This Changes Everything (Referred as TCE)
Rezvaneh Rezapour, Jana Diesner
January 2015

Summary of Project: we study the impact of Naomi Klein's book, *This Changes Everything*, in anticipation of the release of the documentary film of the same name in 2015.

We started by asking what impact the book would like or expect to find. The main argument of the text is that the crisis of climate change is a sign of and an opportunity to address global economic inequality. Klein's book thus connects climate change and economic inequality—two themes that the author argues have not traditionally been linked in media as well as public and academic discourse. We theorized that if we were to see the book have an impact, one would expect to see the relationship of these two themes to change after the release of the book.

If we think of language as a window into our reality and also a shaper of that reality, the way in which topics like global warming or capitalism are talked about, who talks about them, where and how they are discussed are all indicators of the state of those issues in our world today. Utilizing theories and methods from network analysis and natural language processing, we set out to study the issues of climate change and economic inequality as reflected in media and social media, which together form what we refer to as "public discourse."

This report represents the first stage of the analysis of the impact of the book. The purpose of this report is to explain the process by which the text and network analyses were conducted and to share its findings and methods. Based on input from Naomi's team on their expected outcomes of the book in terms of impact, we began by asking the following questions:

1. What was the media and social media discourse on the topics (capitalism and climate change) *before* the book came out? [This is the baseline].
2. What is the media and social media discourse on the topic (capitalism and climate change) as well as on the book *after* the book's release?
3. To what extent do the media discourse on the topics (baseline) and on the book intersect with the actual content of the transcript [we refer to the book as the "the ground truth"]? I.e., does the public discourse pick up on content from the book/ does the content of the book inform the public discourse/ does the content of the book have an impact on the public discourse? Similarly, to what extent do the media discourse on the topics (baseline) intersect with the discourse on the book?
4. Do we see differences in the discourse between news media and social media? [Note: Social media includes Facebook and Twitter. News Media include a variety of outlets as indexed by and available through LexisNexis Academic].

General Procedure of Studying Public Discourse on Topics and Public Interest Media Initiatives (Book & Film):

The procedure of the project is divided into three parts, which are as follows:

- Baseline model: data collection and analysis prior to releasing the book and the film to assess public discourse on the topics addressed in book and film
- Ground truth model: analysis of the book transcript and documentary transcript
- Change assessment: data collection and analysis after releasing the book and the film to assess public discourse on the topics and on the book

We started with creating the **baseline model** of the topic of the book based on media and social media data. In the future, when the documentary is released, the same questions will be asked of the documentary. It is important to recognize that this stage of the project looks only at the public discourse of the topics and the book based on news and social media. Future stages of the project will also analyze the on-the-ground stakeholders of the topics of climate change, economic inequality, and their related networks. This type of analysis will necessitate, however, using a wider set of data, such as governmental data. The planning for this stage is currently in progress.

Before the Release of *This Changes Everything* (Book): Creating a Baseline Model

Step 1: Query Construction

We first needed to identify the keywords that result in the retrieval of documents that best capture to the main topics of the book. These key words are needed for querying media articles. The initial set of keywords was provided by Naomi Klein's team. We converted their input into various Boolean queries and tested for reasonable retrieval rates and relevance of retrieved results. We separated the words into two general themes of three words/phrases each. They were "climate crisis" OR "climate change" OR "renewable energy" and "Neoliberalism" OR "capitalism" OR "economic system." This approach requires concepts from each of the two themes to be present in each retrieved article.

Step 2: Data Collection from News Media (through LexisNexis Academic)

The keywords were used to collect news article data on the topic of the book for the time frame of one year prior to the release of the book. (September 17, 2013 to September 17, 2014). These were found using LexisNexis Academic, one of the world's largest online electronic libraries for legal, business, news, and public information. Note that the time frame is entirely flexible. In theory, we could collect data as far back as one hundred years or more

Step 3: Network Construction and Text Mining to make sense of the data

Next, we employ **a)** network analysis to detect key agents, organizations and themes, and **b)** text mining techniques to find trends in current discussions (topics, sentiments, and dynamics).

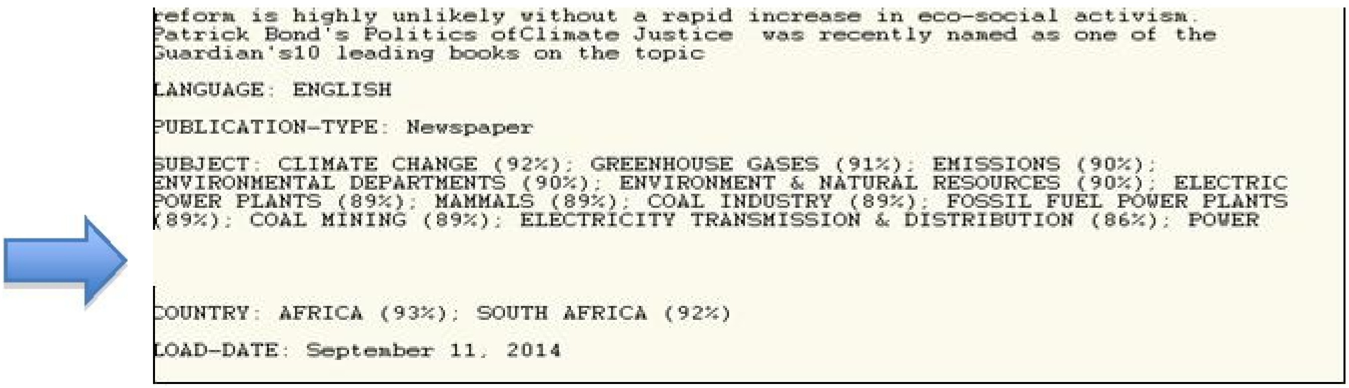
For Steps 2 and 3, we used three tools:

- ConText (http://context.lis.illinois.edu) was used to analyze the data collected from LexisNexis. ConText stands for **Con**nections and **Text**s. This is the short way of saying that ConText supports the construction of network data from structured and unstructured natural language text data. ConText is designed as a general applicable tool for conducting text analysis and network analysis in an integrated, systematic and automated fashion, especially for researchers and practitioners from the digital humanities, computational social sciences and real-world application domains (J. Diesner, 2014; J Diesner & Rezapour, 2015).

