



**University
of Basel**

Center for International Economics
and Business | CIEB

What's it worth?

Predicting the Importance of Patents Using Supervised Learning

Dr. Matthias Niggli and Dr. Christian Rutzer

7th European COST Conference on Artificial Intelligence in Industry and Finance

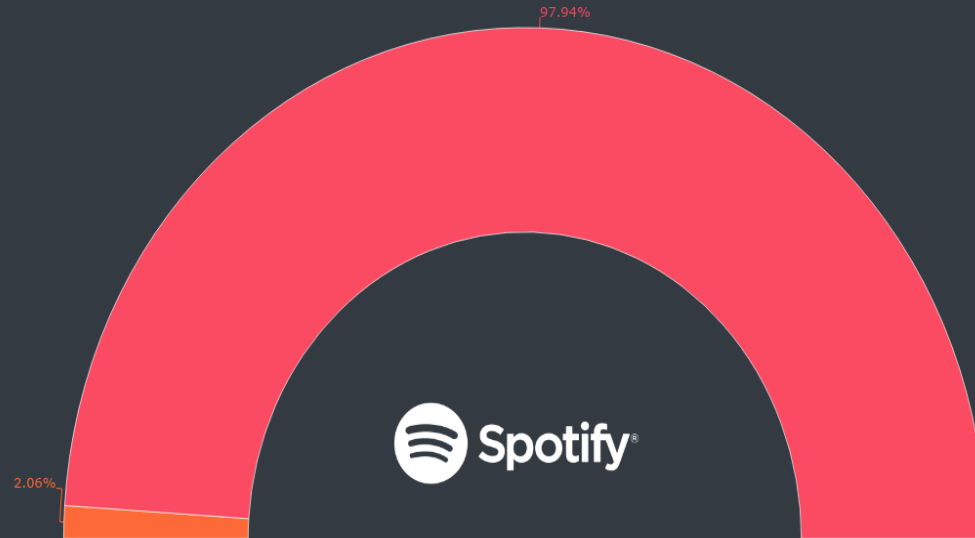
September 28, 2022



Spotify

How many artists have released over 10 tracks on Spotify AND have over 10,000 monthly listeners?

- Artists on Spotify who've released over 10 tracks to date AND have over 10,000 monthly listeners (165k)
- Artists on Spotify who haven't released over 10 tracks to date AND/OR don't have over 10,000 monthly listeners (7.835M)



Source: Loud & Clear data

MBW

Source: Loud&Clear (2022)
<https://loudandclear.byspotify.com/>

- There are about **8 million artists** on Spotify.
- But **most artist don't get a lot of attention.**
- Only **2%** (165'000) have at least 10'000 monthly listeners and more than 10 songs.

Similarly, only a fraction of patents is technologically important

(12) **United States Patent**
Gelfenbeyn et al.

(10) **Patent No.:** US 11,170,415 B2
(45) **Date of Patent:** *Nov. 9, 2021

(54) **ENHANCING FUNCTIONALITIES OF VIRTUAL ASSISTANTS AND DIALOG SYSTEMS VIA PLUGIN MARKETPLACE**

(71) Applicant: Google LLC, Mountain View, CA (US)

(72) Inventors: Ilya Gennadyevich Gelfenbeyn, Sunnyvale, CA (US); Artem Goncharuk, Mountain View, CA (US); Pavel Aleksandrovich Strotin, Sunnyvale, CA (US)

(73) Assignee: GOOGLE LLC, Mountain View, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.
This patent is subject to a terminal disclaimer.

