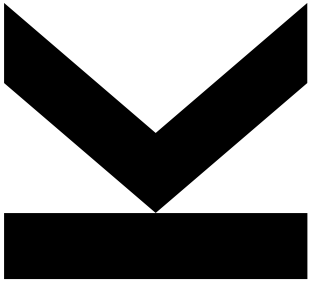
JOHANNES KEPLER
UNIVERSITY LINZ

JOHANNES KEPLER UNIVERSITY LINZ

(Towards) exploring the genesis of competition in economic thought



Spatial Competition and Economic Policies

SPACE

Institute for Comprehensive Analysis of the Economy

Draft paper version prepared for the
Young Economists Conference 2019
1-2 October 2019, Vienna

Matthias Aistleitner, MSc
matthias.aistleitner@jku.at
www.jku.at/icae

FWF Der Wissenschaftsfonds.

JOHANNES KEPLER
UNIVERSITY LINZ
Altenberger Straße 69
4040 Linz, Austria
jku.at

Research background: SPACE project

Research background: SPACE project

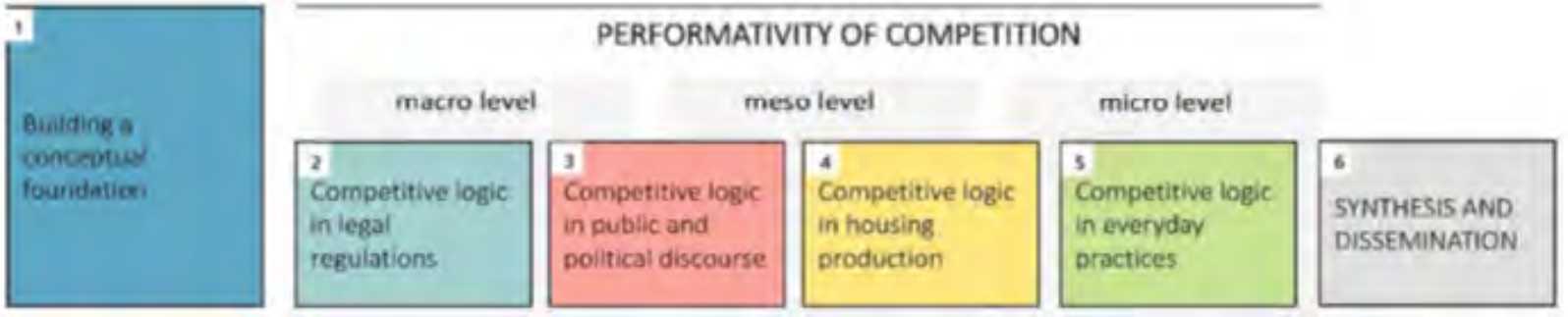
- Spatial Competition and Economic Policies (SPACE): Discourses, Institutions and Everyday Practices

Research background: SPACE project

- Spatial Competition and Economic Policies (SPACE): Discourses, Institutions and Everyday Practices
  - Investigating the impact of an increasingly strong reliance on competition as a prime mode of social organization
  - Competition as a core concept for designing social institutions

Research background: SPACE project

- Spatial Competition and Economic Policies (SPACE): Discourses, Institutions and Everyday Practices
  - Investigating the impact of an increasingly strong reliance on competition as a prime mode of social organization
  - Competition as a core concept for designing social institutions
- Six Working Packages



Research background: SPACE project

Research background: SPACE project

- SPACE Working Package 1: Building a conceptual foundation.

Research background: SPACE project

- SPACE Working Package 1: Building a conceptual foundation.
  - How is C defined and conceptualized in economics?

Research background: SPACE project

- SPACE Working Package 1: Building a conceptual foundation.
  - How is C defined and conceptualized in economics?

Research background: SPACE project

- SPACE Working Package 1: Building a conceptual foundation.
  - How is C defined and conceptualized in economics?
  - Which temporal changes has the dominant meaning of C undergone in the post WWII-period?

Research background: SPACE project

- SPACE Working Package 1: Building a conceptual foundation.
  - How is C defined and conceptualized in economics?
  - Which temporal changes has the dominant meaning of C undergone in the post WWII-period?

Research background: SPACE project

- SPACE Working Package 1: Building a conceptual foundation.
  - How is C defined and conceptualized in economics?
  - Which temporal changes has the dominant meaning of C undergone in the post WWII-period?
  - How do major economic conceptions of C differ between economic paradigms?

Aim of this paper

Aim of this paper

- A contribution in identifying and deconstructing the conceptual foundation/evolution of C in economics and its observed changes over time.

Aim of this paper

- A contribution in identifying and deconstructing the conceptual foundation/evolution of C in economics and its observed changes over time.
  - Analysis of economic research literature via **topic modeling**

Aim of this paper

- A contribution in identifying and deconstructing the conceptual foundation/evolution of C in economics and its observed changes over time.
  - Analysis of economic research literature via **topic modeling**
  - Topic models = algorithms that discover the content of large collections of documents via **topics**

Aim of this paper

- A contribution in identifying and deconstructing the conceptual foundation/evolution of C in economics and its observed changes over time.
  - Analysis of economic research literature via **topic modeling**
  - Topic models = algorithms that discover the content of large collections of documents via **topics**
  - **Latent Dirichlet Allocation (LDA)** as a simple and widely used topic model (Blei et al. 2003)

Aim of this paper

Aim of this paper

- Topic modeling has emerged as an established method in answering specific research questions in economics and economic history.

Aim of this paper

- Topic modeling has emerged as an established method in answering specific research questions in economics and economic history.

  Focus on economic literature

Aim of this paper

- Topic modeling has emerged as an established method in answering specific research questions in economics and economic history.

  Focus on economic literature

  - located in a specific journal/s (e.g. Lüdering and Winker 2016; Wehrheim 2019, Ambrosino et al. 2018)

Aim of this paper

- Topic modeling has emerged as an established method in answering specific research questions in economics and economic history.

  Focus on economic literature

  - located in a specific journal/s (e.g. Lüdering and Winker 2016; Wehrheim 2019, Ambrosino et al. 2018)
  - on a more comprehensive level (e.g. whole sub-fields for a given period (Angrist et al. 2017; Jelveh et al. 2015)

Aim of this paper

- Topic modeling has emerged as an established method in answering specific research questions in economics and economic history.

  Focus on economic literature

  - located in a specific journal/s (e.g. Lüdering and Winker 2016; Wehrheim 2019, Ambrosino et al. 2018)
  - on a more comprehensive level (e.g. whole sub-fields for a given period (Angrist et al. 2017; Jelveh et al. 2015)
- In this paper: applying topic modeling to a part of the literature which is *ex ante* constraint by a specific research topic: competition.

Research background: SPACE project

Research background: SPACE project

- SPACE Working Package 1: Building a conceptual foundation.

Research background: SPACE project

- SPACE Working Package 1: Building a conceptual foundation.
  - How is C defined and conceptualized in economics?

Research background: SPACE project

- SPACE Working Package 1: Building a conceptual foundation.
  - How is C defined and conceptualized in economics?
    - *Which latent topics emerge within the literature that relates to competition?*

Research background: SPACE project

- SPACE Working Package 1: Building a conceptual foundation.
  - How is C defined and conceptualized in economics?
    - *Which latent topics emerge within the literature that relates to competition?*
  - Which temporal changes has the dominant meaning of C undergone in the post WWII-period?

Research background: SPACE project

- SPACE Working Package 1: Building a conceptual foundation.
  - How is C defined and conceptualized in economics?
    - *Which latent topics emerge within the literature that relates to competition?*
  - Which temporal changes has the dominant meaning of C undergone in the post WWII-period?
    - *Focusing on the dominant (mainstream) research literature in each period: Are there differences in latent topics over time?*

Research background: SPACE project

- SPACE Working Package 1: Building a conceptual foundation.
  - How is C defined and conceptualized in economics?
    - *Which latent topics emerge within the literature that relates to competition?*
  - Which temporal changes has the dominant meaning of C undergone in the post WWII-period?
    - *Focusing on the dominant (mainstream) research literature in each period: Are there differences in latent topics over time?*
  - How do major economic conceptions of C differ between economic paradigms?

Research background: SPACE project

- SPACE Working Package 1: Building a conceptual foundation.
  - How is C defined and conceptualized in economics?
    - *Which latent topics emerge within the literature that relates to competition?*
  - Which temporal changes has the dominant meaning of C undergone in the post WWII-period?
    - *Focusing on the dominant (mainstream) research literature in each period: Are there differences in latent topics over time?*
  - How do major economic conceptions of C differ between economic paradigms?
    - *Are there latent topics which can be assigned to economic paradigms?*

The LDA model and its main assumptions

Source: Steyvers and Griffiths 2007

The LDA model and its main assumptions

- LDA is best described by its “imaginary random process by which the model assumes the documents arose” (Blei 2012, 78)

Source: Steyvers and Griffiths 2007

The LDA model and its main assumptions

- LDA is best described by its “imaginary random process by which the model assumes the documents arose” (Blei 2012, 78)
- In the model, topics are give **first** and then the documents are generated.

Source: Steyvers and Griffiths 2007

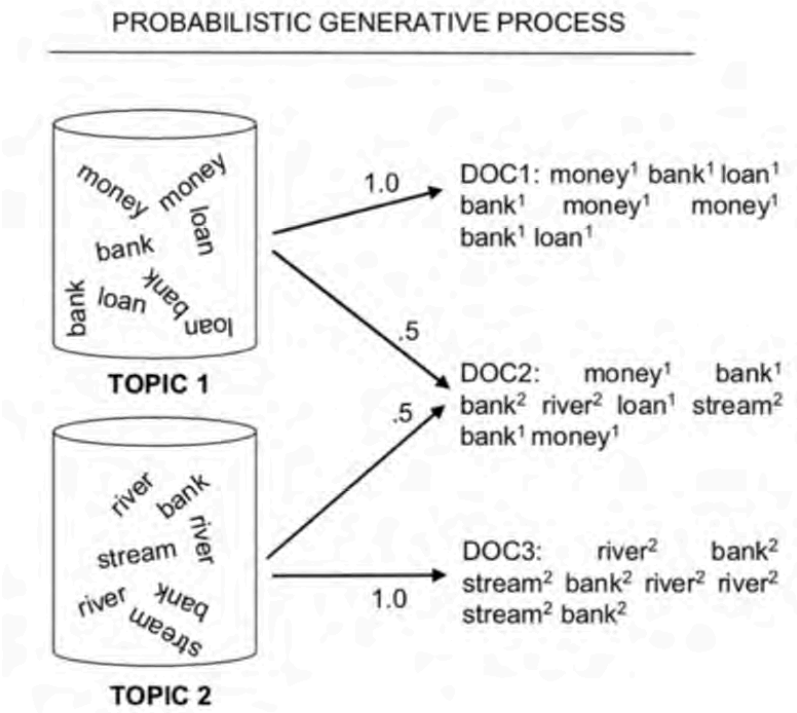
The LDA model and its main assumptions

- LDA is best described by its “imaginary random process by which the model assumes the documents arose” (Blei 2012, 78)
- In the model, topics are give **first** and then the documents are generated.
- Documents = bag of words

Source: Steyvers and Griffiths 2007

The LDA model and its main assumptions

- LDA is best described by its “imaginary random process by which the model assumes the documents arose” (Blei 2012, 78)
- In the model, topics are give **first** and then the documents are generated.
- Documents = bag of words
- For every document in the corpus

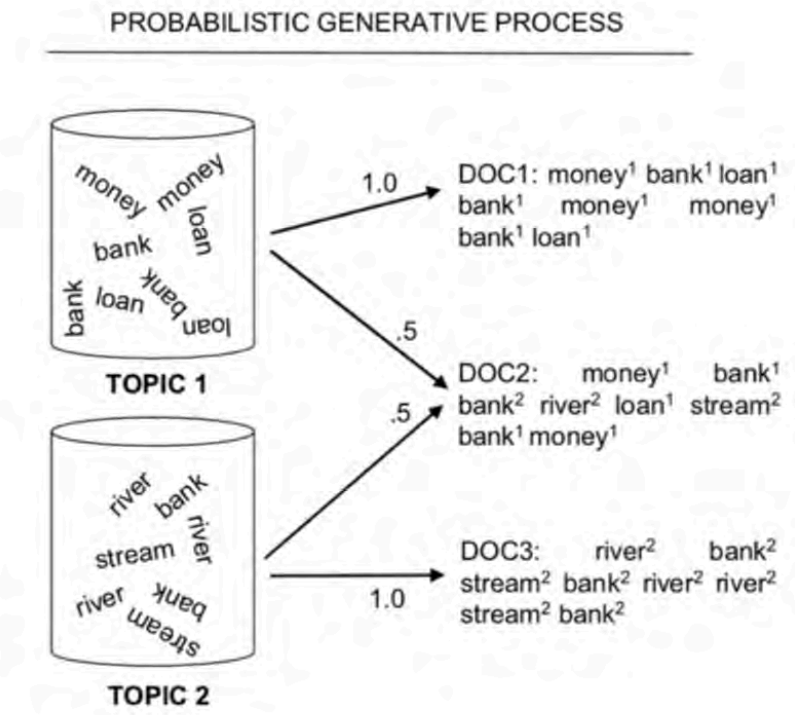


Source: Steyvers and Griffiths 2007

The LDA model and its main assumptions

- LDA is best described by its “imaginary random process by which the model assumes the documents arose” (Blei 2012, 78)
- In the model, topics are give **first** and then the documents are generated.
- Documents = bag of words

- For every document in the corpus
  1. Pick a topic

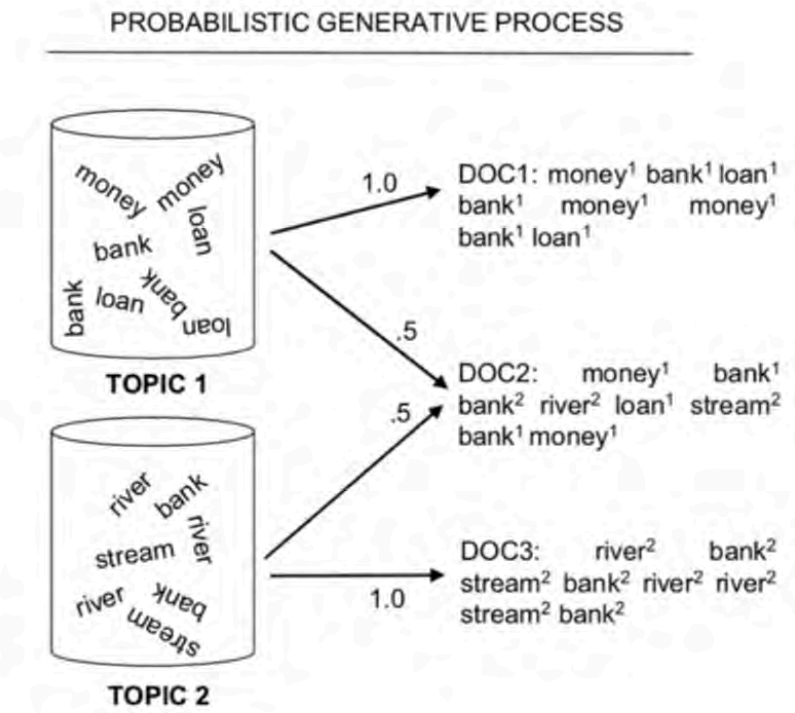


Source: Steyvers and Griffiths 2007

The LDA model and its main assumptions

- LDA is best described by its “imaginary random process by which the model assumes the documents arose” (Blei 2012, 78)
- In the model, topics are give **first** and then the documents are generated.
- Documents = bag of words