- Gephi (http://gephi.github.io) to visualize the networks generated in ConText.
- NodeXL (http://nodexl.codeplex.com) which is a free and open add-in for Excel that supports network overview, discovery and exploration. For this project, NodeXL was used to import and analyze data from Facebook and Twitter. NodeXL identifies quantitative (number of likes, followers), qualitative (text content related) and relational (social networks between users) information from these sources. By using this software, we are able to map different types of relationships between users and find out to what degree a user's account has impact.

Step 2 continued: Keyword Combinations & Data Acquisition from LexisNexis: Meta-Data

News articles consist of a header, a body, and meta-data. Meta-data include time stamps, name of newspaper or journal, and various index terms. Those index terms are typically automatically assigned to each article, and represent the main, high-level individuals, organizations, locations, subjects and other types of information addressed in an article, if applicable. Figure 1 shows an example of a news article as provided by LexisNexis entailing a portion of a text body plus some meta-data (language, publication type, load date, index terms for the categories of subject and country). Each index term is associated with a percentage value that indicates the relevance of the keyword for the article. Index terms per category are listed in decreasing order of relevance.


reform is highly unlikely without a rapid increase in eco-social activism.
Patrick Bond's Politics ofClimate Justice was recently named as one of the
Guardian's10 leading books on the topic

LANGUAGE: ENGLISH

PUBLICATION-TYPE: Newspaper

SUBJECT: CLIMATE CHANGE (92%); GREENHOUSE GASES (91%); EMISSIONS (90%);
ENVIRONMENTAL DEPARTMENTS (90%); ENVIRONMENT & NATURAL RESOURCES (90%); ELECTRIC
POWER PLANTS (89%); MAMMALS (89%); COAL INDUSTRY (89%); FOSSIL FUEL POWER PLANTS
(89%); COAL MINING (89%); ELECTRICITY TRANSMISSION & DISTRIBUTION (86%); POWER

COUNTRY: AFRICA (93%); SOUTH AFRICA (92%)

LOAD-DATE: September 11, 2014

Figure 1: A typical example of part of a news article as provided through LexisNexis

ConText contains routines for:

- Accurately and efficiently splitting each retrieved set of documents into individual, disambiguated documents
- Splitting individual documents into header, body, meta-data
- Populating a database for managing the meta-data on all articles

Note that ConText does not facilitate the collection of data from LexisNexis (for that, a subscription to LexisNexis is needed), but supports the management and analysis of data retrieved through this service.

After searching through LexisNexis Academic, we found 868 documents with the terms related to climate change and 877 documents related to the theme of capitalism for the year prior to the release of the film. These combined 1745 articles serve as the database for future analysis and is what we will refer to as the **dataset** for press on the topic before the release of the book or film (short: press on topic before). Using ConText, the downloaded data were disambiguated and the meta-data were written into a database that can be further use for data management, search and retrieval of the downloaded data.

Step 3 continued: Network Construction and Text Mining to make sense of the data: Creating, Visualizing and Analyzing Meta-Data Networks

Using ConText, we constructed semantic networks from the keywords of the type subject based on their co-occurrence per article. In these networks, a link indicates that one or more articles jointly addressed a pair of concepts. In addition to the consideration of subjects, similar networks can be easily created for connections among and between people, organizations, countries, cities, companies, and more.

In the following network visualizations (all produced in Gephi), similar colors indicate cluster affiliation (based on modularity), node size is scaled by degree (number of direct neighbors per node), and tie width reflects link frequency (i.e. number of articles in which a subject index terms pair co-occurred).

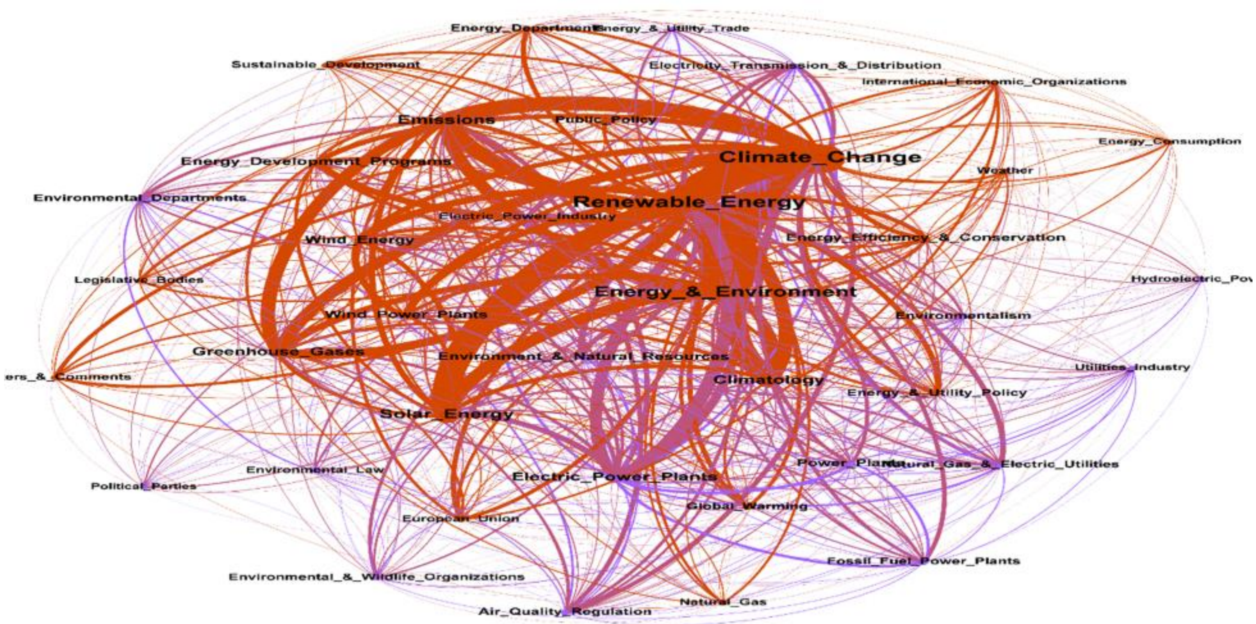


Figure 2: Subject-subject network from meta-data (Climate Change)

Figure 2: Semantic or Subject Network (Climate Change): As stated, the baseline model aims to capture the debate around the main topics of the book (e.g, climate change versus economic

inequality/ capitalism). The book asserts that the traditional public and academic discourse around the topic of climate change is separate from that of capitalism or economic inequality. Does the analysis of the metadata of social and news media from the year prior to the release of the book bear out this assertion?