(21) Appl. No.: 16/424,783
(22) Filed: May 29, 2019

(65) **Prior Publication Data**
US 2019/0279264 A1 Sep. 12, 2019

Related U.S. Application Data
(63) Continuation of application No. 15/166,293, filed on May 27, 2016, now Pat. No. 10,311,492. (Continued)

(51) **Int. Cl.**
G06Q 30/06 (2012.01)

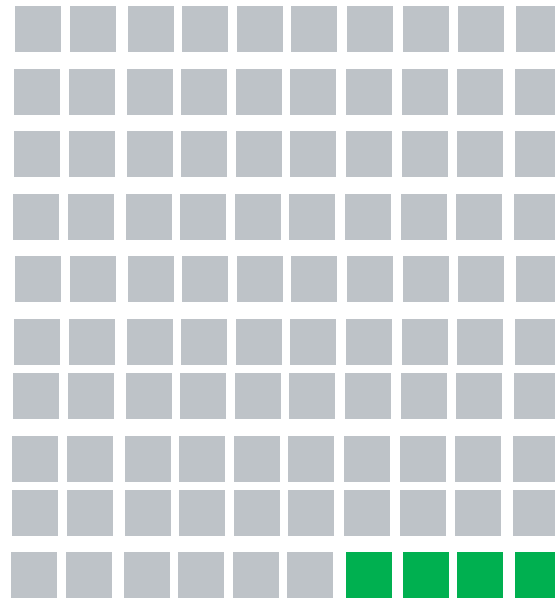
(52) **U.S. Cl.**
CPC: G06Q 30/0601 (2013.01)

(58) **Field of Classification Search**
CPC: G06Q 30/0601; G06Q 30/0601
See application file for complete search history.

(56) **References Cited**
U.S. PATENT DOCUMENTS
6,565,640 B1 * 12/2003 Bennett G09B 5/04 704/317
6,565,650 B1 * 12/2003 O'Grady, Jr. G05B 19/042 706/12
(Continued)
FOREIGN PATENT DOCUMENTS
CN 102668499 9/2012
CN 103119557 5/2013
(Continued)
OTHER PUBLICATIONS
China National Intellectual Property Administration; Notification of First Office Action issue in Application No. 201680018980.5, dated Mar. 18, 2020.
(Continued)
Primary Examiner — Cao H Nguyen
(74) Attorney, Agent, or Firm — Middleton Reutlinger

(57) **ABSTRACT**
A method for enhancing dialog systems is disclosed herein. The method may include maintaining an online marketplace that may have a plurality of dialog system extension elements. The plurality of dialog system extension elements may include at least one of a dialog system plugin, a dialog system add-on, a dialog system update, and a dialog system upgrade. The method may further include receiving a selection of one of the plurality of dialog system extension elements from an end user. The end user may be associated with a dialog system. The method may continue with associating the one of the plurality of dialog system extension elements with the dialog system of the end user.

14 Claims, 9 Drawing Sheets



From a sample of patents filed in 2015, only 4% have received more than 10 citations.

This presentation is about **how to separate important patents from unimportant ones?**

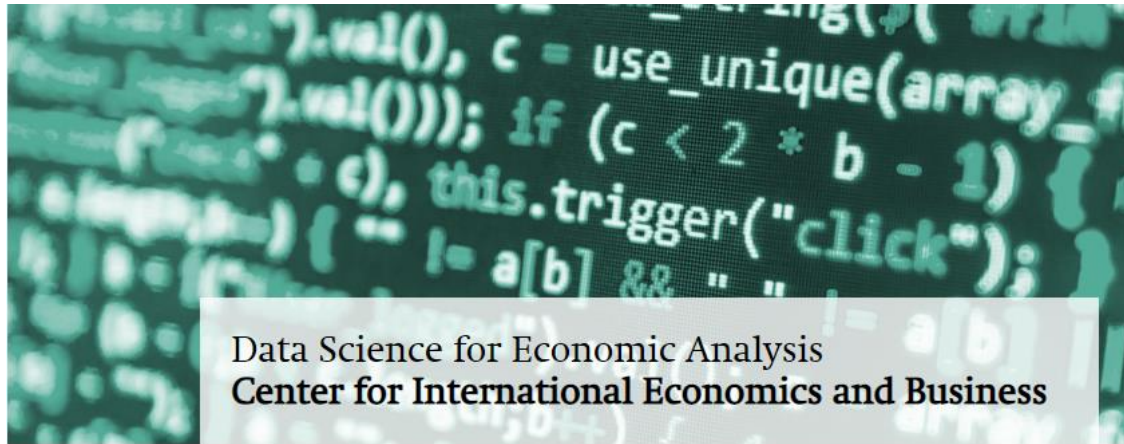


Source: Flickr, Tony Hammond ©
<https://www.flickr.com/photos/8525214@N06/6124451226>

Outline

- 1. Background**
- 2. Patents and Patent Data**
- 3. Data Engineering**
- 4. Model and Performance**
- 5. Takeaways**

Who we are



Dr. Christian Rutzer
Deputy Director &
Economist CIEB



Dr. Matthias Niggli
Research Data Scientist &
Economist CIEB

Why Bother Predicting the Importance of Patents?