- For every document in the corpus
  1. Pick a topic
  2. Pick a word

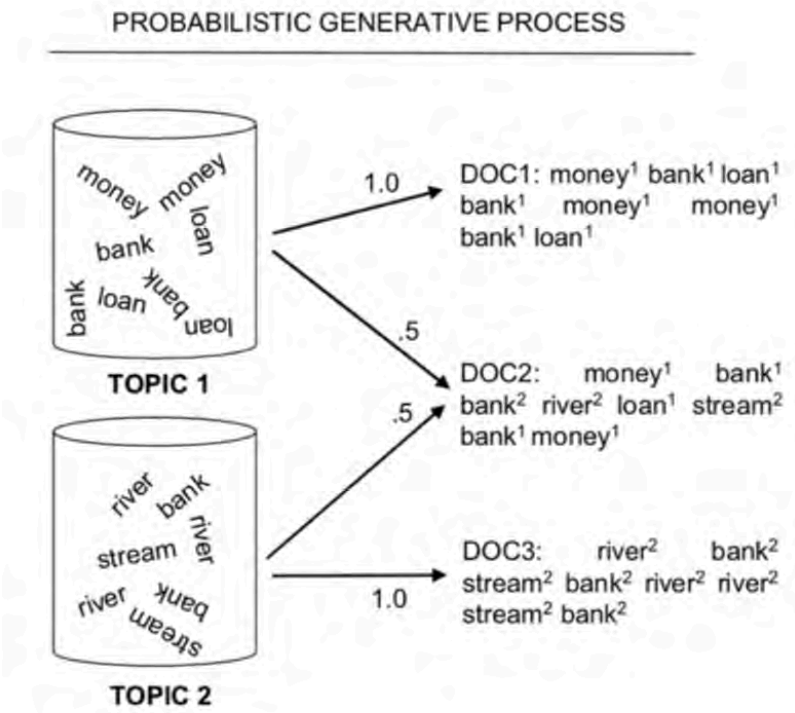


Source: Steyvers and Griffiths 2007

The LDA model and its main assumptions

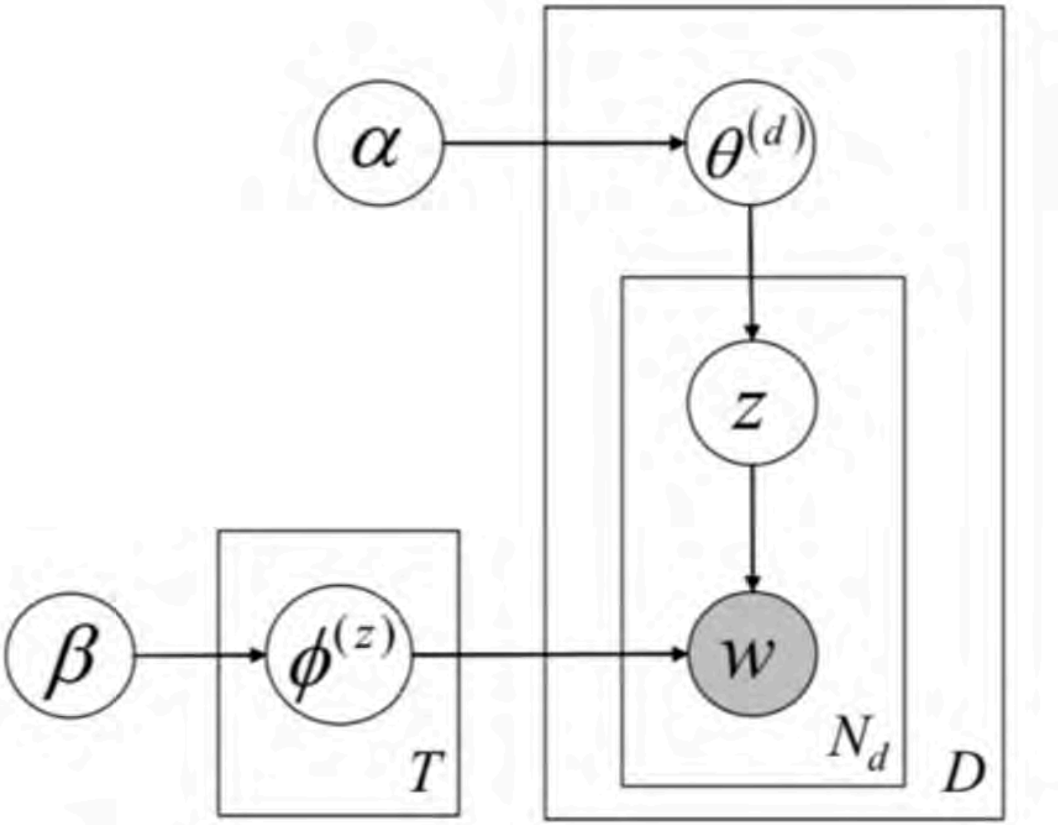
- LDA is best described by its “imaginary random process by which the model assumes the documents arose” (Blei 2012, 78)
- In the model, topics are give **first** and then the documents are generated.
- Documents = bag of words

- For every document in the corpus
  1. Pick a topic
  2. Pick a word
  3. Place it in the bag
     until the document is complete



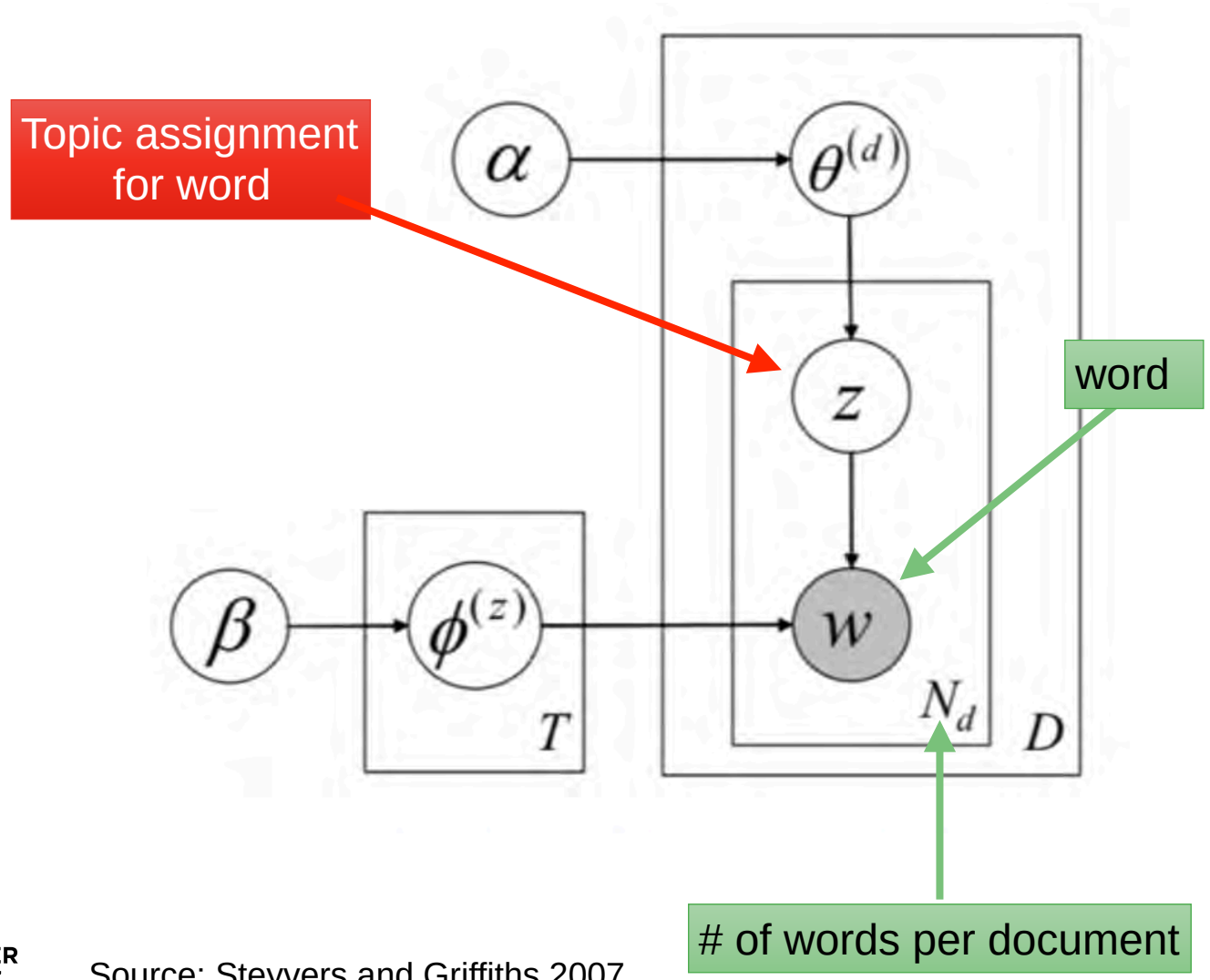
Source: Steyvers and Griffiths 2007

The LDA model and its parameters



Source: Steyvers and Griffiths 2007

The LDA model and its parameters

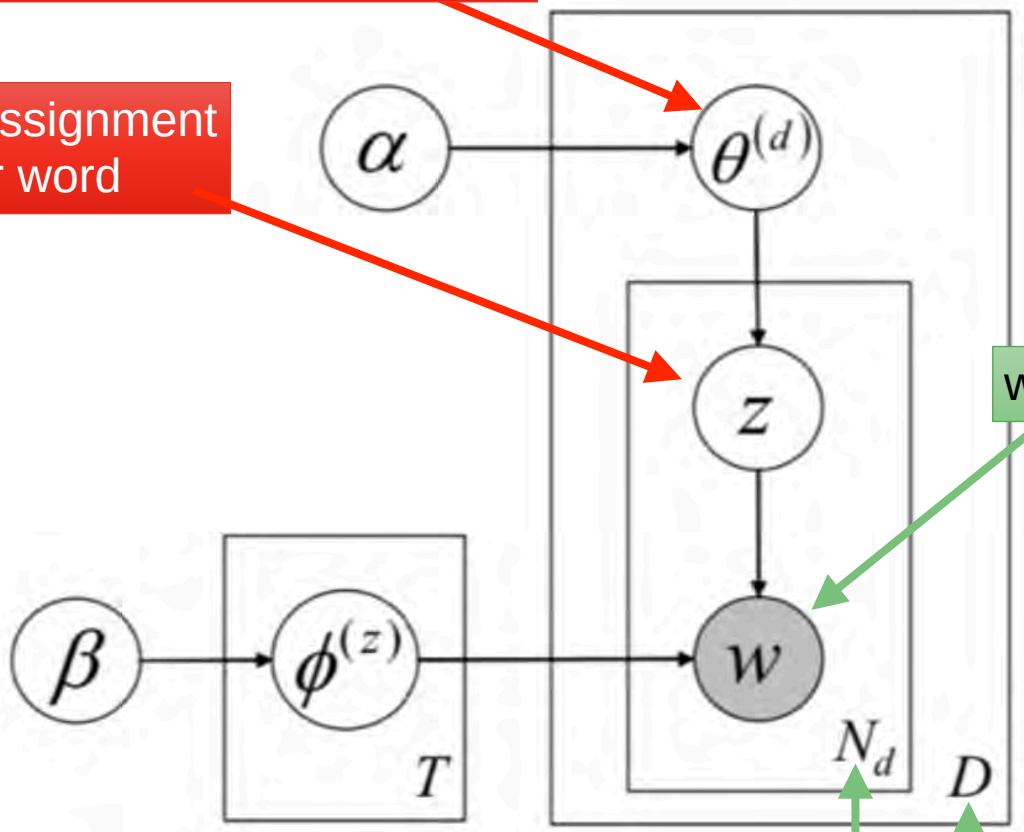


Source: Steyvers and Griffiths 2007

The LDA model and its parameters

topic-document distribtution
(which topics for which document?)

Topic assignment
for word

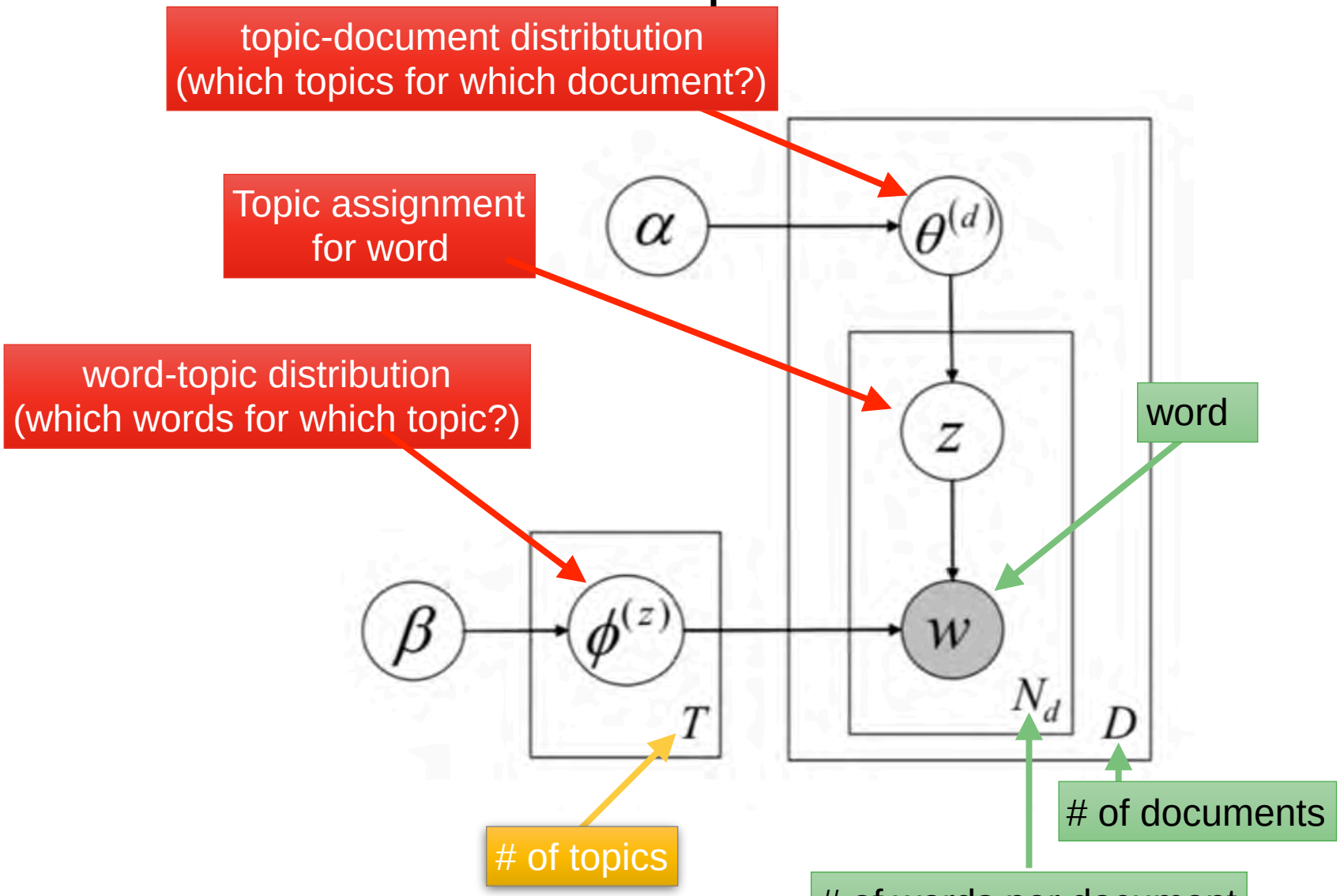


# of documents

# of words per document

Source: Steyvers and Griffiths 2007

The LDA model and its parameters



Source: Steyvers and Griffiths 2007

The LDA model and its parameters

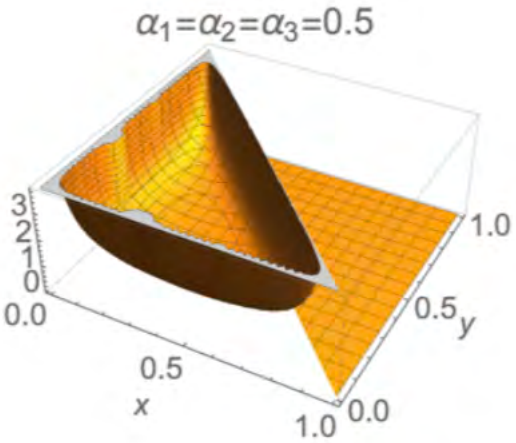
topic-document distribtution
(which topics for which document?)

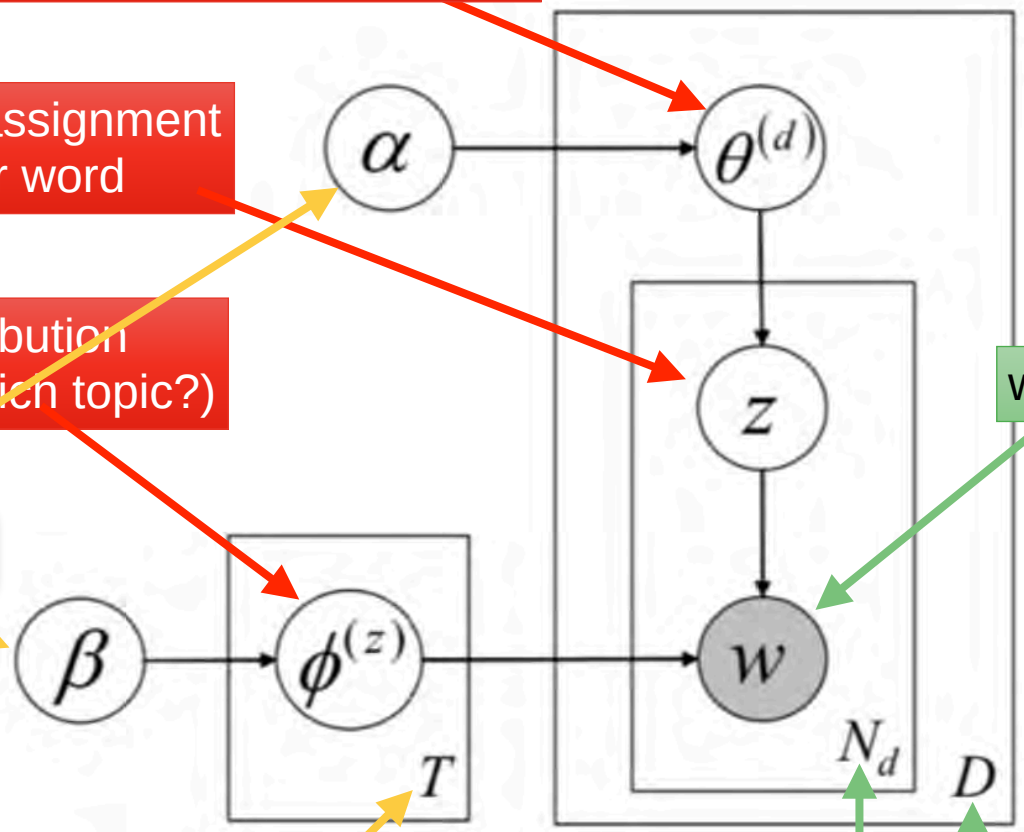
Topic assignment
for word

word-topic distribution
(which words for which topic?)

Hyperparameters (priors)

word





# of topics

# of documents

# of words per document

Source: Steyvers and Griffiths 2007

The LDA model and its parameters

topic-document distribtution
(which topics for which document?)