<u>Findings:</u> Based on the meta-data analysis, the answer is yes. The results from the semantic network analysis tell us that in the traditional news media focused on the climate change and climatology the year before the film, issues like economic systems or capitalism were a relatively peripheral topic. Main themes in this public debate are climate change, energy – especially from renewable sources – and the environment.

Figure 3 (<u>Semantic or Subject Network: Capitalism):</u> These findings are confirmed by performing the same analysis with the capitalism articles where the environmental theme is only very peripheral. See Appendix I for a list of the most common subjects that appeared in connection with climate change.

<u>Findings</u>: In general, this issue (capitalism) is discussed in the context of politics and religion, not the environment.

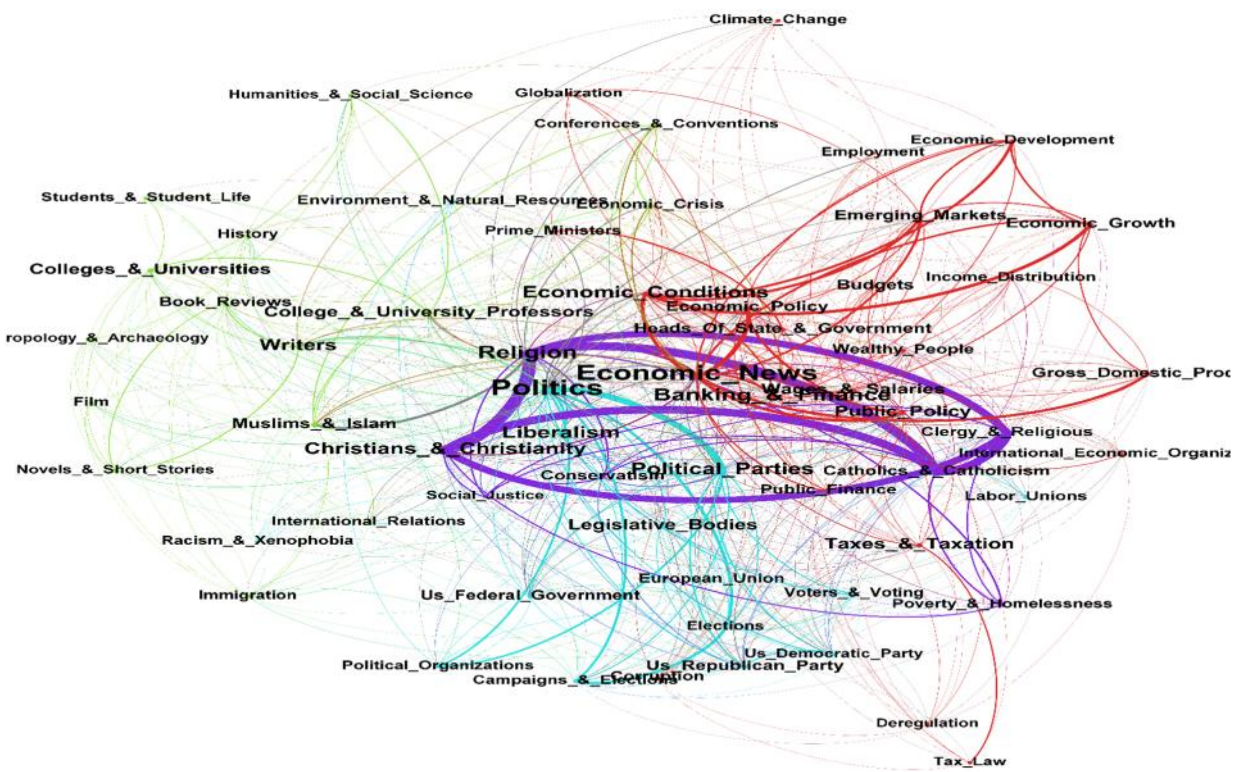


Figure 3: Subject-subject network from meta-data (Capitalism)

<u>What nodes are key in these networks?</u>

In network analysis, people often identify the most important nodes with respect to different dimensions of prominence, power and influence. In the next step of the analysis, we identified the most prominent nodes according to three common network analysis metrics.

The first is **degree centrality**, which represents the number of direct neighbors per node. A high degree means that a concept co-occurs with a large number of other concepts, i.e. this concept has multiple meanings or is used in a diverse set of contexts. Second is **betweenness centrality**, which represents a concept's ability to link other concepts. Concepts high on betweenness act as bridges between (clusters of) themes. The third is **eigenvector centrality**, which identifies nodes that are close to nodes that have a high degree (i.e. it is a recursive function based on degree). High eigenvector represents influential nodes that have a high number of neighbors, who are also neighbors with influential nodes.

Findings: The results for key player analysis (Table 1, Table 2) show that the discourse on climate change centers on that theme itself plus renewable energy, which includes solar energy and the environment. The discourse on capitalism, by contrast, is driven by the concepts of politics, the economy, banking and religion. These findings confirm the prior observation that the topics of climate change and capitalism do not intersect.

