and small group activity. The more stable the more grouped the network, and the social activity of the cohort will be more in small groups ( $\geq 3$  participants) as opposed to dyads.

The third paper, "Measurement error and model instability in automated text analysis", is a purely methodological piece, demonstrating issues with the use of unsupervised methods as measurement devices for automated content analysis, and especially the current practice of indirect validation in the use of Topic Models for "text as data" research. We argue that once an unsupervised topic model has been interpreted and labelled, it becomes a measurement device, that should be evaluated in the same way as supervised models. Using a simulation design, we show how current standards of validation allow a very high variability in both the classification performance of the topic model, as well as the actual research conclusions drawn based on its measurements.

The fourth paper, "Consequences of Group Style for Differential Participation", compares different regional groups of the refugee solidarity movements in Denmark. Using a supervised learning framework for doing computational content analysis, we measure the contentiousness of the interaction (posts and comments) in 119 online facebook groups organizing the humanitarian aid of refugees coming to Denmark. Combining the online data with a detailed survey of the activists it seeks to show how group-level degree of contentiousness, i.e. the group style, has consequences for individual-level political mobilization.

The fifth and final paper, "The Solidary Relationship's Consequences for the Ebb and Flow of Activism", studies activist retention in social movements in relation to interactions with the "deprived other". The research design experiments with the juxtaposition of evidence from qualitative life-history interviews with quantitative analysis of interactional social media data. From the interviews, meeting the deprived other, was described as a point of no return for the activists engagement in the movement - intensifying and prolonging their commitment to the cause. However, to generalize the finding to the whole movement, and to ameliorate valid criticisms of potential retrospective biases, we compared this to an analysis of the behavioral traces in the refugee solidarity facebook groups. The online activity, was used to construct individual activist careers allowing us to investigate temporal dynamics of retention and activity levels. Using name-based method to infer the ethnicity of each activist, we compared the careers of activists in relation to how much they made contact with a refugee. Using an elaborate case-control design the analysis of the online social media corroborated the findings from the qualitative interviews that meeting / the online interaction with the "deprived other" was an important