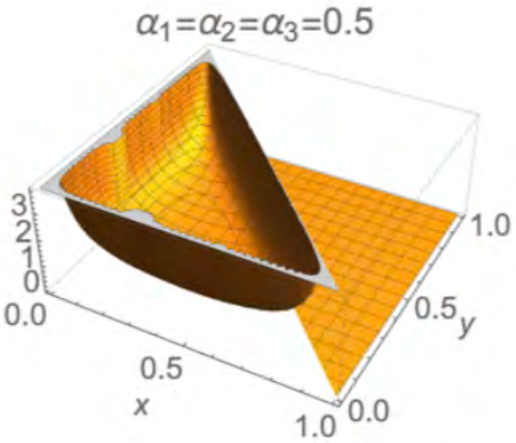
Text preprocessing

Topic assignment
for word

word-topic distribution
(which words for which topic?)

word

Hyperparameters (priors)



# of topics

# of documents

# of words per document

Source: Steyvers and Griffiths 2007

The puzzle of statistical inference

$$p(\theta, \phi, \boldsymbol{z}|\boldsymbol{w}, \alpha, \beta) = \frac{p(\theta, \phi, \boldsymbol{z}, \boldsymbol{w}|\alpha, \beta)}{p(\boldsymbol{w}|\alpha, \beta)}$$

The puzzle of statistical inference

- The posterior (conditional) distribution is

$$p(\theta, \phi, \mathbf{z}|\mathbf{w}, \alpha, \beta) = \frac{p(\theta, \phi, \mathbf{z}, \mathbf{w}|\alpha, \beta)}{p(\mathbf{w}|\alpha, \beta)}$$

The puzzle of statistical inference

- The posterior (conditional) distribution is

$$p(\theta, \phi, \mathbf{z}|\mathbf{w}, \alpha, \beta) = \frac{p(\theta, \phi, \mathbf{z}, \mathbf{w}|\alpha, \beta)}{p(\mathbf{w}|\alpha, \beta)}$$

The puzzle of statistical inference

- The posterior (conditional) distribution is

$$p(\theta, \phi, \mathbf{z}|\mathbf{w}, \alpha, \beta) = \frac{p(\theta, \phi, \mathbf{z}, \mathbf{w}|\alpha, \beta)}{p(\mathbf{w}|\alpha, \beta)}$$

Joint distribution of all hidden and observed variables

The puzzle of statistical inference

- The posterior (conditional) distribution is

$$p(\theta, \phi, \mathbf{z}|\mathbf{w}, \alpha, \beta) = \frac{p(\theta, \phi, \mathbf{z}, \mathbf{w}|\alpha, \beta)}{p(\mathbf{w}|\alpha, \beta)}$$

Joint distribution of all hidden and observed variables

Marginal probability of the observations

The puzzle of statistical inference

- The posterior (conditional) distribution is

$$p(\theta, \phi, \mathbf{z}|\mathbf{w}, \alpha, \beta) = \frac{p(\theta, \phi, \mathbf{z}, \mathbf{w}|\alpha, \beta)}{p(\mathbf{w}|\alpha, \beta)}$$

Joint distribution of all hidden and observed variables

Marginal probability of the observations

"[The marginal probability] is the probability of seeing the observed corpus under any topic model." (Blei 2012, 81)

The puzzle of statistical inference

- The posterior (conditional) distribution is

$$p(\theta, \phi, \mathbf{z}|\mathbf{w}, \alpha, \beta) = \frac{p(\theta, \phi, \mathbf{z}, \mathbf{w}|\alpha, \beta)}{p(\mathbf{w}|\alpha, \beta)}$$

Joint distribution of all hidden and observed variables

Marginal probability of the observations

"[The marginal probability] is the probability of seeing the observed corpus under any topic model." (Blei 2012, 81)

➡ Summing up all possible ways of assigning each observed word (usually in the order of millions!) to one of the topics

The puzzle of statistical inference

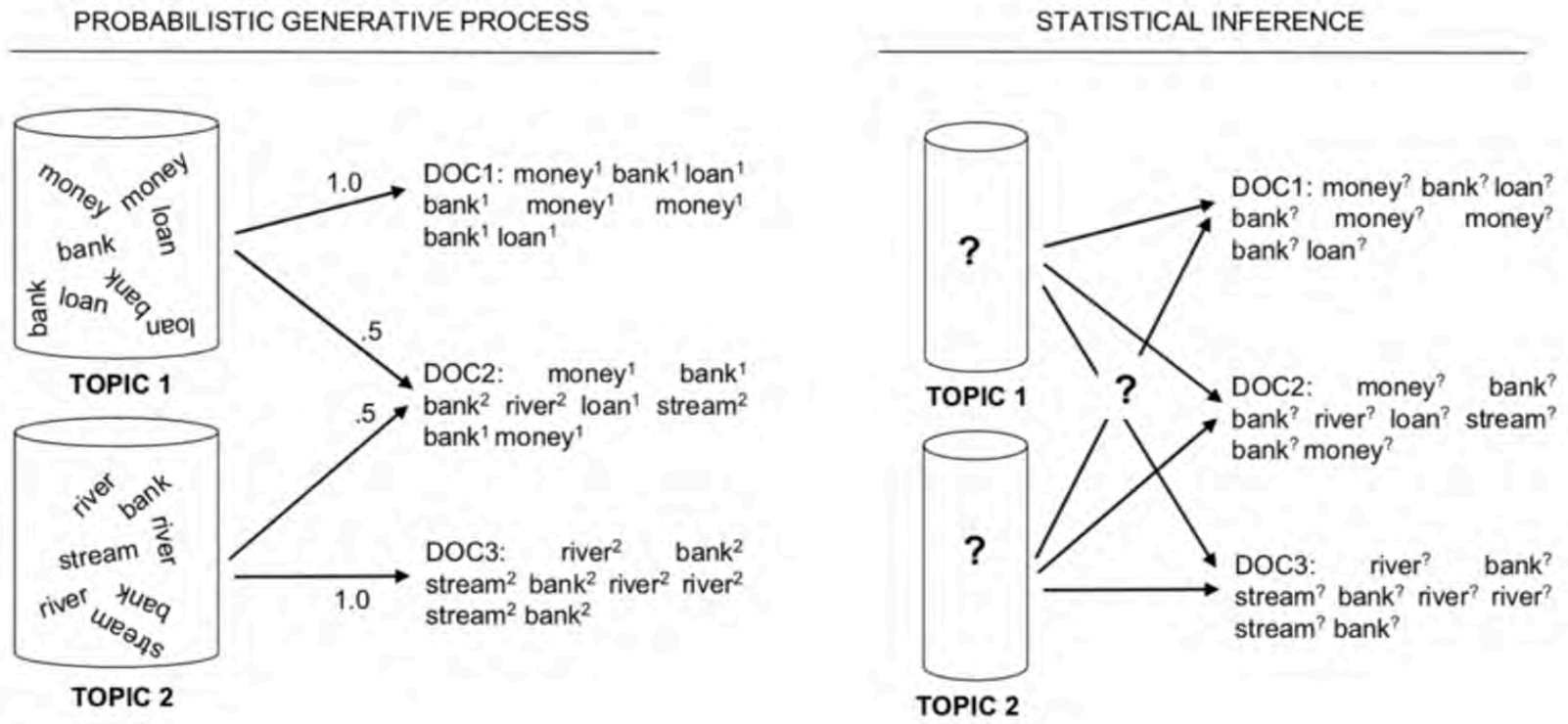


Figure 3: Illustration of the generative process of LDA and the corresponding puzzle of statistically inferring the topic structure. Source: [Steyvers and Griffiths 2007](), 3.

The puzzle of statistical inference

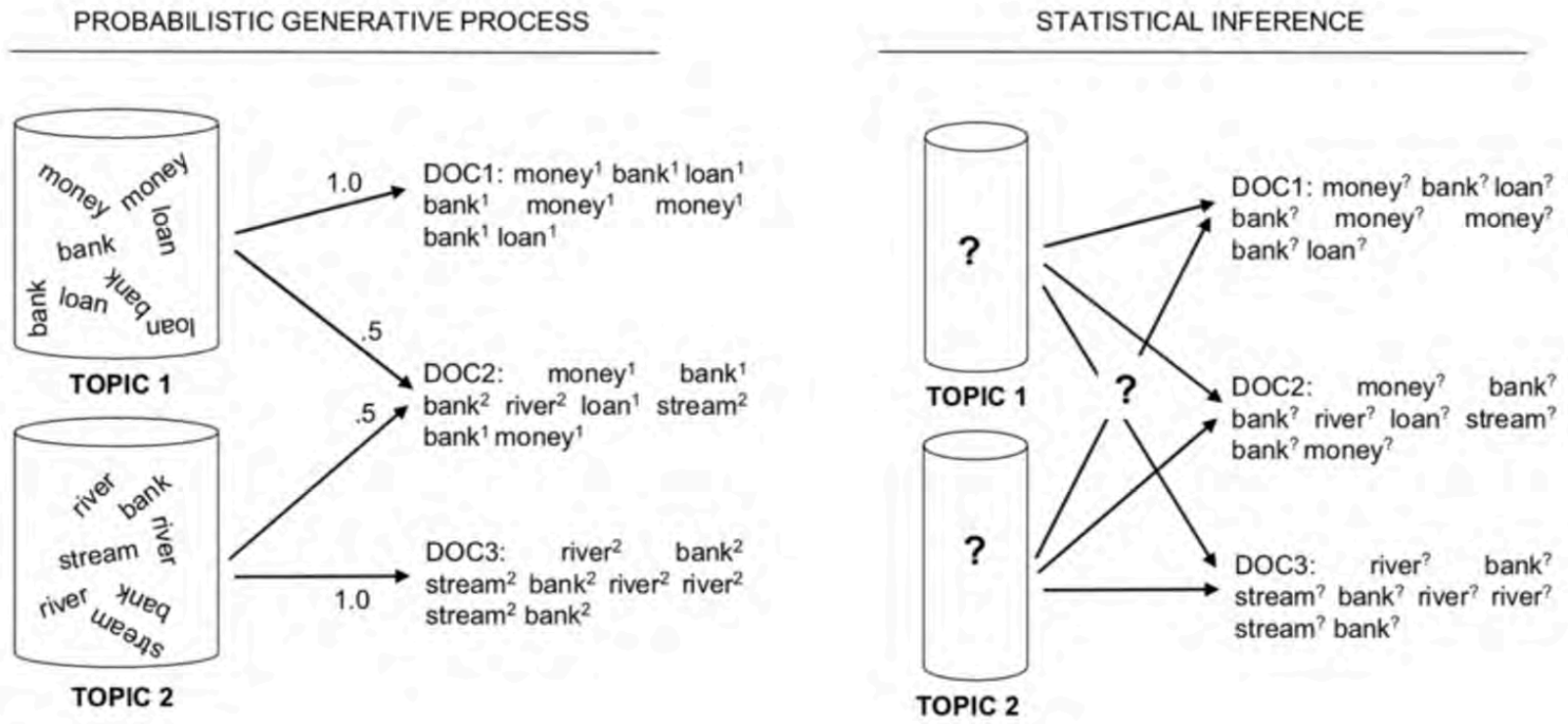


Figure 3: Illustration of the generative process of LDA and the corresponding puzzle of statistically inferring the topic structure. Source: [Steyvers and Griffiths 2007](), 3.

The puzzle of statistical inference

The puzzle of statistical inference

- Collapsed Gibbs sampling (CGS) as inference algorithm (Griffiths and Steyvers 2004)

The puzzle of statistical inference

- Collapsed Gibbs sampling (CGS) as inference algorithm (Griffiths and Steyvers 2004)
- Markov Chain Monte Carlo (MCMC) based method

The puzzle of statistical inference

- Collapsed Gibbs sampling (CGS) as inference algorithm (Griffiths and Steyvers 2004)
- Markov Chain Monte Carlo (MCMC) based method
- Inferring the conditional distribution via estimating z (assignments of words to topics)

The puzzle of statistical inference

- Collapsed Gibbs sampling (CGS) as inference algorithm (Griffiths and Steyvers 2004)
- Markov Chain Monte Carlo (MCMC) based method
- Inferring the conditional distribution via estimating z (assignments of words to topics)

The puzzle of statistical inference

- Collapsed Gibbs sampling (CGS) as inference algorithm (Griffiths and Steyvers 2004)
- Markov Chain Monte Carlo (MCMC) based method
- Inferring the conditional distribution via estimating z (assignments of words to topics)

$$P(z_i = j | \mathbf{z}_{-i}, w_i, d_i, \cdot) \propto \frac{C_{w_i j}^{WT} + \beta}{\sum_{w=1}^{W} C_{wj}^{WT} + W\beta} \frac{C_{d_i j}^{DT} + \alpha}{\sum_{t=1}^{T} C_{d_i t}^{DT} + T\alpha}$$

The puzzle of statistical inference

- Collapsed Gibbs sampling (CGS) as inference algorithm (Griffiths and Steyvers 2004)
- Markov Chain Monte Carlo (MCMC) based method
- Inferring the conditional distribution via estimating z (assignments of words to topics)

$$P(z_i = j | \mathbf{z}_{-i}, w_i, d_i, \cdot) \propto \frac{C_{w_i j}^{WT} + \beta}{\sum_{w=1}^{W} C_{wj}^{WT} + W\beta} \frac{C_{d_i j}^{DT} + \alpha}{\sum_{t=1}^{T} C_{d_i t}^{DT} + T\alpha}$$

The puzzle of statistical inference

- Collapsed Gibbs sampling (CGS) as inference algorithm (Griffiths and Steyvers 2004)
- Markov Chain Monte Carlo (MCMC) based method
- Inferring the conditional distribution via estimating z (assignments of words to topics)

$$P(z_i = j | \mathbf{z}_{-i}, w_i, d_i, \cdot) \propto \frac{C_{w_i j}^{WT} + \beta}{\sum_{w=1}^{W} C_{wj}^{WT} + W\beta} \frac{C_{d_i j}^{DT} + \alpha}{\sum_{t=1}^{T} C_{d_i t}^{DT} + T\alpha}$$

The puzzle of statistical inference

- Collapsed Gibbs sampling (CGS) as inference algorithm (Griffiths and Steyvers 2004)
- Markov Chain Monte Carlo (MCMC) based method
- Inferring the conditional distribution via estimating z (assignments of words to topics)

$$P(z_i = j | \mathbf{z}_{-i}, w_i, d_i, \cdot) \propto \frac{C^{WT}_{w_i j} + \beta}{\sum_{w=1}^{W} C^{WT}_{wj} + W\beta} \frac{C^{DT}_{d_i j} + \alpha}{\sum_{t=1}^{T} C^{DT}_{d_i t} + T\alpha}$$

- Calculating ϕ (word-topic distribution)

$$\phi'^{(j)}_i = \frac{C^{WT}_{ij} + \beta}{\sum_{k=1}^{W} C^{WT}_{kj} + W\beta}$$

The puzzle of statistical inference

- Collapsed Gibbs sampling (CGS) as inference algorithm (Griffiths and Steyvers 2004)
- Markov Chain Monte Carlo (MCMC) based method
- Inferring the conditional distribution via estimating z (assignments of words to topics)

$$P(z_i = j | \mathbf{z}_{-i}, w_i, d_i, \cdot) \propto \frac{C^{WT}_{w_i j} + \beta}{\sum_{w=1}^{W} C^{WT}_{wj} + W\beta} \frac{C^{DT}_{d_i j} + \alpha}{\sum_{t=1}^{T} C^{DT}_{d_i t} + T\alpha}$$

- Calculating ϕ (word-topic distribution)

$$\phi'^{(j)}_i = \frac{C^{WT}_{ij} + \beta}{\sum_{k=1}^{W} C^{WT}_{kj} + W\beta}$$

- Calculating θ (topic-document distribution)

$$\theta'^{(d)}_j = \frac{C^{DT}_{dj} + \alpha}{\sum_{k=1}^{T} C^{DT}_{dk} + T\alpha}$$

Data

Data

- Database: JSTOR Data for Research

Data

- Database: JSTOR Data for Research
- Search term “competition“

Data

- Database: JSTOR Data for Research
- Search term “competition“
- 124.749 items published in 260 journals between 1851-2017

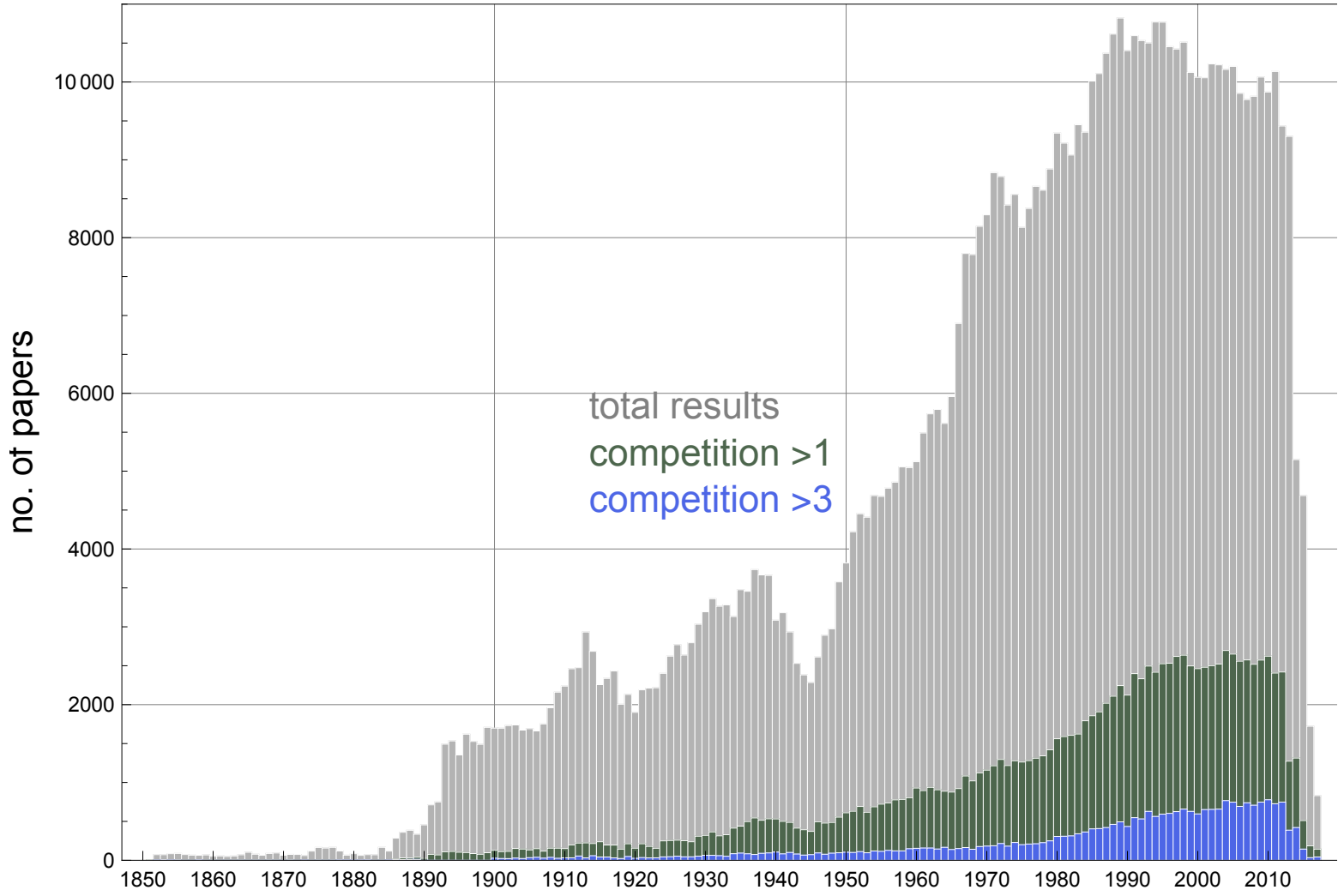
Data

- Database: JSTOR Data for Research
- Search term “competition“
- 124.749 items published in 260 journals between 1851-2017
- Full text ngram counts + bibliometric meta-data