Table 1:Key nodes: Climate Change (bold: concept occurs in all three metrics)

Degree	Betweenness	Eigenvector Centrality
Climate Change	**Climate Change**	**Climate Change**
Renewable Energy	**Renewable Energy**	**Renewable Energy**
Energy_&_Environment	**Energy_&_Environment**	**Energy_&_Environment**
Solar Energy	**Solar Energy**	**Solar Energy**
Emissions	Climatology	Electric_Power_Plants

Table 2: Key nodes: Capitalism

Degree	Betweenness	Eigenvector Centrality
Politics	**Politics**	**Politics**
Economic News	**Economic News**	**Economic News**
Religion	**Banking_&_Finance**	**Religion**
Banking_&_Finance	**Religion**	**Banking_&_Finance**
Political Parties	Writers	Liberalism

Step 3 continued: Network Construction and Text Mining to make sense of the data:
Text Mining of News Article Bodies (Studying the Actual Content of the Articles)

Index term based analyses provide a fast, high level overview on the gist of a body of information. This bird eye view needs to be complemented with the salient themes and concepts that emerge from text bodies as the latter can provide a more nuanced and culturally sensitive understanding of the main themes that are explicitly or implicitly mentioned in some data (J. Diesner, 2012). We used ConText to study the actual content (bodies) of the 1745 articles in the dataset; digging deeper into the substance of the information.

We first used topic modeling in order to concisely summarize the text sets. Topic modeling is an unsupervised machine learning technique that summarizes the content of a corpus of unstructured, natural language text data in terms of the most salient topics that are explicitly or implicitly contained in the data (Blei, 2012). Each topic is represented by a fit value that indicates how strongly a topic describes a text set, as well as the most salient terms per topic in the underlying data. The terms per topic are sorted by their fit with a topic.

Topic modeling is a parameterized method, i.e. the user has to set the number of topics to be identified, and the iteration rate for the routine and a list with non-content bearing terms to be excluded from the analysis (stop word list). We identified the best settings per data set by running topic modeling with different parameter configurations multiple times and comparing the results and then gave a label to each topic, that's the middle row in the following two tables. The average fit columns indicates how strongly a topic describes a text set; the fits per table add up to 100%.

Table 3: Topics, Climate Change

Average Fit	Topic	Top terms
36%	Climate change	energy climate change renewable emissions
20%	Renewable energy	energy renewable climate Energy power
18%	Carbon emission regulations	energy renewable carbon government target
16%	Alternative energies	climate change people wind future
10%	Sustainable energy	energy Project sustainable project renewable

Table 4: Topics, Capitalism

Average Fit	Topic	Top terms
32%	Economic growth	economic capitalism system economy growth
21%	Governmental role	government people Labour business party
20%	Politics	political social state South power
15%	Capitalism	capitalism Francis Pope people economic
12%	Scholarly work	University book pages Press work

<u>Findings:</u> As evident in Table 3 (Topic Modeling: Climate Change), media articles from the dataset on climate change most frequently center on the topics of climate change, renewable energy, carbon emission regulations, and alternative as well as sustainable energy. The articles on capitalism are about economic growth, governmental influence, capitalism, politics and the consideration of scholarly work (Table 4). There is little to no overlap between these two main themes, which reinforces the findings from the meta-data analysis. Both levels of analysis for the dataset then confirm Naomi Klein's thesis that for the year before the release of the book, traditional news media on the topic of climate change do not discuss the topic in relationship to capitalism and vice versa.

<u>Step 3 continued: Network Construction and Text Mining to make sense of the data:</u>
<u>Semantic Networks constructed from the Content of the News Articles</u>

To be consistent with the meta-data analysis, we also built semantic networks of salient terms and themes that represent issues occurring in the content of the text bodies (J. Diesner & Carley, 2011a, 2011b). To identify these themes, we used ConText to apply the following techniques to the text bodies:

- Reusing the terms occurring in the meta-data of type subject, even if they might not match terms (including multi-word units) in the text data
- Terms resulting from topic modeling
- Top 250 terms based on term weighting techniques (tf*idf)
- Top 100 bigrams (e.g. World Bank)

All entries were added into a codebook, which is a data structure that consists of three columns: text terms, node labels, entity type. The node label column is there to handle entity resolution issues, such as consolidation of synonyms. The codebook was applied to the data such that any matches that co-occurred within a distance of seven terms from each other in the original article got linked.

<u>Findings:</u> The results are fairly consistent with the meta-data analysis: The semantic network on climate change (Figure 4) centers on renewable energy (red cluster) and global warming (blue cluster); lacking references to economic issues. The semantic network on capitalism (Figure 5) are about politics, money and power; not intersecting with ecological issues.