- **Patents are booming:** 188'000 new patent applications submitted to the European Patent Office (EPO) in 2021.
- **Patents** are an **early indicator for inventive activity** (of firms, regions or even countries) and **approximate future productivity** (see, e.g., Kogan 2017; Hall 2007, 2005).
- But **patents' importance** can only be **observed retrospectively** - with a time lag of up to 5 years!
- **Too late for business analytics** that are obviously interested in much shorter time spans.
 - **Can we predict, which patents will become important in near-realtime?**

The Project

Project funded by **Innosuisse**:

- Partners: **ZHAW** and **Econsight**.
- Proof-of-Concept.
- Implemented for 2 different technology fields («renewable energy», «mobility & storage»).

Patents and Patent Data



US010872024B2

(12) **United States Patent**
Cranfill et al. (10) **Patent No.:** **US 10,872,024 B2**
(45) **Date of Patent:** **Dec. 22, 2020**

(54) **USER INTERFACES FOR CONTROLLING OR PRESENTING DEVICE USAGE ON AN ELECTRONIC DEVICE** (58) **Field of Classification Search**
CPC H04L 12/00; H04L 12/02; H04L 12/14; H04L 63/00; H04L 63/10; H04L 63/102; (Continued)

(71) Applicant: **Apple Inc.**, Cupertino, CA (US) (56) **References Cited**
(72) Inventors: **Elizabeth Caroline Cranfill**, San Francisco, CA (US); **Christopher P. Foss**, San Francisco, CA (US); **David C. Graham**, Cupertino, CA (US) U.S. PATENT DOCUMENTS
5,483,261 A 1/1996 Yasutake
5,488,204 A 1/1996 Mend et al.
(Continued)

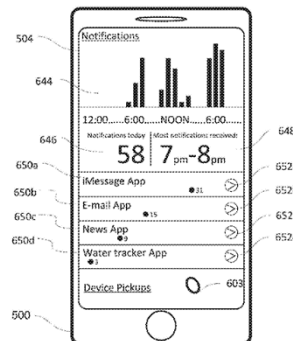
(73) Assignee: **Apple Inc.**, Cupertino, CA (US) FOREIGN PATENT DOCUMENTS
(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days. CN 103294965 A 9/2013
CN 104471521 A 3/2015
(Continued)

(21) Appl. No.: **16/147,069** OTHER PUBLICATIONS
(22) Filed: **Sep. 28, 2018** Unknown Author—My Data Manager Track your mobile data usage and save money—appPicker—Aug. 18, 2015 (Year: 2015).*
(65) **Prior Publication Data** (Continued)

US 2019/0347180 A1 Nov. 14, 2019
Related U.S. Application Data
(60) Provisional application No. 62/692,831, filed on Jul. 1, 2018, provisional application No. 62/679,927, filed (Continued)
Primary Examiner—Tauqir Hussain
Assistant Examiner—Boris D Grijalva Lobos
(74) *Attorney, Agent, or Firm*—Kubota & Basol LLP

(51) **Int. CL** (57) **ABSTRACT**
G06F 3/0481 (2013.01) In some embodiments, an electronic device presents indications of usage metrics for the device. In some embodiments, an electronic device sets, configures and/or enforces device usage limits. In some embodiments, an electronic device limits access to certain applications during certain periods of time. In some embodiments, an electronic device suppresses auxiliary functions of certain applications when an application usage limit or restriction criteria associated with those applications is reached. In some embodiments, an electronic device manages restriction settings with permission optionally provided by another electronic device.
G06F 9/54 (2006.01) (Continued)
(52) **U.S. CL**
CPC **G06F 11/3438** (2013.01); **G06F 3/04812** (2013.01); **G06F 3/04817** (2013.01); (Continued)

78 Claims, 282 Drawing Sheets



Patents and Patent Data

- Patent data contain a lot of information that can be exploited.
- Patent data is rather messy and subject to many caveats.
 - Requires a good understanding of the patent corpus and how to handle it.
- For some issues there exist best practices in the literature, but a lot remains up to the scientist.
 - Many degrees of freedom.