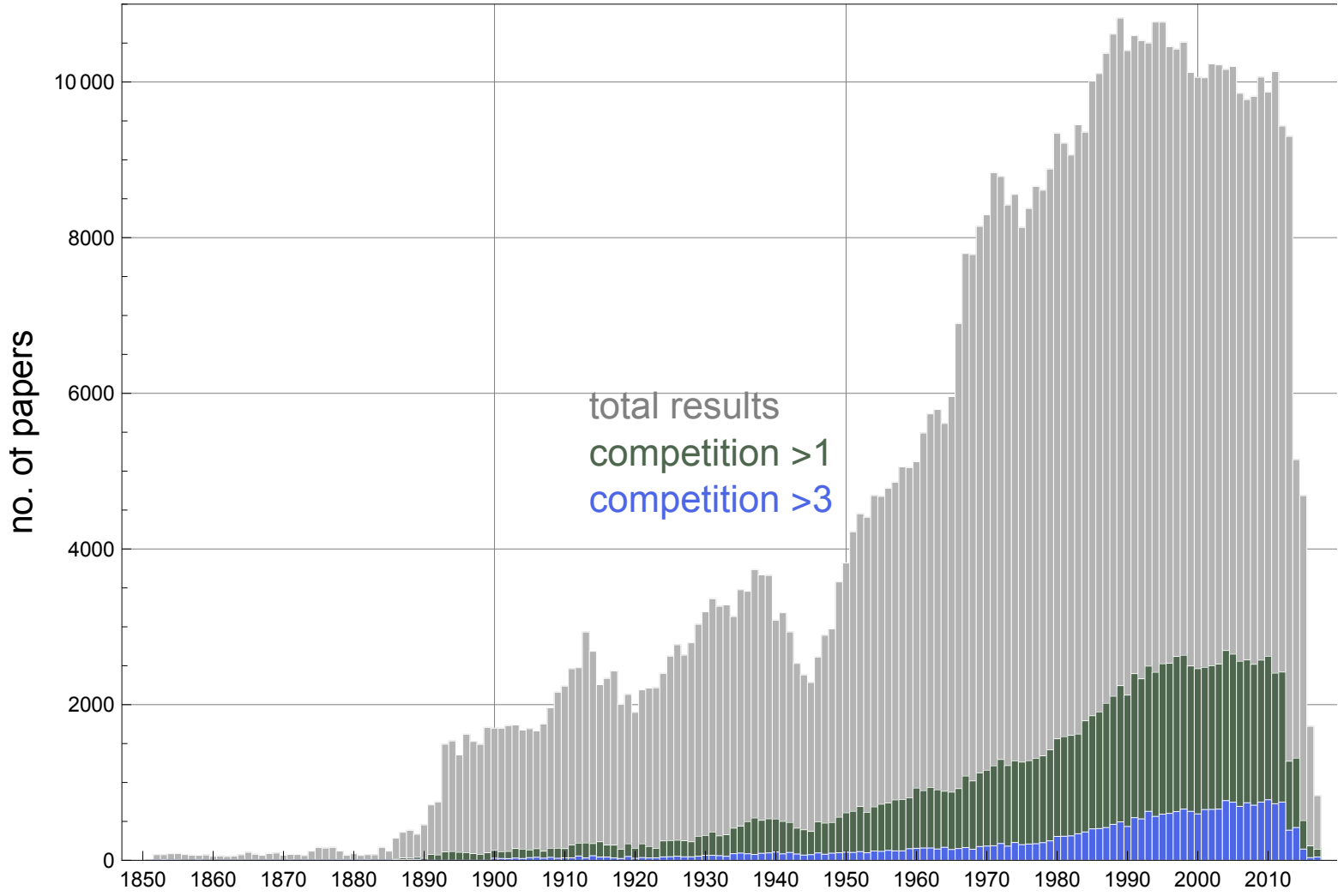
Data

- Database: JSTOR Data for Research
- Search term “competition“
- 124.749 items published in 260 journals between 1851-2017
- Full text ngram counts + bibliometric meta-data
- “Narrow” sample: items containing “competition” > 3
  - 27.488 articles
  - 227 journals (thereof 34 heterodox journals)
  - Text preprocessing + “pseudo” abstracts as final input for the model
  - 98.572 unique words (5.496.608 total words in the corpus)

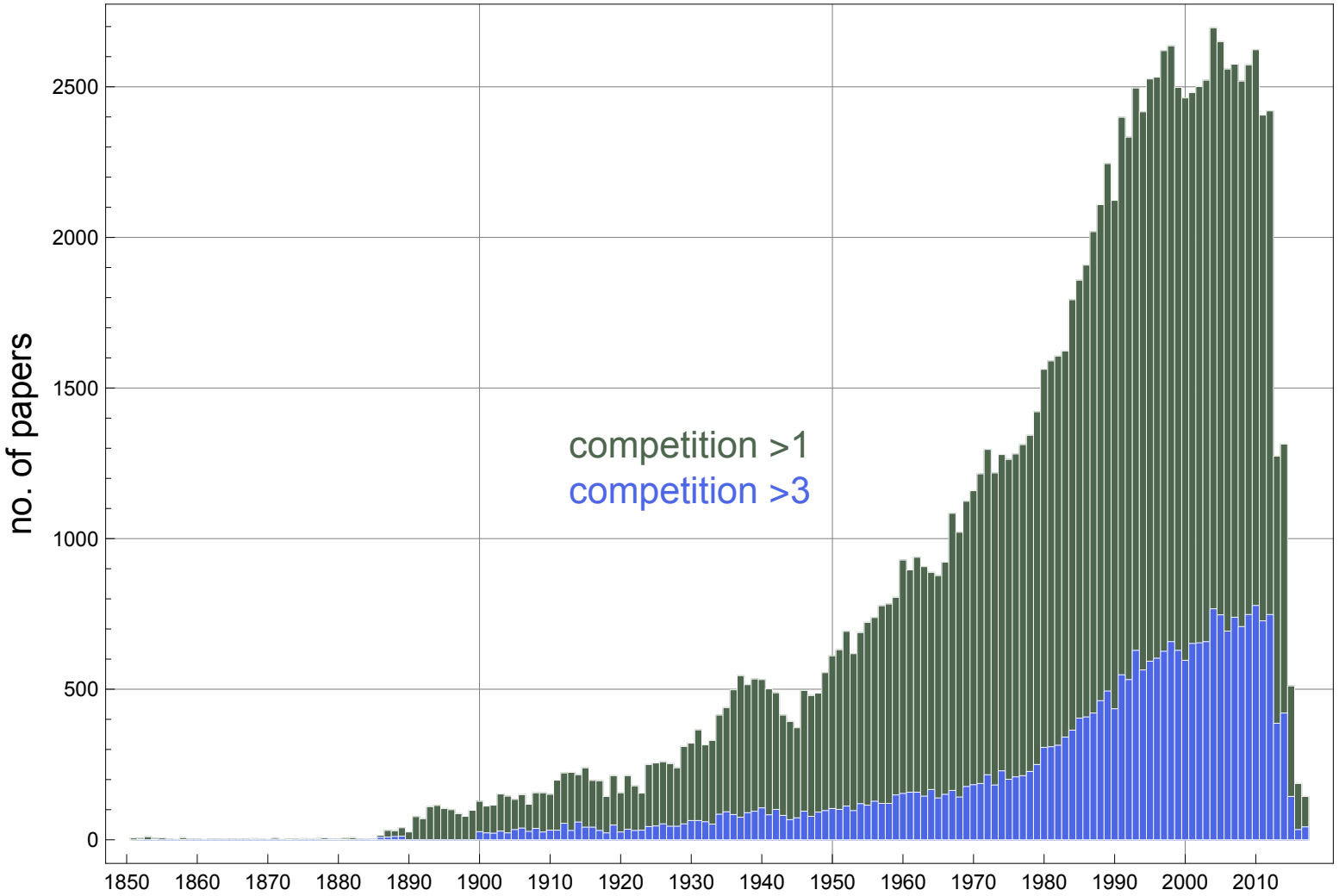
(preliminary) results I: descriptives



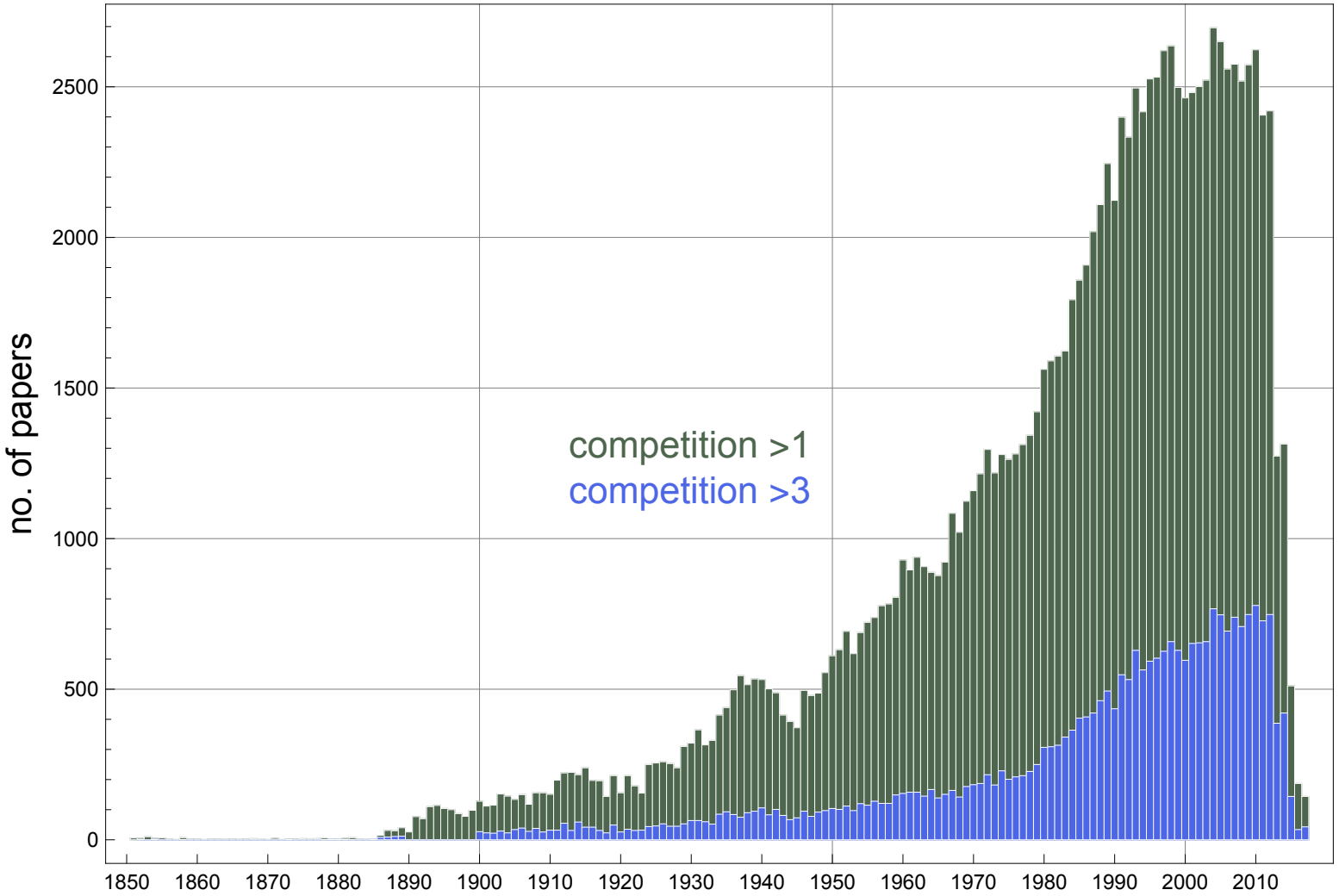
(preliminary) results I: descriptives



(preliminary) results I: descriptives



(preliminary) results I: descriptives



(preliminary) results I: descriptives

Table 1. The 20 most frequent journals in the data sample.

	journal (listed in JSTOR)	no. of articles (27.488)	relative share (100%)	cumulative share (100%)
1	The American Economic Review	1679	6.1	6.1
2	Economic and Political Weekly	1036	3.8	9.9
3	The Economic Journal	871	3.2	13.1
4	The Quarterly Journal of Economics	773	2.8	15.9
5	The Journal of Industrial Economics	739	2.7	18.6
6	Journal of Political Economy	711	2.6	21.2
7	Public Choice	710	2.6	23.8
8	Ann. Am. Acad. Political Soc. Sci.	700	2.5	26.3
9	Southern Economic Journal	640	2.3	28.6
10	The RAND Journal of Economics	602	2.2	30.8
11	Journal of Economic Issues	463	1.7	32.5
12	The Journal of Law & Economics	442	1.6	34.1
13	The Canadian Journal of Economics	437	1.6	35.7
14	Journal of Farm Economics	393	1.4	37.1
15	The Review of Economic Studies	371	1.3	38.4
16	Managerial and Decision Economics	357	1.3	39.7
17	American Journal of Agricultural Economics	354	1.3	41.0
18	Cambridge Journal of Economics	349	1.3	42.3
19	The Review of Economics and Statistics	348	1.3	43.6
20	The Business History Review	335	1.2	44.8

(preliminary) results I: descriptives

Table 1. The 20 most frequent journals in the data sample.

	journal (listed in JSTOR)	no. of articles (27.488)	relative share (100%)	cumulative share (100%)
1	The American Economic Review	1679	6.1	6.1
2	Economic and Political Weekly	1036	3.8	9.9
3	The Economic Journal	871	3.2	13.1
4	The Quarterly Journal of Economics	773	2.8	15.9
5	The Journal of Industrial Economics	739	2.7	18.6
6	Journal of Political Economy	711	2.6	21.2
7	Public Choice	710	2.6	23.8
8	Ann. Am. Acad. Political Soc. Sci.	700	2.5	26.3
9	Southern Economic Journal	640	2.3	28.6
10	The RAND Journal of Economics	602	2.2	30.8
11	Journal of Economic Issues	463	1.7	32.5
12	The Journal of Law & Economics	442	1.6	34.1
13	The Canadian Journal of Economics	437	1.6	35.7
14	Journal of Farm Economics	393	1.4	37.1
15	The Review of Economic Studies	371	1.3	38.4
16	Managerial and Decision Economics	357	1.3	39.7
17	American Journal of Agricultural Economics	354	1.3	41.0
18	Cambridge Journal of Economics	349	1.3	42.3
19	The Review of Economics and Statistics	348	1.3	43.6
20	The Business History Review	335	1.2	44.8

(preliminary) results I: descriptives

4 out of the 'Big 5'

Table 1. The 20 most frequent journals in the data sample.

	journal (listed in JSTOR)	no. of articles (27.488)	relative share (100%)	cumulative share (100%)
1	The American Economic Review	1679	6.1	6.1
2	Economic and Political Weekly	1036	3.8	9.9
3	The Economic Journal	871	3.2	13.1
4	The Quarterly Journal of Economics	773	2.8	15.9
5	The Journal of Industrial Economics	739	2.7	18.6
6	Journal of Political Economy	711	2.6	21.2
7	Public Choice	710	2.6	23.8
8	Ann. Am. Acad. Political Soc. Sci.	700	2.5	26.3
9	Southern Economic Journal	640	2.3	28.6
10	The RAND Journal of Economics	602	2.2	30.8
11	Journal of Economic Issues	463	1.7	32.5
12	The Journal of Law & Economics	442	1.6	34.1
13	The Canadian Journal of Economics	437	1.6	35.7
14	Journal of Farm Economics	393	1.4	37.1
15	The Review of Economic Studies	371	1.3	38.4
16	Managerial and Decision Economics	357	1.3	39.7
17	American Journal of Agricultural Economics	354	1.3	41.0
18	Cambridge Journal of Economics	349	1.3	42.3
19	The Review of Economics and Statistics	348	1.3	43.6
20	The Business History Review	335	1.2	44.8

(preliminary) results I: descriptives

4 out of the 'Big 5'