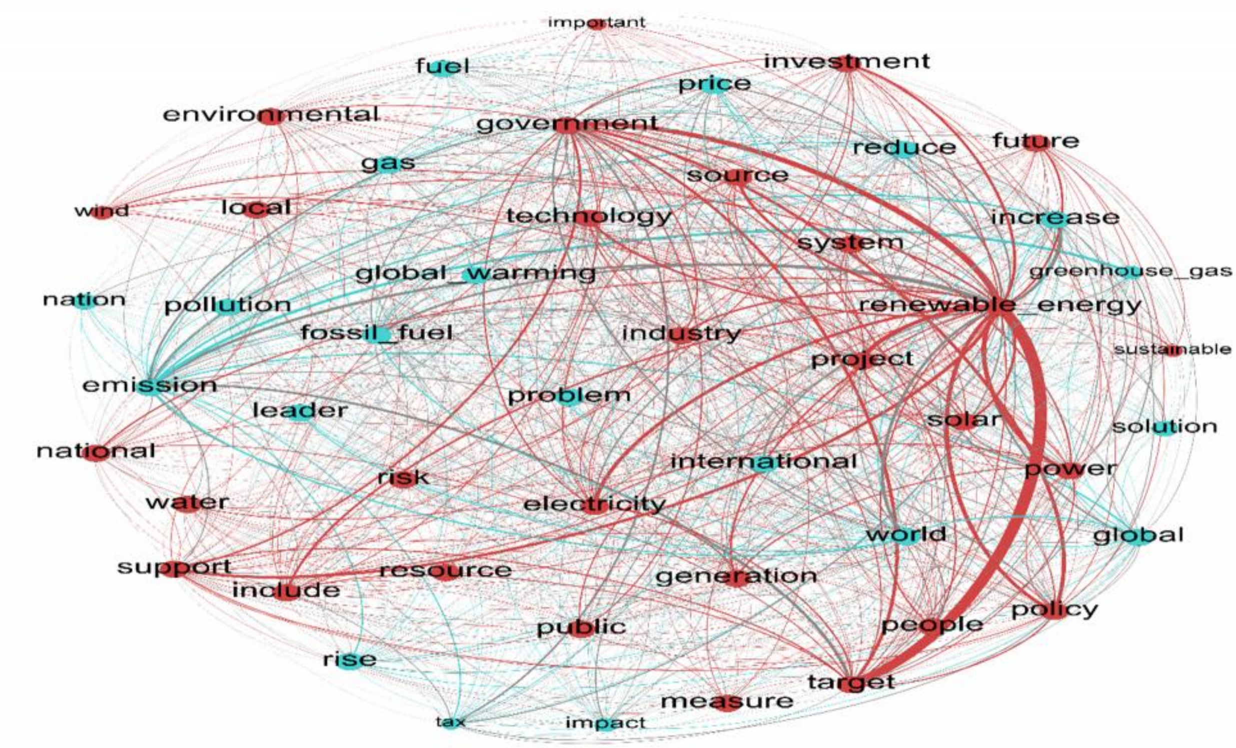


Figure 4: Semantic network from the body of the article (Climate Change)

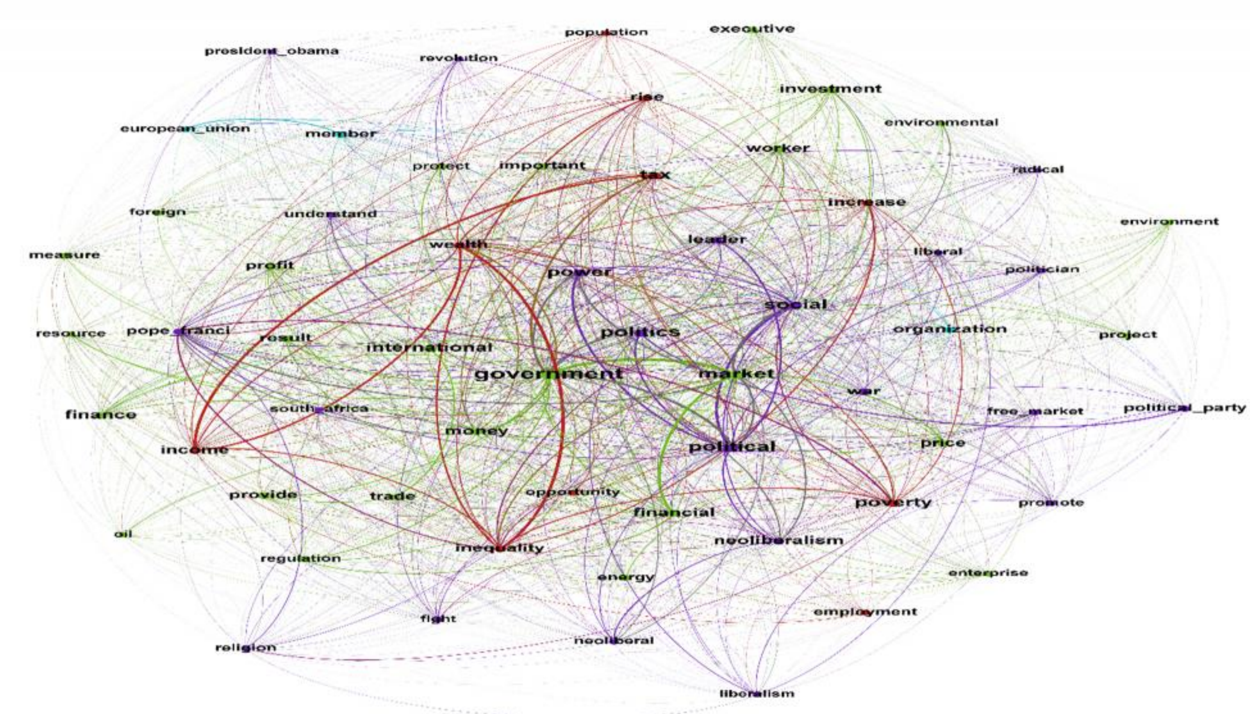


Figure 5: Semantic network of the body of the article (Capitalism)

<u>Further Use of Dataset: Identification of Journalists and Media by Name and County:</u>
In addition to aggregate features, the dataset can be further mined for particular information. Note that these data may be useful to the filmmakers in that they identify where - and even by which journalists - the issue of climate change is being discussed (and where it is not).

In the dataset, for example, the clearest identification of climate change as connected to the issue of economic inequality comes from an editorial in a Welsh paper called *The Western Mail* on April 7, 2014, which was written by Pippa Bartolotti, Wales Green Party Leader. Another one comes from *The Canberra Times* in Australia on September 6, 2014 though this article is concerned with the economic problems that Australia will face in the form of trade penalties from other countries if it does not reduce carbon emissions. Interestingly, climate change seems to be an important issue in Australia (judging from the number of articles published) relative to other countries. Australia is often associated with climate change issues, but not because it is a leader in trying to stop climate change. According to the articles, Australia seems quite resistant to policies that curb carbon tax, for example. (See Appendix I for a list of the most common subjects that appeared in connection with climate change).

Appendix III lists the top newspapers that have published articles on the themes of climate change, renewable energy, and the environment. In addition to Australia, England and India are also key countries with China to a lesser extent. Interestingly, within the United States, outside of US Officials News (newswire), only *The New York Times* seems to make the list (and only for 8 mentions during the year). This would indicate a point of concern for anyone interested in the issue of climate change and wanting to see more media attention on the issue in the United States.