How to *Measure* the Importance of Patents?

- **Patents cite other patents.**

(56)

References Cited

U.S. PATENT DOCUMENTS

5,483,261	A	1/1996	Yasutake
5,488,204	A	1/1996	Mead et al.

(Continued)

FOREIGN PATENT DOCUMENTS

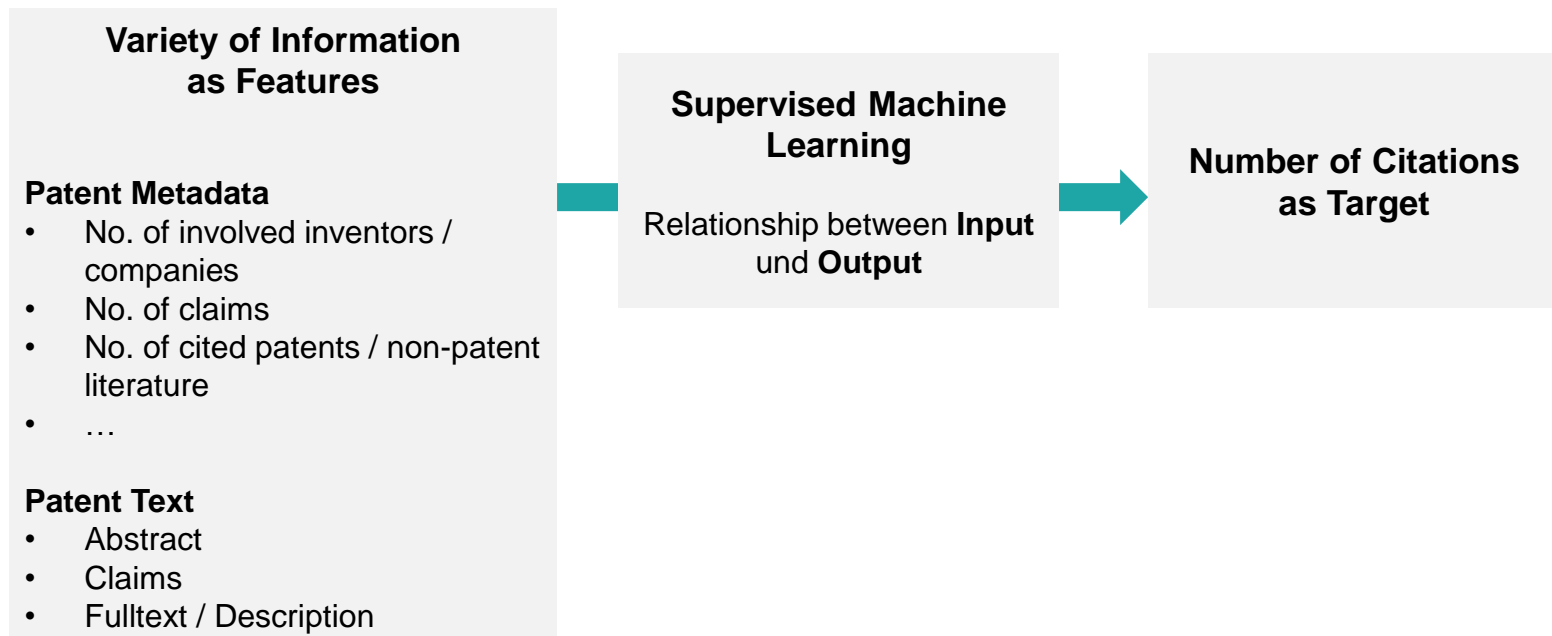
CN	103294965	A	9/2013
CN	104471521	A	3/2015

(Continued)

- Number of received citations is positively related to technological innovation and the economic value of a patent (see Kogan et al. 2017 or Hall et al. 2005, for example)

How To *Predict* the Importance of Patents?