HET journals

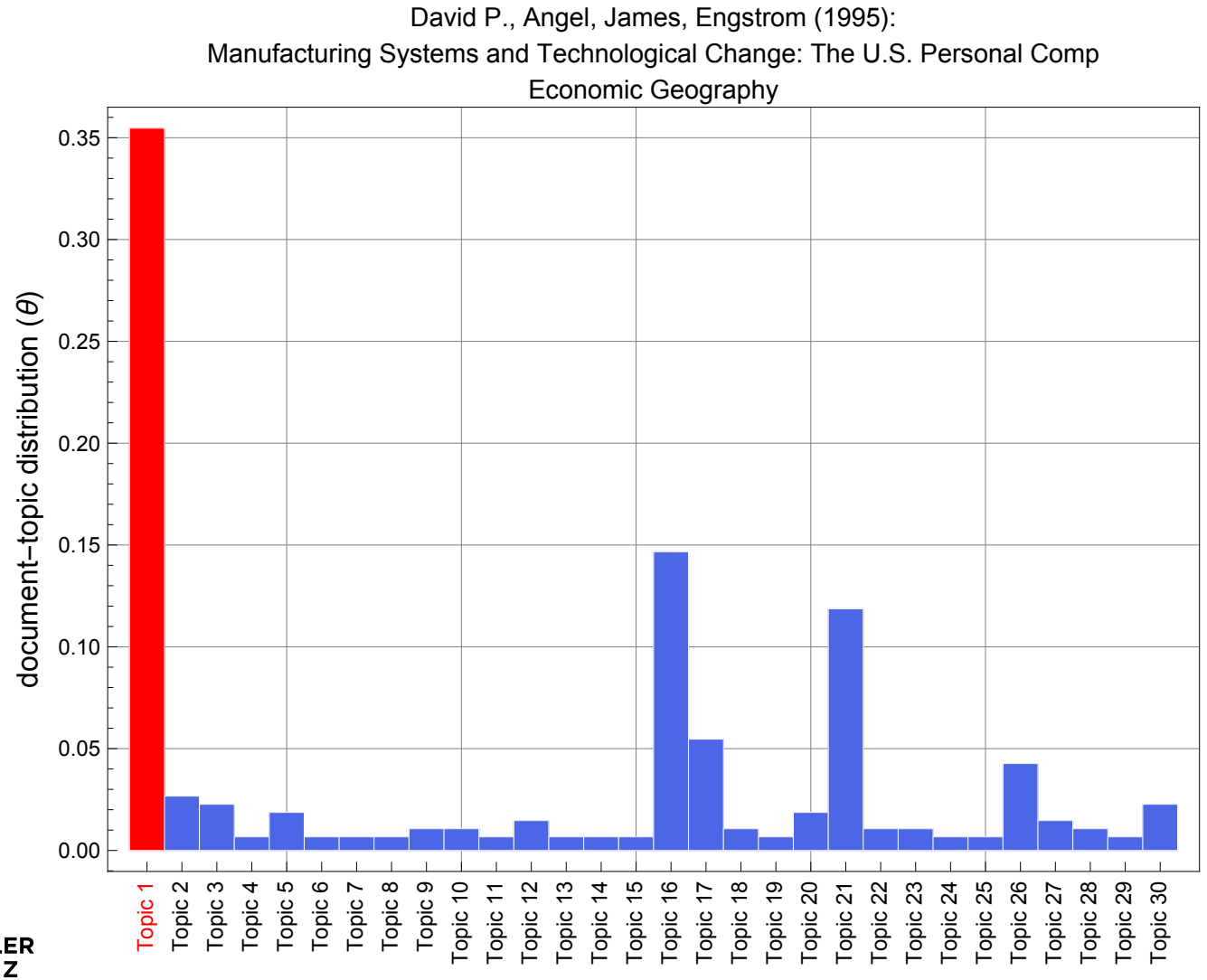
Table 1. The 20 most frequent journals in the data sample.

	journal (listed in JSTOR)	no. of articles (27.488)	relative share (100%)	cumulative share (100%)
1	The American Economic Review	1679	6.1	6.1
2	Economic and Political Weekly	1036	3.8	9.9
3	The Economic Journal	871	3.2	13.1
4	The Quarterly Journal of Economics	773	2.8	15.9
5	The Journal of Industrial Economics	739	2.7	18.6
6	Journal of Political Economy	711	2.6	21.2
7	Public Choice	710	2.6	23.8
8	Ann. Am. Acad. Political Soc. Sci.	700	2.5	26.3
9	Southern Economic Journal	640	2.3	28.6
10	The RAND Journal of Economics	602	2.2	30.8
11	Journal of Economic Issues	463	1.7	32.5
12	The Journal of Law & Economics	442	1.6	34.1
13	The Canadian Journal of Economics	437	1.6	35.7
14	Journal of Farm Economics	393	1.4	37.1
15	The Review of Economic Studies	371	1.3	38.4
16	Managerial and Decision Economics	357	1.3	39.7
17	American Journal of Agricultural Economics	354	1.3	41.0
18	Cambridge Journal of Economics	349	1.3	42.3
19	The Review of Economics and Statistics	348	1.3	43.6
20	The Business History Review	335	1.2	44.8

(preliminary) results II: a bird's eye's view on competition in economic research

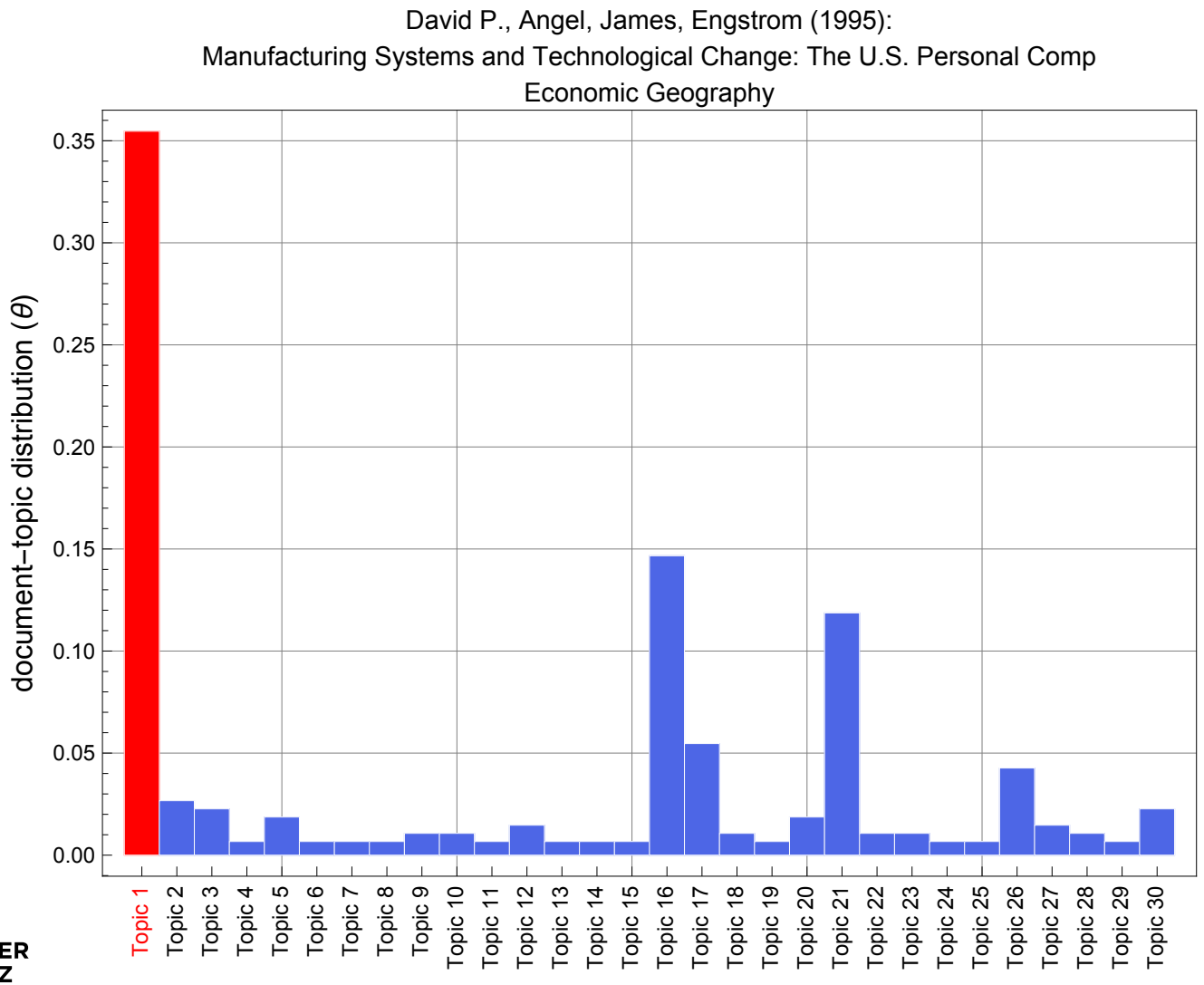
- CGS with overall "narrow" sample
  - Number of topics T=*30*
  - Dirichlet hyperparameters α = *50/T*, β = *0.01*
  - Number of iterations: 500

(preliminary) results II: some examples



(preliminary) results II: some examples

Topic 1	Weight (ϕ)
technology	0.0247
industry	0.0172
innovation	0.0164
firms	0.0158
product	0.0153
technological	0.0151
research	0.0147
production	0.013
process	0.0128
firm	0.0119
investment	0.0113
development	0.0108
industries	0.0104
products	0.0104
industrial	0.0103

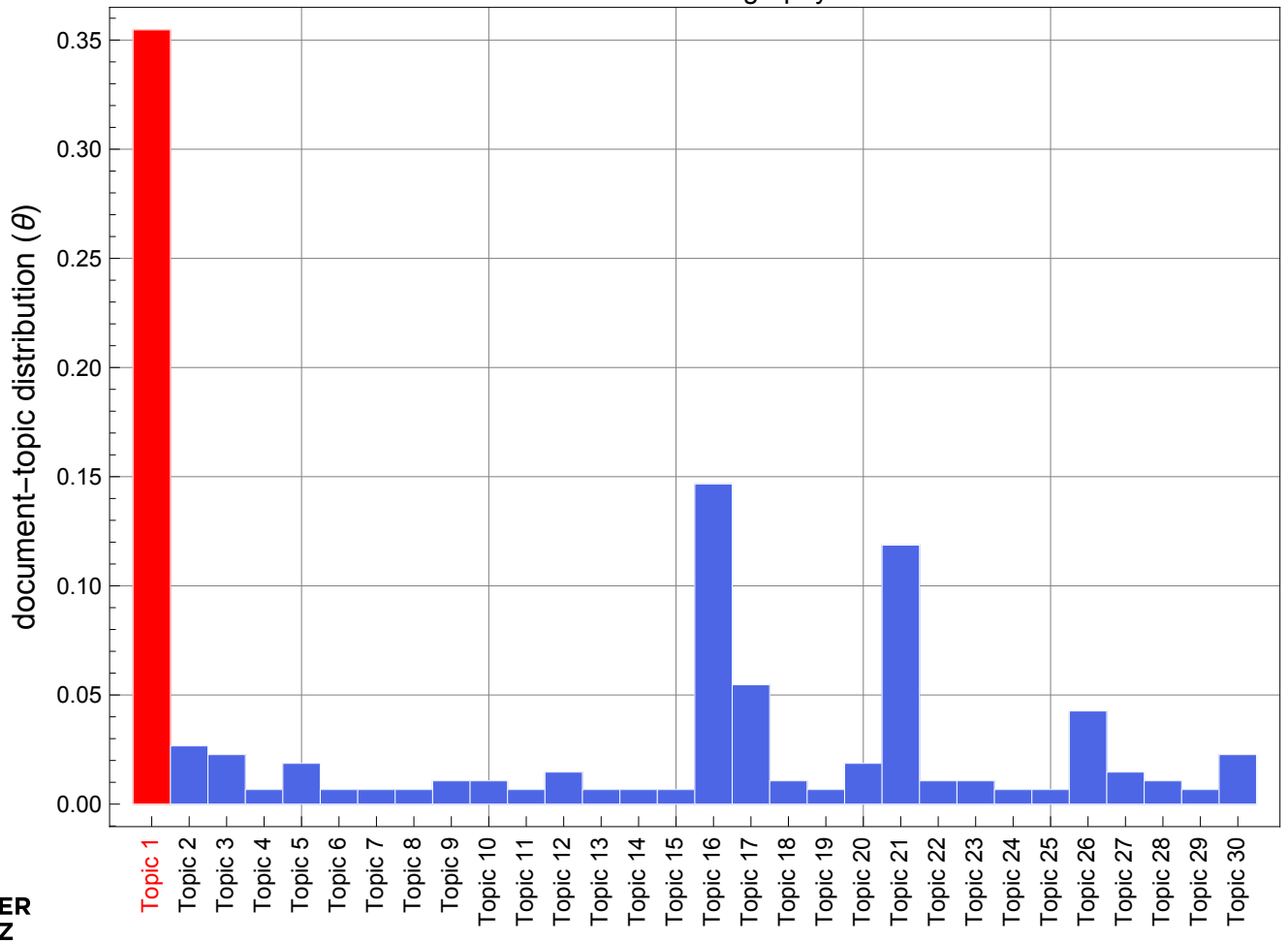


(preliminary) results II: some examples

- Topic1: technological innovation and industrial development

Topic 1	Weight (ϕ)
technology	0.0247
industry	0.0172
innovation	0.0164
firms	0.0158
product	0.0153
technological	0.0151
research	0.0147
production	0.013
process	0.0128
firm	0.0119
investment	0.0113
development	0.0108
industries	0.0104
products	0.0104
industrial	0.0103

David P., Angel, James, Engstrom (1995):
Manufacturing Systems and Technological Change: The U.S. Personal Comp
Economic Geography

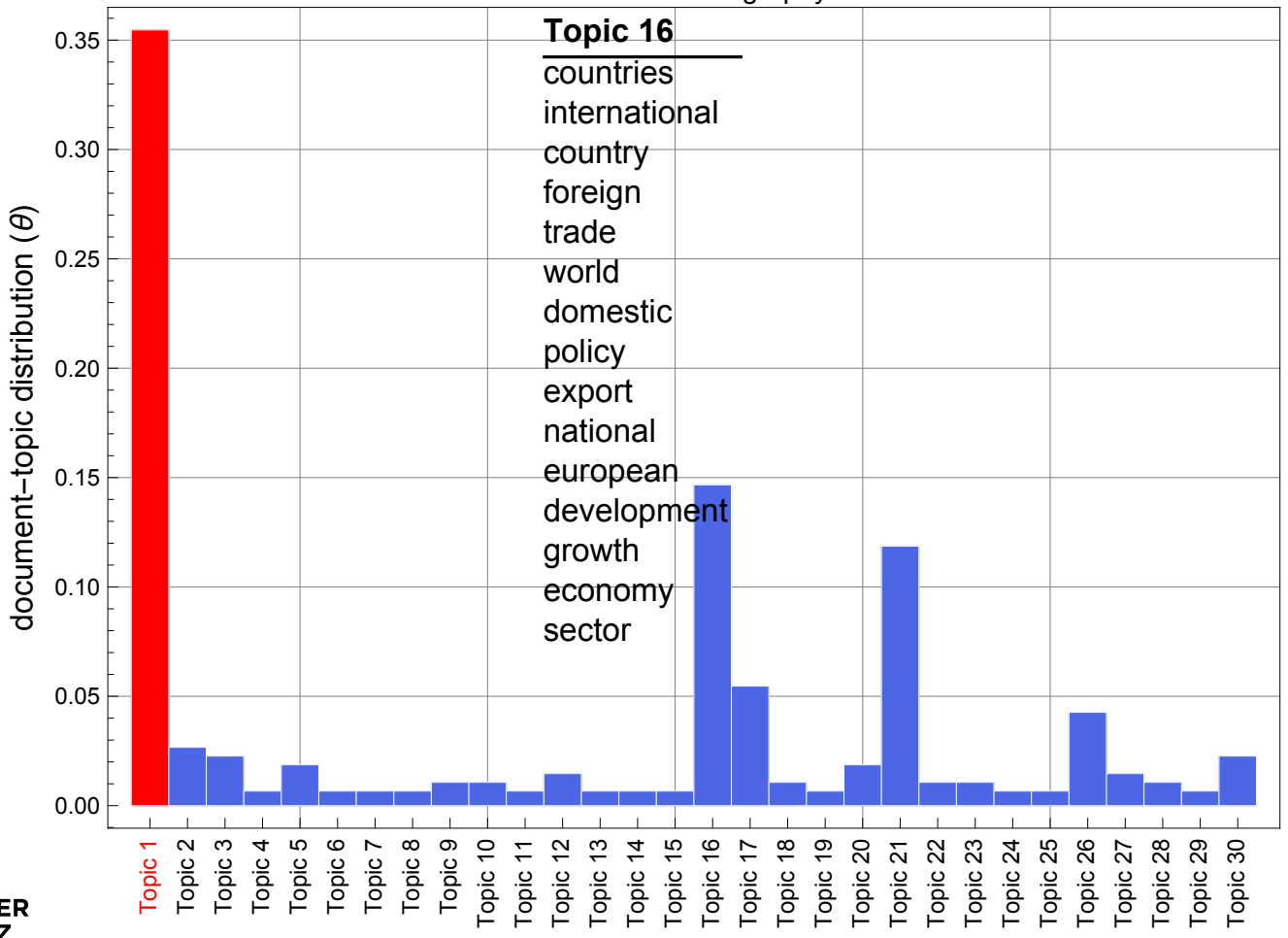


(preliminary) results II: some examples

- Topic1: technological innovation and industrial development

Topic 1	Weight (ϕ)
technology	0.0247
industry	0.0172
innovation	0.0164
firms	0.0158
product	0.0153
technological	0.0151
research	0.0147
production	0.013
process	0.0128
firm	0.0119
investment	0.0113
development	0.0108
industries	0.0104
products	0.0104
industrial	0.0103

David P., Angel, James, Engstrom (1995):
Manufacturing Systems and Technological Change: The U.S. Personal Comp
Economic Geography