A closer look at the meta data database helps to identify the journalists interested in or reporting on the issue of climate change. In the case of *The Guardian*, for example, it does not appear that only one reporter is reporting on the story but that several writers are involved. However, with the exception of Bartolotti's editorial, these articles do not directly tie the issue of climate change with economic inequality as Klein argues needs to happen.

The few mentions of climate change in the capitalism dataset come from pieces such as a "Letter to the Editor" of *The Observer* (England) by Kevin Albertson of Manschester Metropolitan University on May 4, 2014 or to the editor of *The Toronto Star* on December 24, 2013 by a reader named Ken Ranney, or a quote by Nick Robins, the head of the climate change centre of excellence of the bank HSBC in *The Guardian* (London) on February 13, 2014 when he participated in a Guardian Roundtable on capitalism. There is a brief mention in an article by Colin Hines in *Guardian Weekly* on November 29, 2013, which identified climate change as a cross-border issue, and more significantly, in a July 25, 2014 article on the green lobby's Margarita Declaration, which calls "for the death of capitalism," in *Investor's Business Daily*.

In general, however, discussion of climate change in the dataset on capitalism is dominated by mentions of Prince Charles, who publicly spoke about how those issues were linked in May 2014. In an article by Simon Walker, Director General of the Institute of Directors, in *City A.M.* (London) on May 29, 2014, in an article in *The Daily Telegraph* (London) on May 28, 2014 by Emily Gosden, or in an article by US Official News on May 29, 2014, Prince Charles's activities indicate that he has the potential and the reach to be a person of interest in Klein's goal for creating a stronger connection between climate change and economic inequality in the public discourse. Again, in addition to using the dataset to identify where and when the issues of climate change are being connected, one could also use it to identify potential key players—both detractors and supporters—and their networks.

Social Media Analysis

Because traditional media discourse and social media discourse cannot be assumed to be the same, we created a baseline analysis for both networks. For the social media of the book (*This Changes Everything*), we collected data relevant to two separate time period. The first one was on September 27th and the second time was on November 28/29th. Note that the book and the film share a social media account.

1. Twitter:

As of November 28, 2014, a total of **3,838** twitter users (followers) are following the film's Twitter account, while the account itself is following **118** users (followees/friends/followed). The account is following and followed by **108** users (reciprocated followers/intersection). The table below shows a comparison between the analysis of the movie in September 27 and November 28 in the number of followers, number of followees, and reciprocated followers.

Table 5: Number of associated users on Twitter

	September 27th	November 28th
Followers (Minus Intersection)	2,146	3,838 (+79%)
Followees (Minus Intersection)	21	118 (+562%)
Reciprocated followers (Intersection)	85	108 (+27%)
Total	2,425	4,064 (+67%)

NOTE: The table above shows how the book's Twitter account has changed in a two-month period. The number of followers of the account has increased by 79%.
As of November 28, of the **3,838** users following the book (but not followed by the book), **15** of them are power users. Below is a table of the 15 power users with their number of followers and a short description of their identity.

2. Facebook Fan Page:

We did not assume all social media platforms have the same reach or lead to the same public perception of a person, organization or topic structure. Thus, we complement the Twitter analysis with an analysis of users' activity on Facebook. On Facebook Fan Pages, users can provide comments to posts. This provides a valuable source for analyzing stimulus (posts) and responses (comments to posts), which together form a public discourse.

As of November 29, 2014, there were **248** posts on the Facebook page. Most contain some text data. On this page, **767** people have clicked **2,644** "Likes" and **59** users have posted **213** "Comments" since the page was created Table 6. Overall, there was a steep growth in all categories from Sept to Nov 2014.

Table 6: TCE on Facebook

	September 27th	November 29th
Overall likes	954	2,644 (+177%)
Users	394	826 (+110%)
Comments	53	213 (+402%)
Posts	53	248 (+368%)

We analyzed the socio-demographics of users based on the information provided in their profiles:

Table 7: Likes per gender

	September 27th	November 29th
Males	188 (47.7%)	360 (47%)
Females	173 (43.9%)	360 (47%)
Not specified	33 (8.3%)	47 (6%)

Table 8: Comments per gender

	September 27th	November 29th
Males	9 (27.2%)	32 (54.2%)
Females	8 (24.2%)	24 (40.6%)
Not specified	16 (48.4%)	3 (5%)

Location: (Top three)

- 553 are English US-68.7%
- 171 are English British- 21.2%
- 15 are French France- 1.86%

Gender differences:
The distribution of men and women is about equal in terms of liking, while men seem to be more active than women in stating their opinion in form of comments.

Next, we set out to map the different themes occurring in the comments. Figure 6 shows a clustered semantic network generated from the posts on this page, where nodes are words in the posts and links are formed if any two words co-occur at least five times within and/ or across posts. The width of the ties is proportional to the frequency of co-occurrence. Each emerging cluster is indicated by a separate box and color. These clusters represent the different themes that emerge from the discussion on this page.

The posts (Figure 6, Figure 7) focus on the author and announcements of the book. The comments are more plentiful and addressing the theme of the book with more nuances and also add additional personal thoughts. <u>We concluded that the Facebook page has been successful in stimulating a public debate on the core topics of the book. The amount of the comments and hence density of the network increased from September to November. To some extent, this is to be expected given that prior to the release of the book, readers would not have the same stimulus or information. But if this change in structure were not present, this would indicate a worrisome pattern.</u>

The key issue, in addition to followers accessed, is the actual nature of the discussion and their relationship to the change in the network of that issue. What the data does indicate is the relationship between followers who post. (See next section).