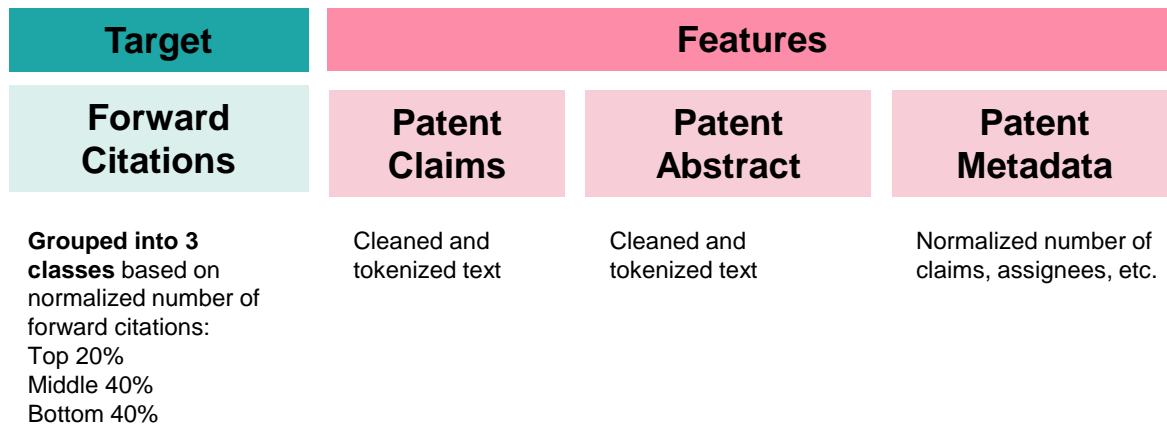
- Patents are cited only over time.
- Assessment of technological significance is not possible for new patents.