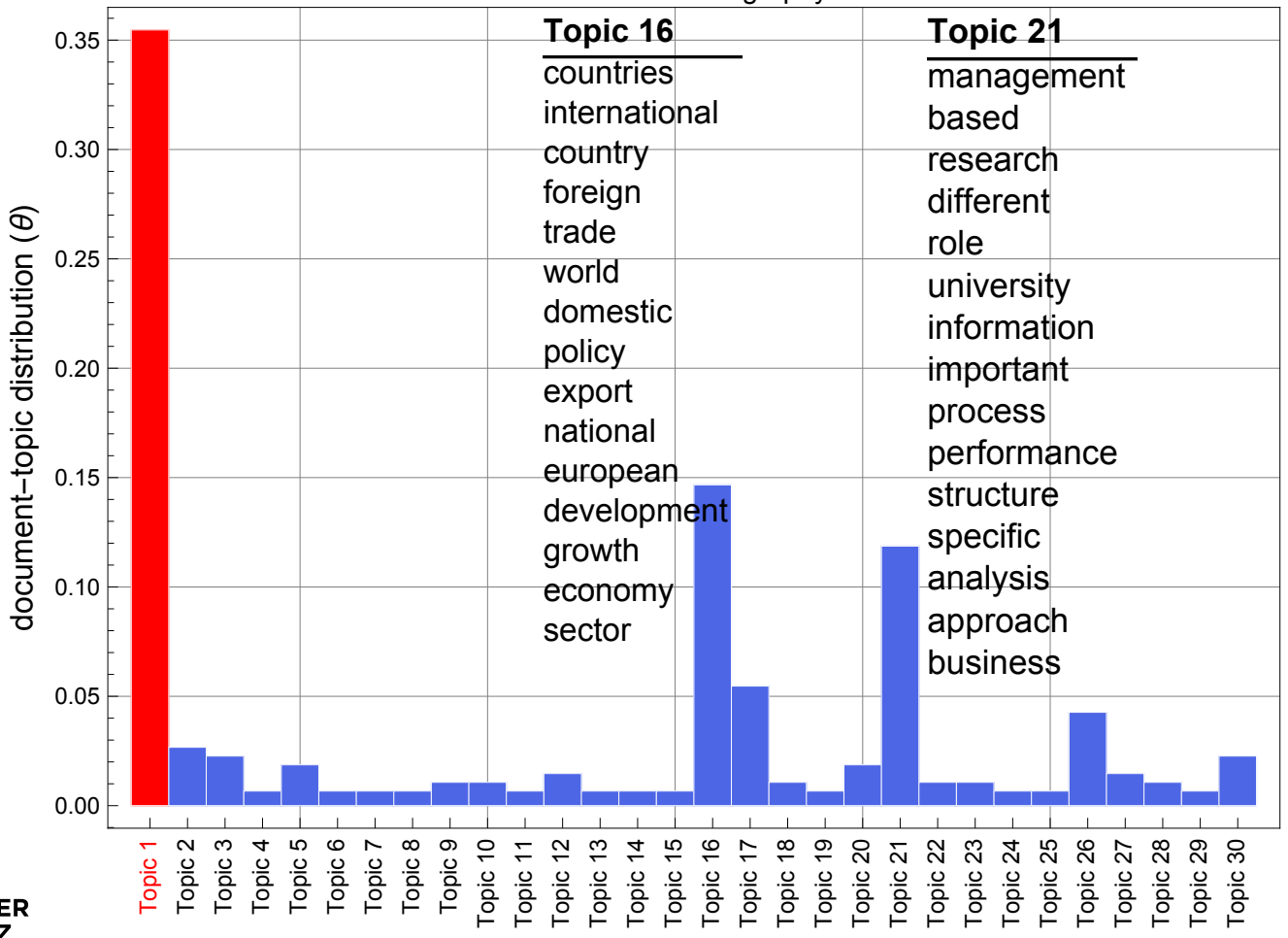


(preliminary) results II: some examples

- Topic1: technological innovation and industrial development

Topic 1	Weight (ϕ)
technology	0.0247
industry	0.0172
innovation	0.0164
firms	0.0158
product	0.0153
technological	0.0151
research	0.0147
production	0.013
process	0.0128
firm	0.0119
investment	0.0113
development	0.0108
industries	0.0104
products	0.0104
industrial	0.0103

David P., Angel, James, Engstrom (1995):
Manufacturing Systems and Technological Change: The U.S. Personal Comp
Economic Geography

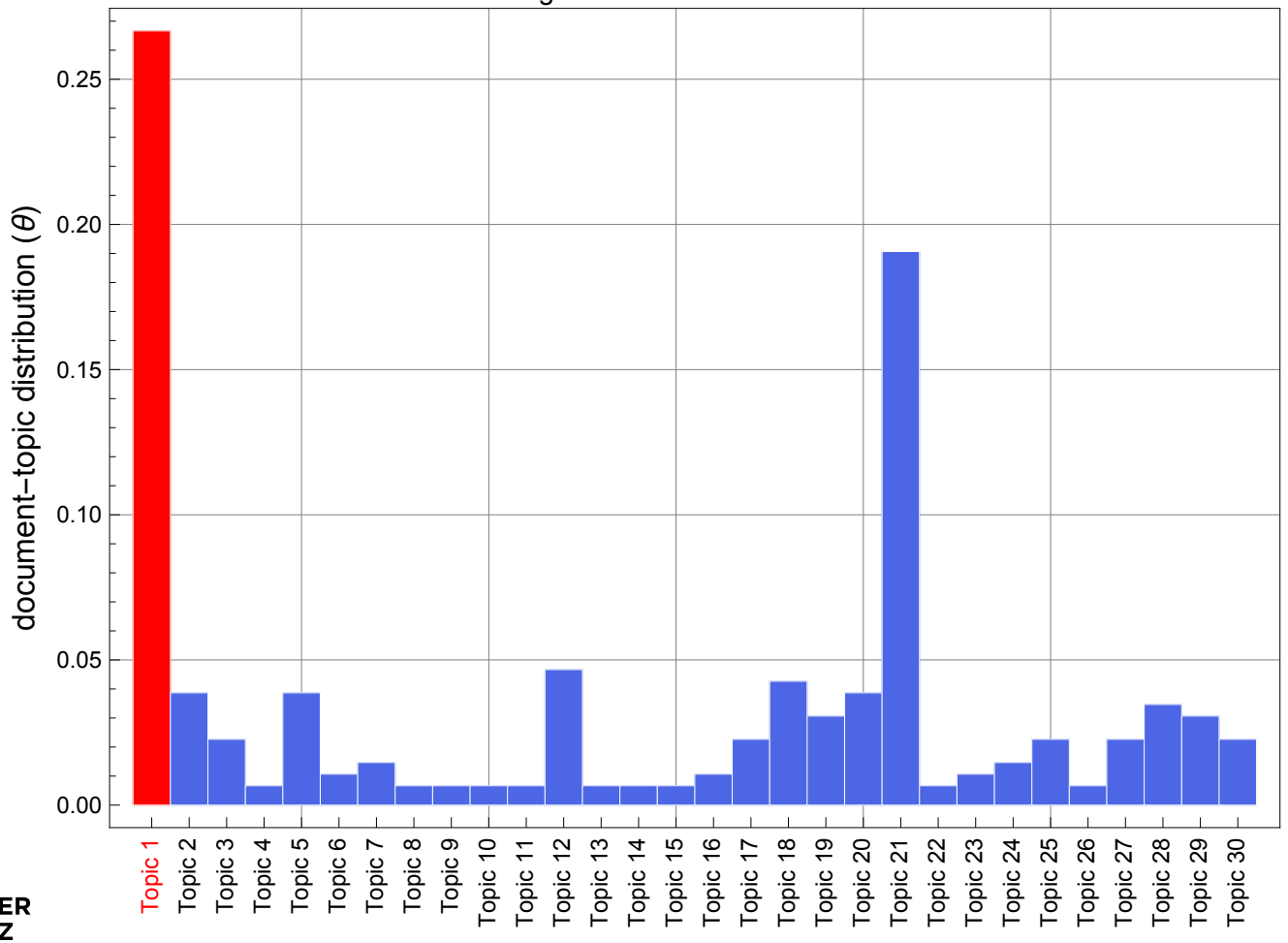


(preliminary) results II: some examples

- Topic1: technological innovation and industrial development

Topic 1	Weight (ϕ)
technology	0.0247
industry	0.0172
innovation	0.0164
firms	0.0158
product	0.0153
technological	0.0151
research	0.0147
production	0.013
process	0.0128
firm	0.0119
investment	0.0113
development	0.0108
industries	0.0104
products	0.0104
industrial	0.0103

David J., Teece (1989):
Inter–Organizational Requirements of the Innovation Process
Managerial and Decision Economics

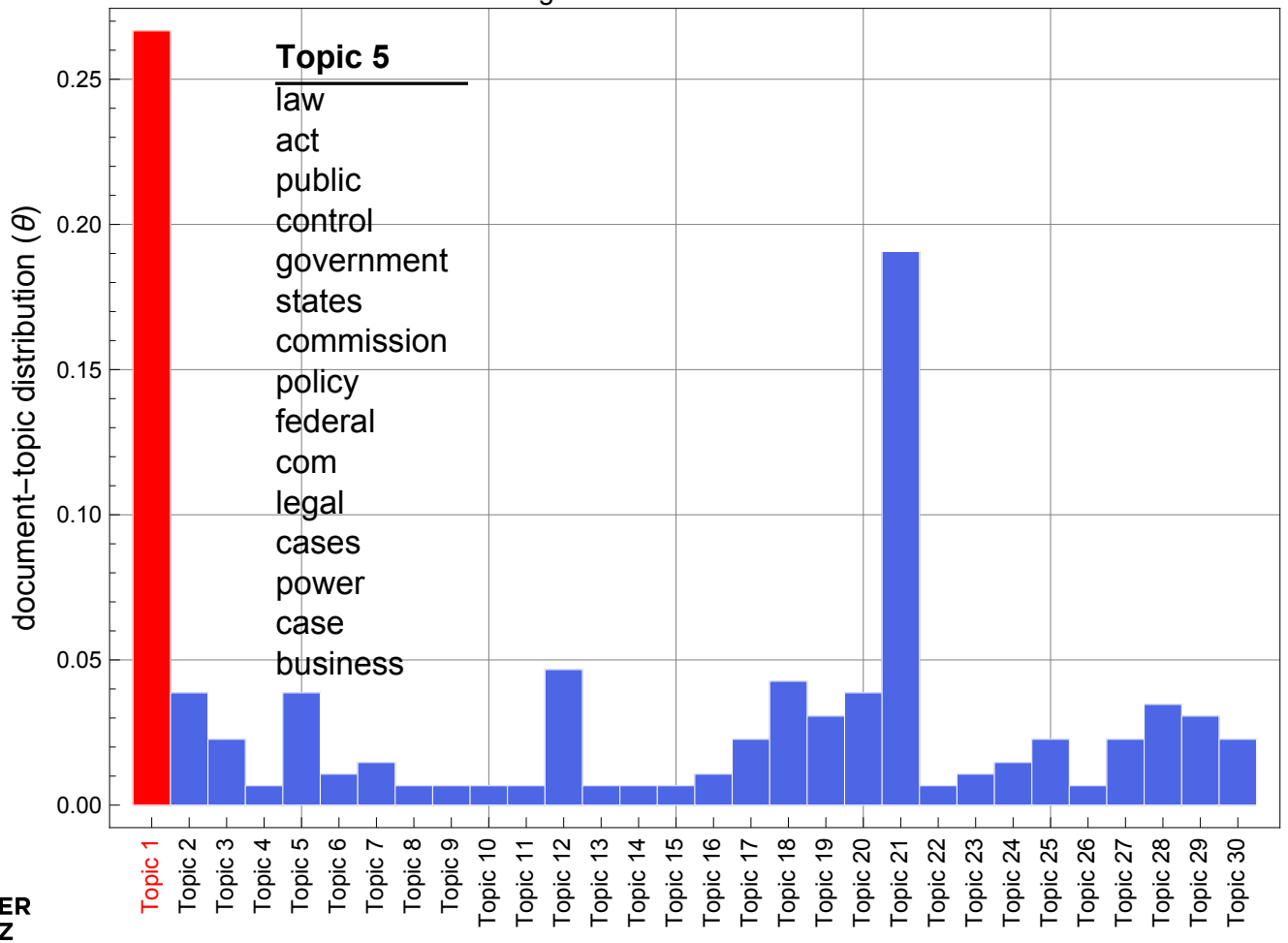


(preliminary) results II: some examples

- Topic1: technological innovation and industrial development

David J., Teece (1989):
Inter–Organizational Requirements of the Innovation Process
Managerial and Decision Economics

Topic 1	Weight (ϕ)
technology	0.0247
industry	0.0172
innovation	0.0164
firms	0.0158
product	0.0153
technological	0.0151
research	0.0147
production	0.013
process	0.0128
firm	0.0119
investment	0.0113
development	0.0108
industries	0.0104
products	0.0104
industrial	0.0103

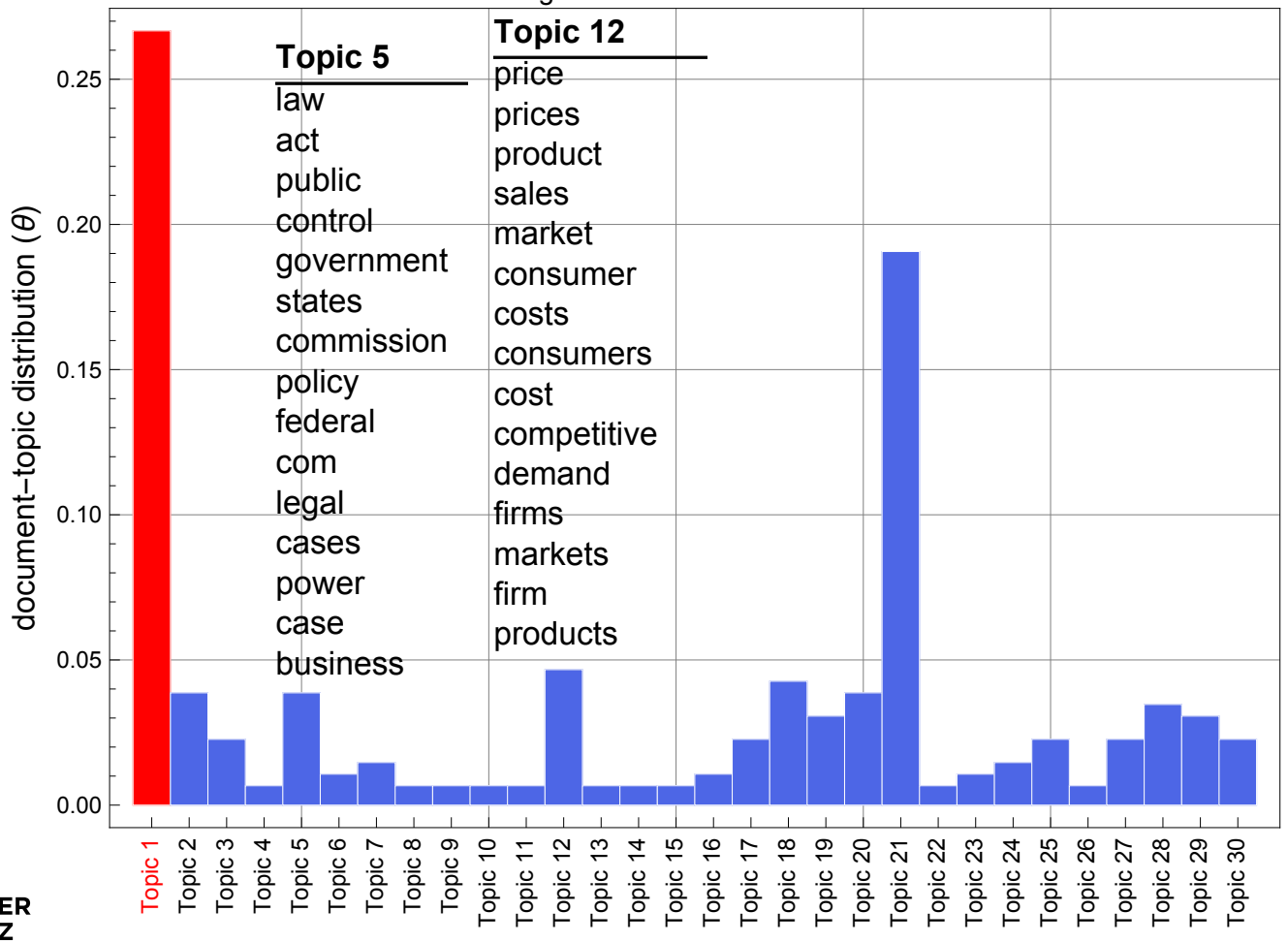


(preliminary) results II: some examples

- Topic1: technological innovation and industrial development

David J., Teece (1989):
Inter–Organizational Requirements of the Innovation Process
Managerial and Decision Economics

Topic 1	Weight (ϕ)
technology	0.0247
industry	0.0172
innovation	0.0164
firms	0.0158
product	0.0153
technological	0.0151
research	0.0147
production	0.013
process	0.0128
firm	0.0119
investment	0.0113
development	0.0108
industries	0.0104
products	0.0104
industrial	0.0103