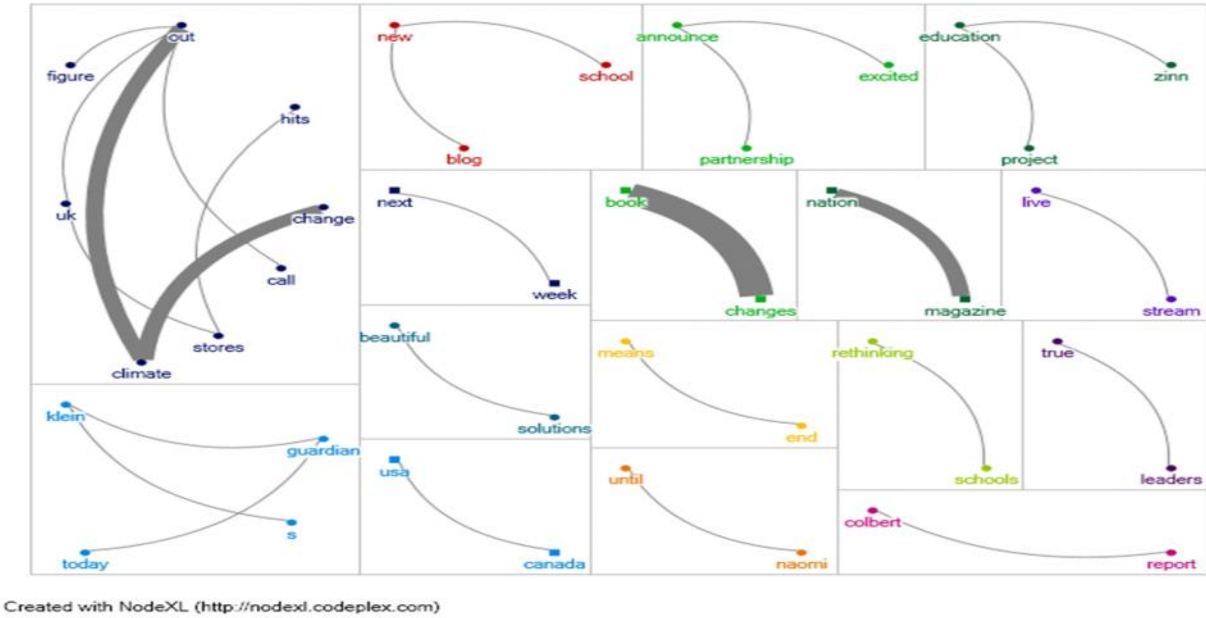


*Figure 6: Semantic map of **posts** on the Facebook fan page (September 27, 2014)*

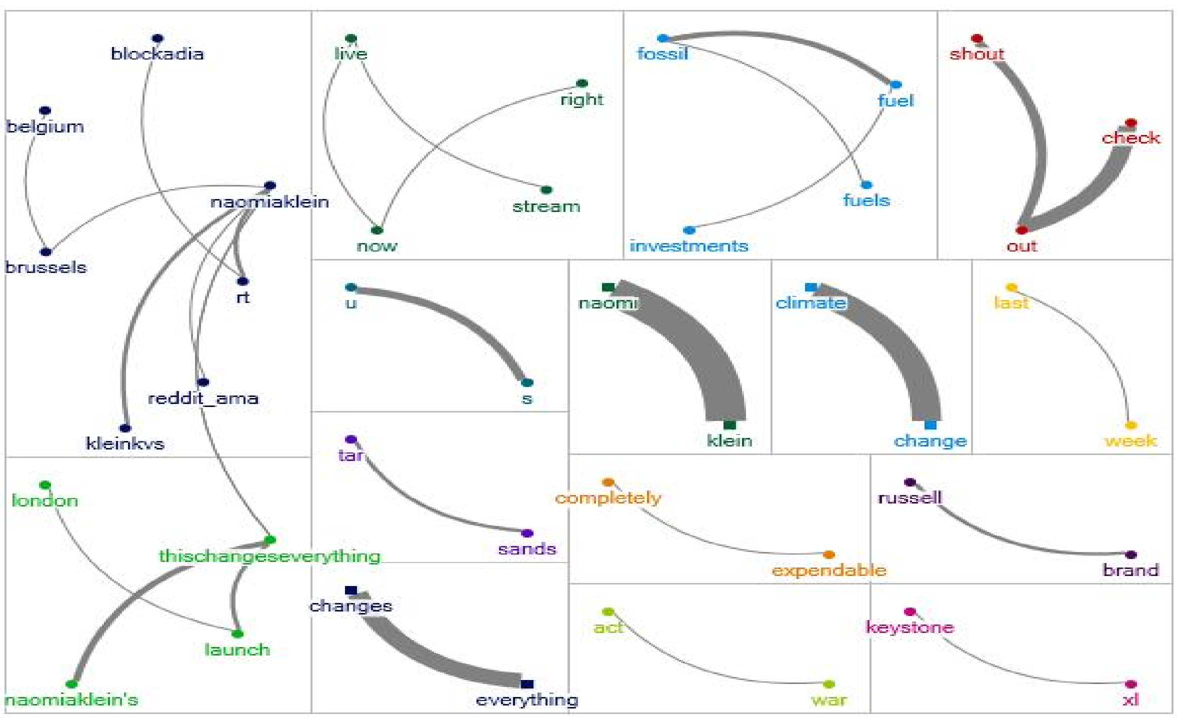


*Figure 7: Semantic map of **posts** on the Facebook fan page (November 29, 2014)*