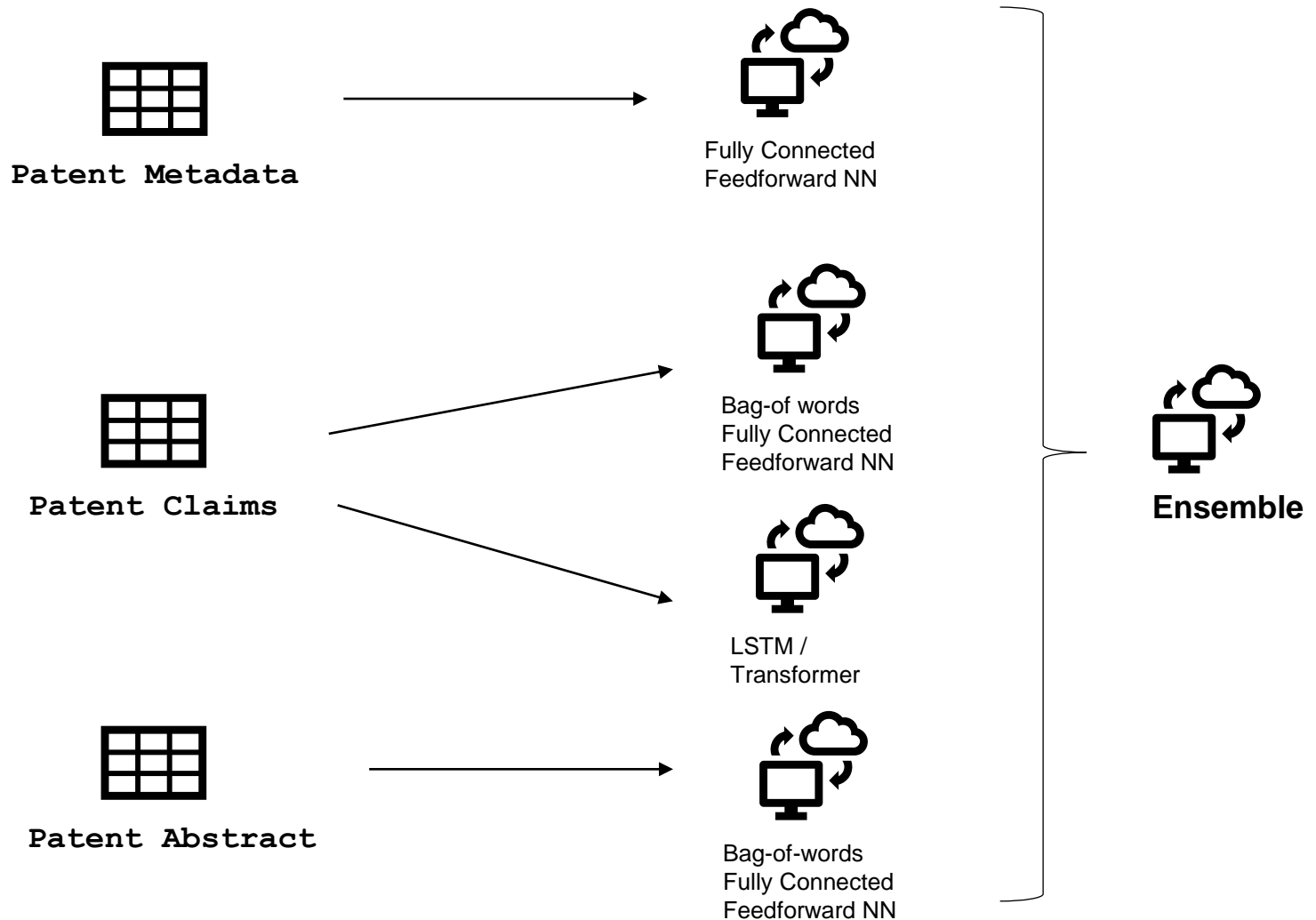
Data Engineering

- The same invention is often codified in different patents (patent families).
 - If patent families are not considered correctly, data leakage!
- **Features**
 - Meta-data and text from the most cited patent of a family.
 - Standardization by priority year and technological field.
 - Standard NLP cleaning.
- **Target**
 - Family-to-family citations.
 - Citation quantiles of publication year and technological field.

Training a Classifier: Data Overview



Training a Classifier: Architecture



Training a Classifier: Performance

Technology Field	Number of Samples	Weighted F1 Score (All Patent Groups)	F1 Score (Top Patents)
Renewable Energy	4415	51.9%	59.5%
Mobility and Storage	4021	52.8%	61.1%

Takeaways

What we have learned

- Decent knowledge of patent data is crucial
- Many degrees of freedom
- Proof-of-concept successful (Performance in line with SOTA e.g., Chung & Sohn 2020; Lee, Kwon, Kim & Kwon 2018)
- Scalable framework to other technologies

Takeaways

What questions remain

- Is a classifier trained on patents from 2010-2016 reliable to predict new patents in 2022?
- How to best handle the trade-off between narrowly defined technology fields vs. the smaller number of corresponding samples?
- How to more efficiently leverage patent text for classification (preprocessing, embeddings, language models)?

Find Us Online



@cieb_unibas



www.innoscape.ch



www.cieb.unibas.ch

LinkedIn

<https://www.linkedin.com/company/cieb-unibas>

Literature

Chung, P., & Sohn, S. Y. (2020). Early detection of valuable patents using a deep learning model: Case of semiconductor industry. *Technological Forecasting and Social Change*, 158, 120146.

Hall, B. H., Jaffe, A., & Trajtenberg, M. (2005). Market value and patent citations. *RAND Journal of Economics*, 36(1), 16-38.

Hall, B. H., Thoma, G., & Torrisi, S. (2007). The market value of patents and R&D: Evidence from European firms. In *Academy of Management Proceedings*, 2007(1), 1-6.

Lee, C., Kwon, O., Kim, M., & Kwon, D. (2018). Early identification of emerging technologies: A machine learning approach using multiple patent indicators. *Technological Forecasting and Social Change*, 127, 291-303.