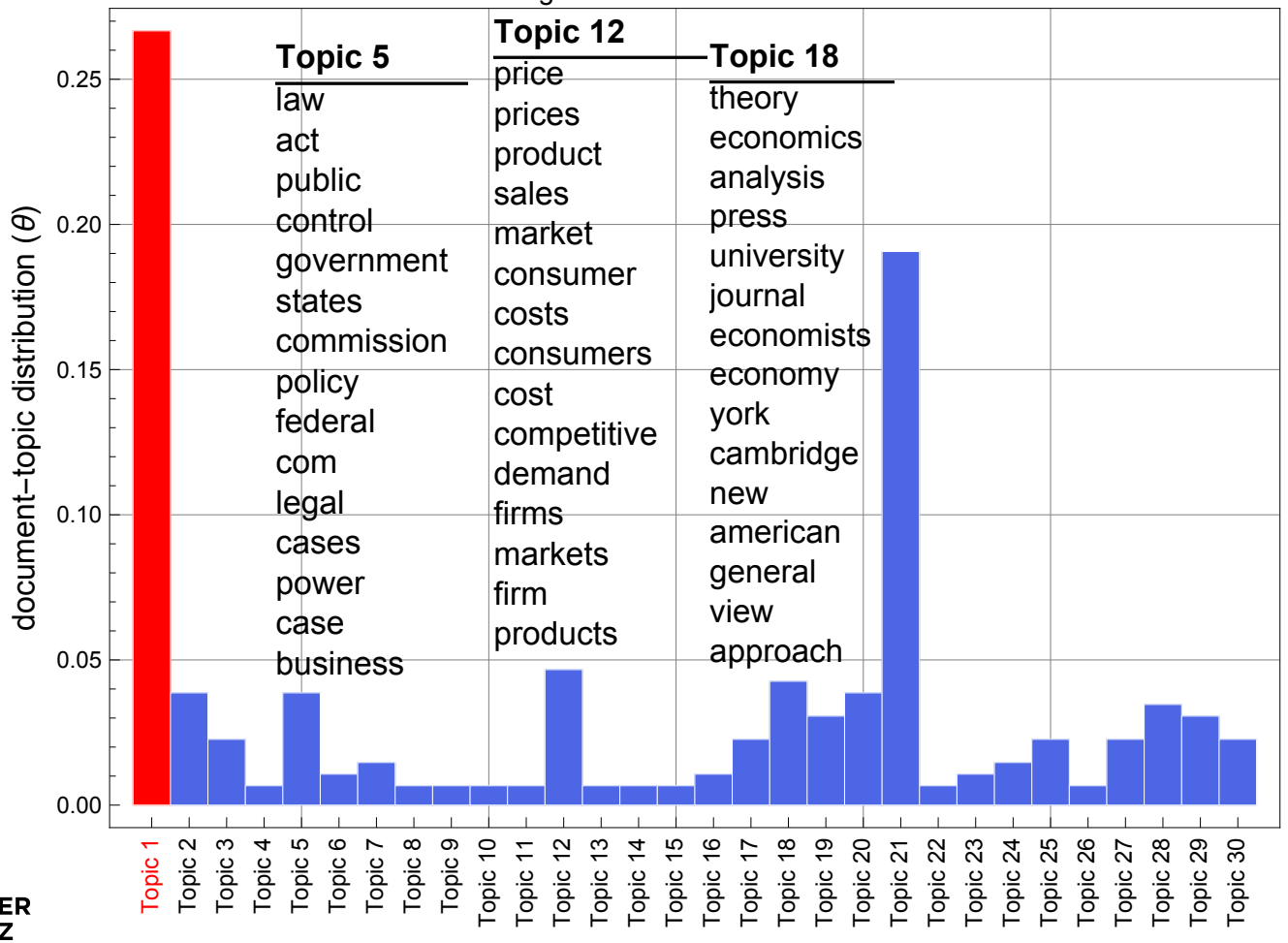


(preliminary) results II: some examples

- Topic1: technological innovation and industrial development

Topic 1	Weight (ϕ)
technology	0.0247
industry	0.0172
innovation	0.0164
firms	0.0158
product	0.0153
technological	0.0151
research	0.0147
production	0.013
process	0.0128
firm	0.0119
investment	0.0113
development	0.0108
industries	0.0104
products	0.0104
industrial	0.0103

David J., Teece (1989):
Inter–Organizational Requirements of the Innovation Process
Managerial and Decision Economics

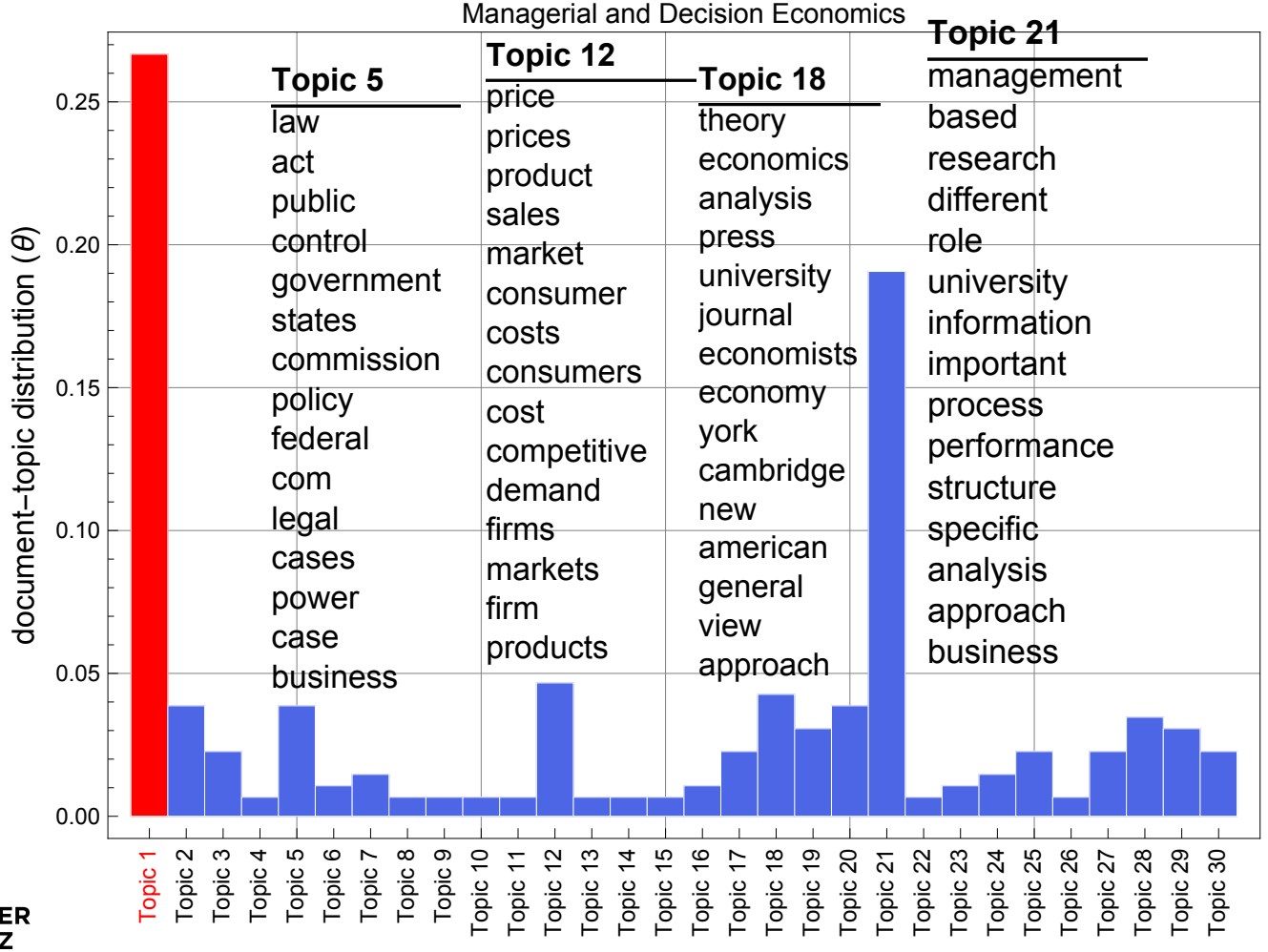


(preliminary) results II: some examples

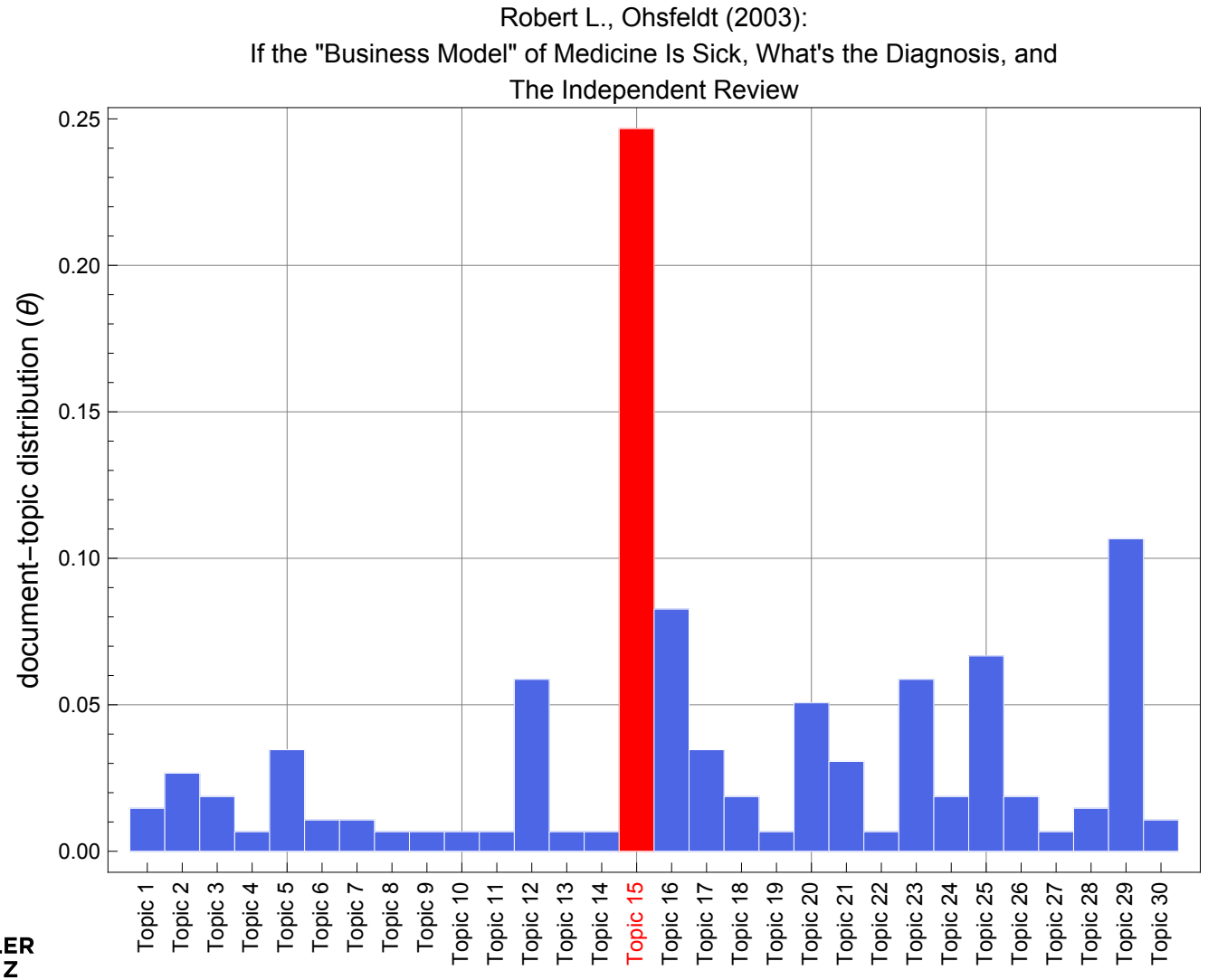
- Topic1: technological innovation and industrial development

Topic 1	Weight (ϕ)
technology	0.0247
industry	0.0172
innovation	0.0164
firms	0.0158
product	0.0153
technological	0.0151
research	0.0147
production	0.013
process	0.0128
firm	0.0119
investment	0.0113
development	0.0108
industries	0.0104
products	0.0104
industrial	0.0103

David J., Teece (1989):
Inter–Organizational Requirements of the Innovation Process
Managerial and Decision Economics

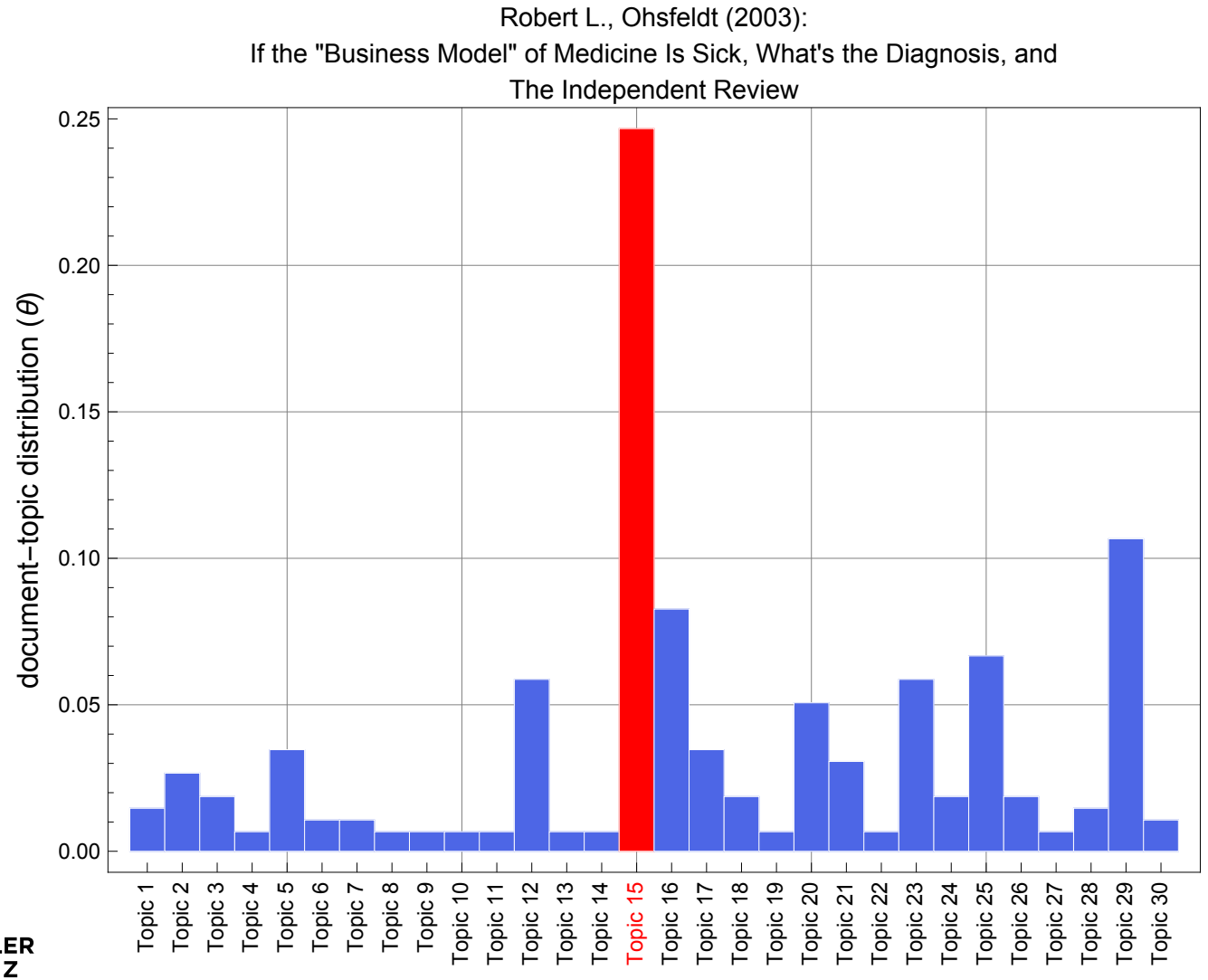


(preliminary) results II: some examples



(preliminary) results II: some examples

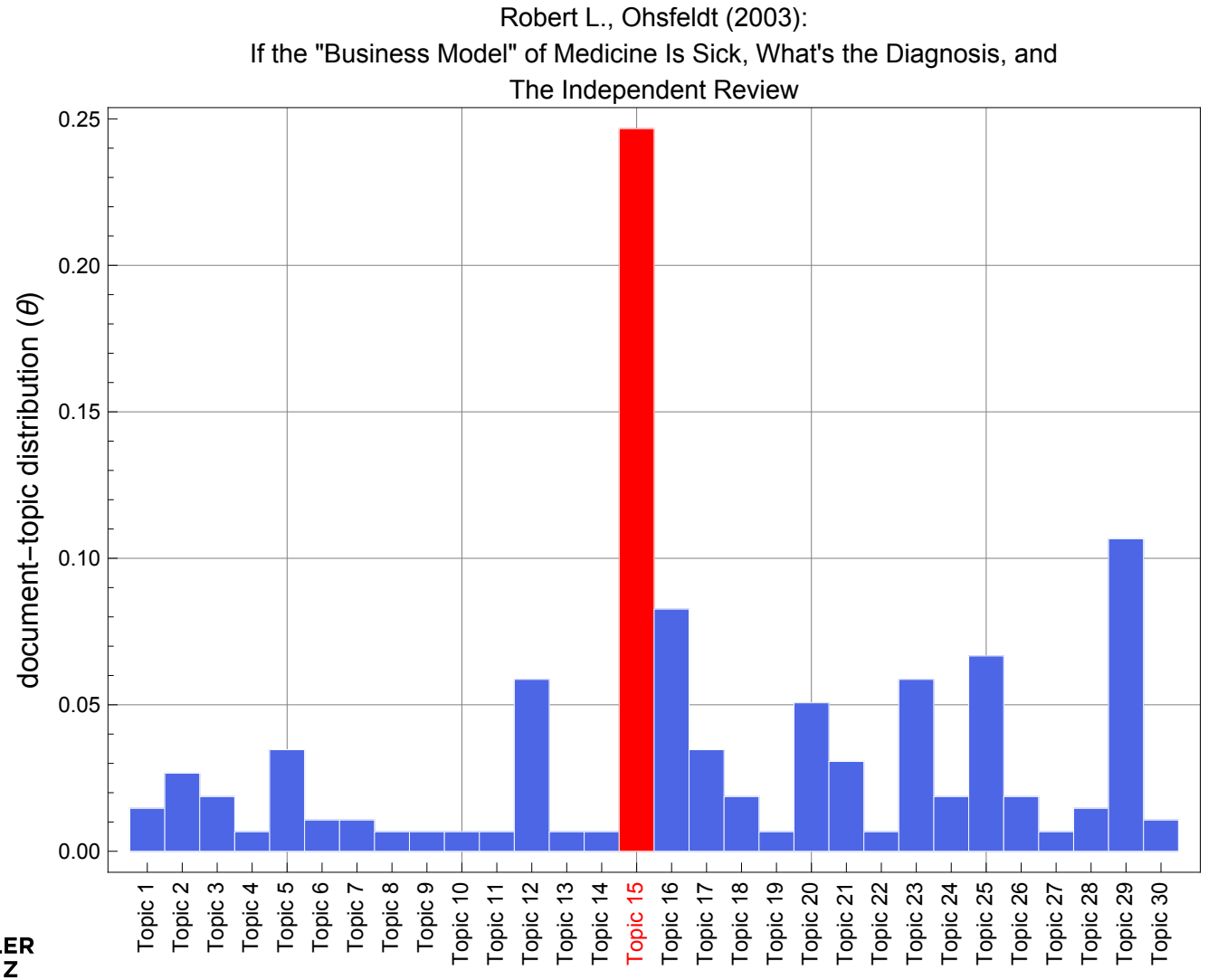
Topic 15	Weight (ϕ)
health	0.0199
insurance	0.0166
care	0.0164
services	0.0146
service	0.0109
cost	0.0108
quality	0.0101
medical	0.01
coverage	0.0096
premium	0.0093
risk	0.0085
airline	0.0077
insurers	0.0077
airlines	0.0074
plans	0.0068