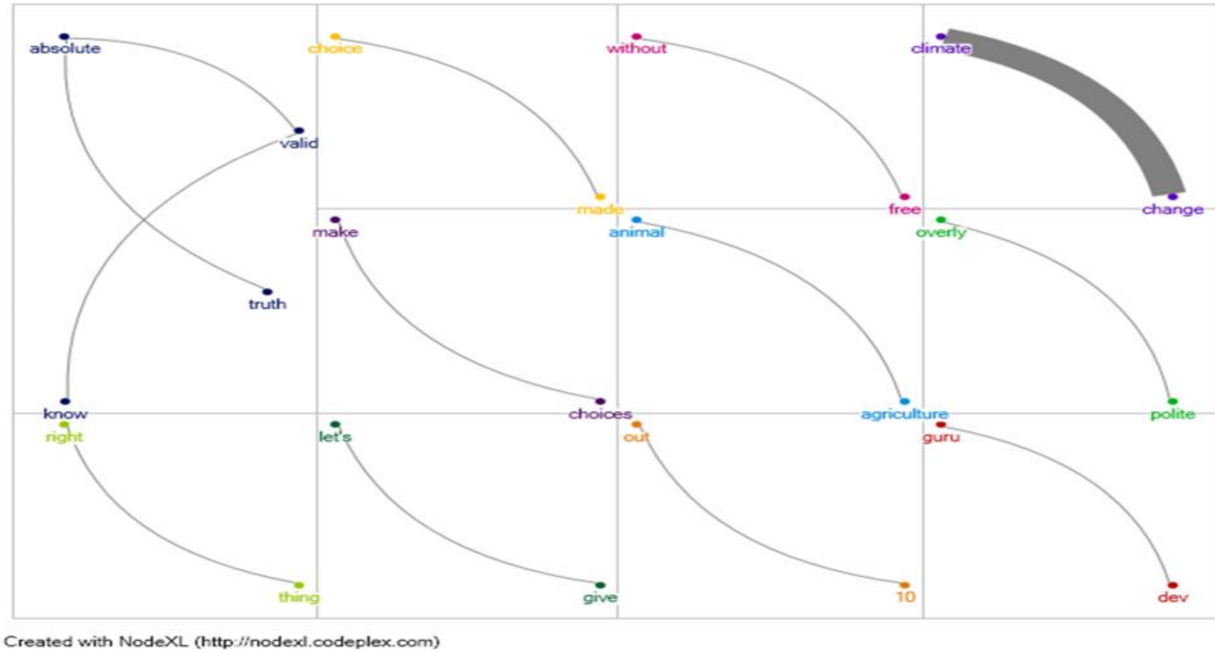


*Figure 8: Semantic map of **comments** to posts on the Facebook fan page (September 27, 2014)*

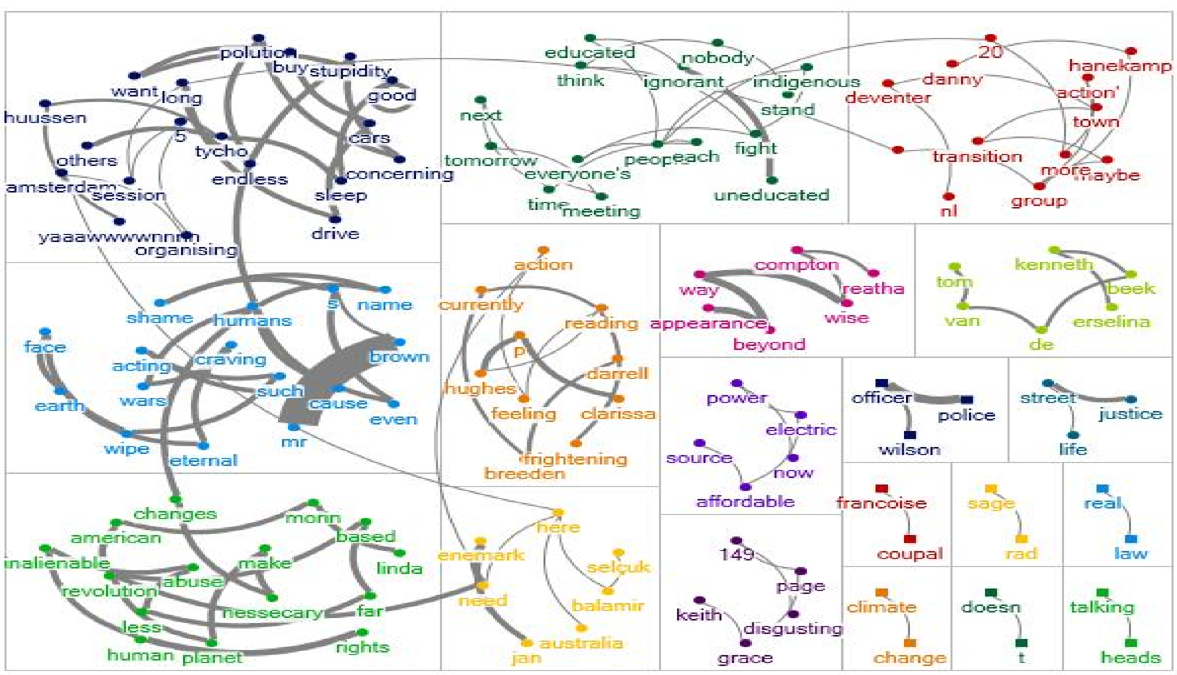


*Figure 9: Semantic map of **comments** to posts on the Facebook fan page (November 29, 2014)*

Acknowledgment

This work is supported by the FORD Foundation's JustFilms Fund and Naomi Klein and her team.

Bibliography

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77-84.

Diesner, J. (2012). *Uncovering and managing the impact of methodological choices for the computational construction of socio-technical networks from texts.* Carnegie Mellon University. (CMU-ISR-12-101, PhD Thesis)

Diesner, J. (2014). *ConText: Software for the Integrated Analysis of Text Data and Network Data*. Paper presented at the Social and Semantic Networks in Communication Research. Preconference at Conference of International Communication Association (ICA), Seattle, WA.

Diesner, J., & Carley, K. M. (2011a). Semantic Networks. In G. Barnett & J. G. Golson (Eds.), *Encyclopedia of Social Networking* (pp. 766-769): Sage.

Diesner, J., & Carley, K. M. (2011b). Words and Networks. In G. Barnett & J. G. Golson (Eds.), *Encyclopedia of Social Networking* (pp. 958-961): Sage.

Diesner, J., & Rezapour, R. (2015). *Social Computing for Impact Assessment of Social Change Projects*. Paper presented at the Social Computing, Behavioral-Cultural Modeling, and Prediction (SBP), Washington, DC, USA.