(preliminary) results II: some examples

- Topic15: health care and insurance

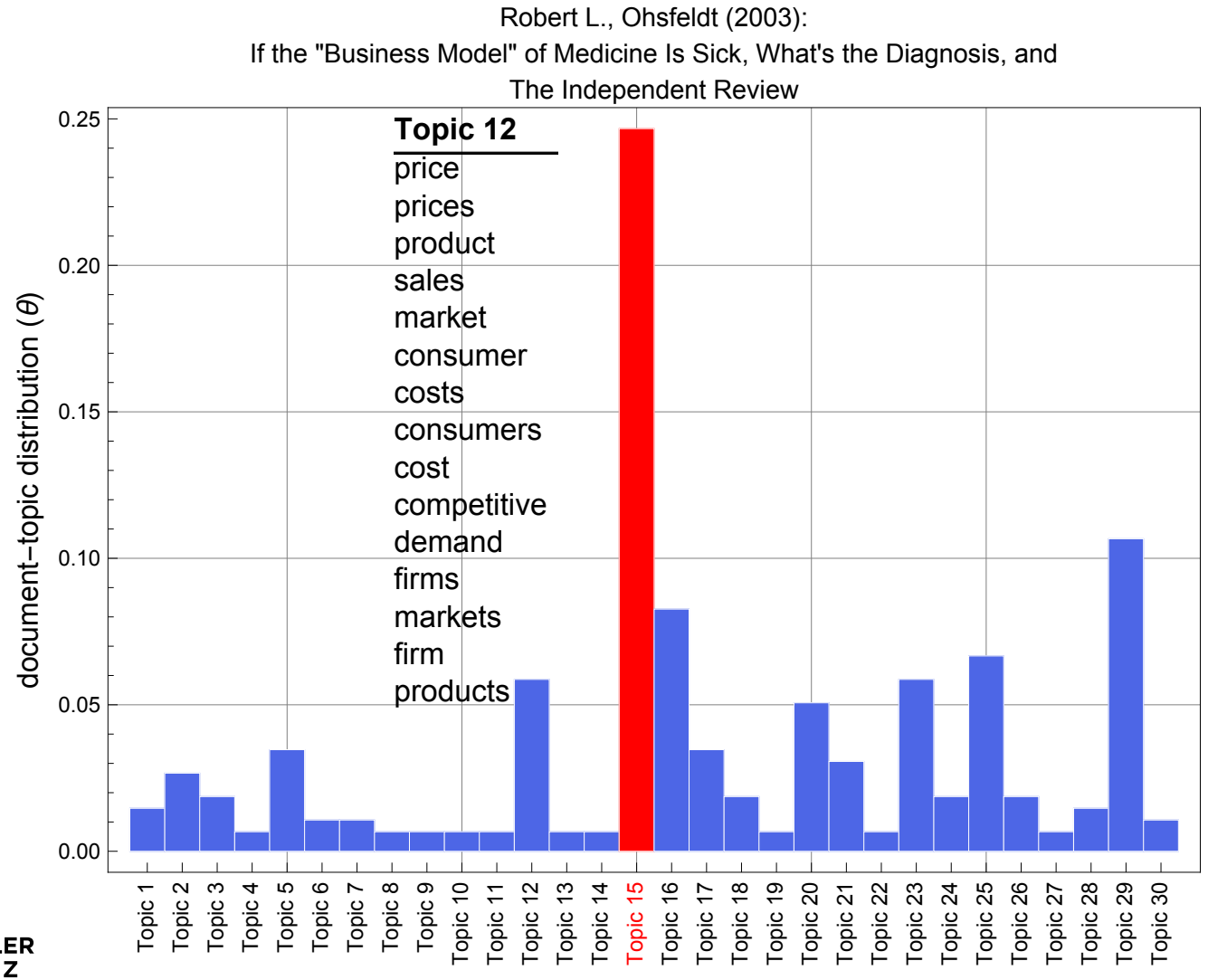
Topic 15	Weight (ϕ)
health	0.0199
insurance	0.0166
care	0.0164
services	0.0146
service	0.0109
cost	0.0108
quality	0.0101
medical	0.01
coverage	0.0096
premium	0.0093
risk	0.0085
airline	0.0077
insurers	0.0077
airlines	0.0074
plans	0.0068



(preliminary) results II: some examples

- Topic15: health care and insurance

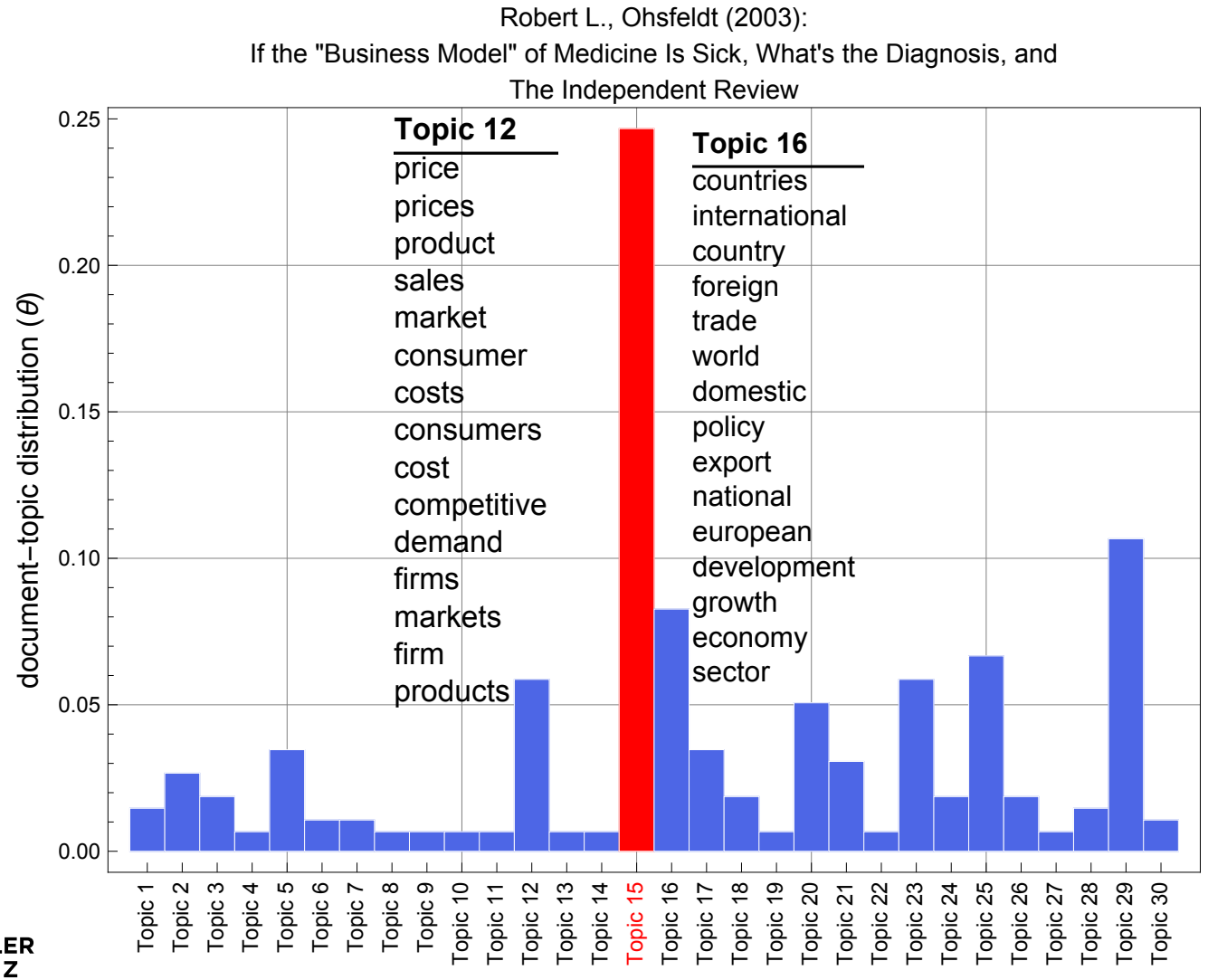
Topic 15	Weight (ϕ)
health	0.0199
insurance	0.0166
care	0.0164
services	0.0146
service	0.0109
cost	0.0108
quality	0.0101
medical	0.01
coverage	0.0096
premium	0.0093
risk	0.0085
airline	0.0077
insurers	0.0077
airlines	0.0074
plans	0.0068



(preliminary) results II: some examples

- Topic15: health care and insurance

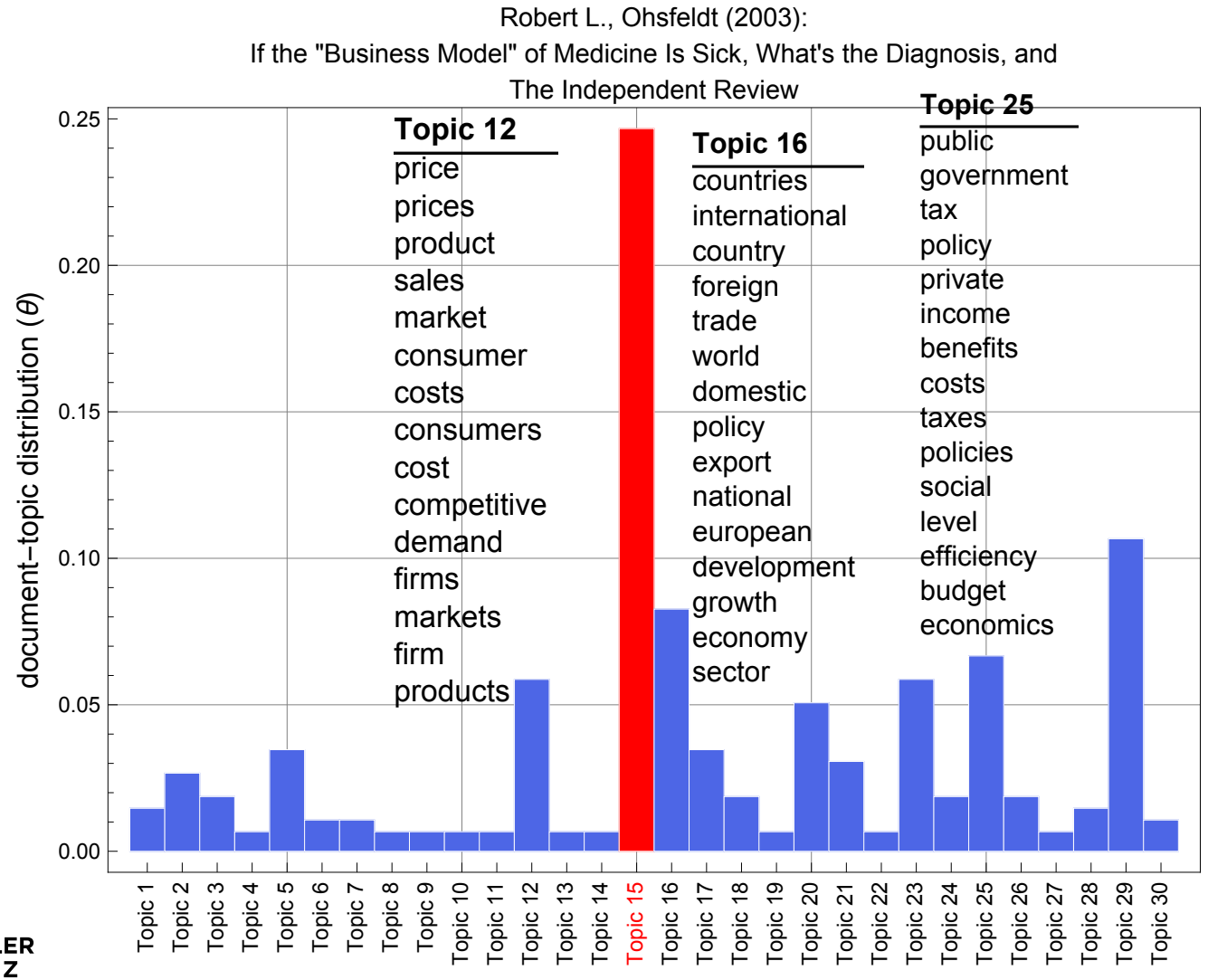
Topic 15	Weight (ϕ)
health	0.0199
insurance	0.0166
care	0.0164
services	0.0146
service	0.0109
cost	0.0108
quality	0.0101
medical	0.01
coverage	0.0096
premium	0.0093
risk	0.0085
airline	0.0077
insurers	0.0077
airlines	0.0074
plans	0.0068



(preliminary) results II: some examples

- Topic15: health care and insurance

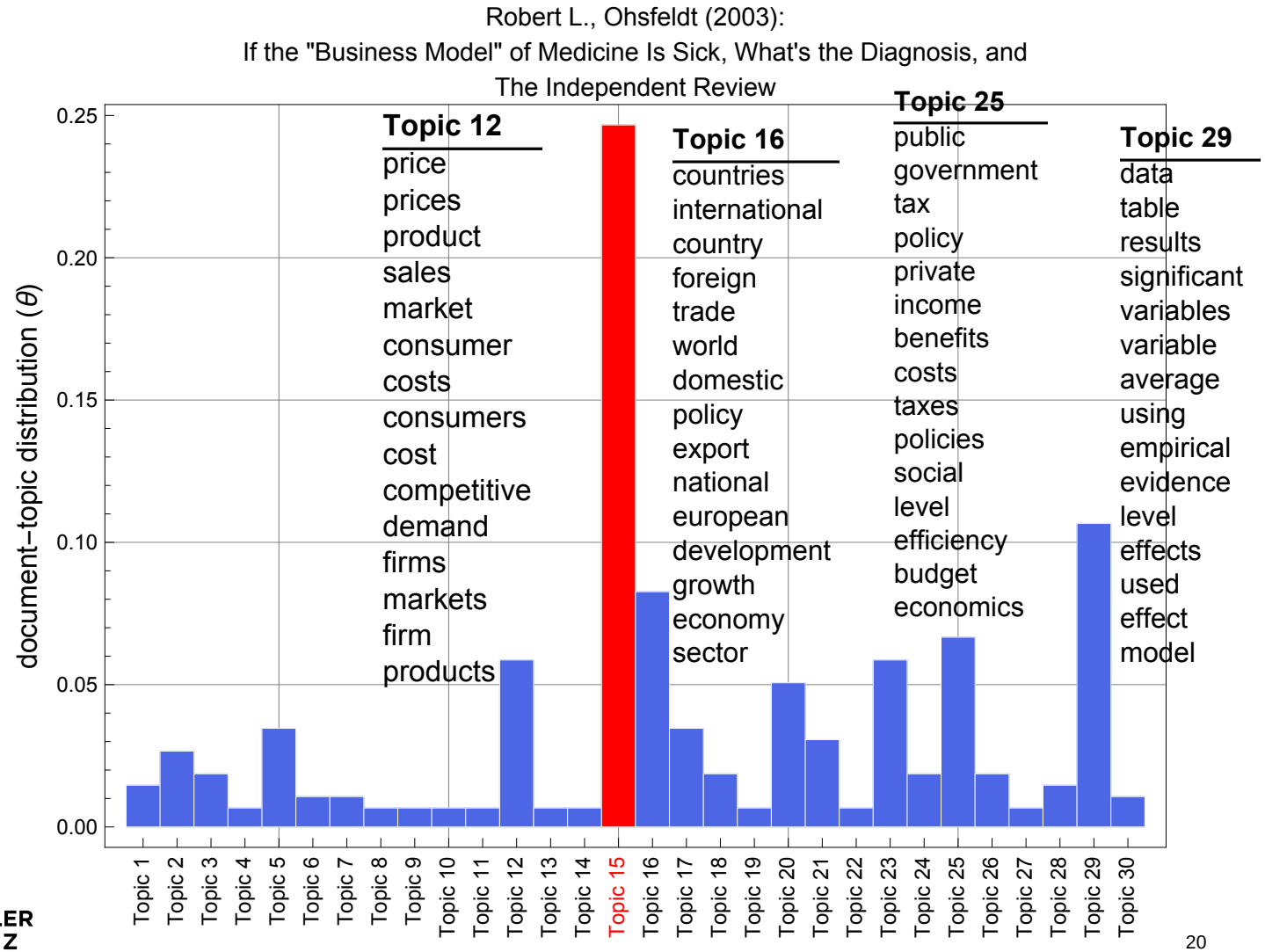
Topic 15	Weight (ϕ)
health	0.0199
insurance	0.0166
care	0.0164
services	0.0146
service	0.0109
cost	0.0108
quality	0.0101
medical	0.01
coverage	0.0096
premium	0.0093
risk	0.0085
airline	0.0077
insurers	0.0077
airlines	0.0074
plans	0.0068



(preliminary) results II: some examples

- Topic15: health care and insurance

Topic 15	Weight (ϕ)
health	0.0199
insurance	0.0166
care	0.0164
services	0.0146
service	0.0109
cost	0.0108
quality	0.0101
medical	0.01
coverage	0.0096
premium	0.0093
risk	0.0085
airline	0.0077
insurers	0.0077
airlines	0.0074
plans	0.0068

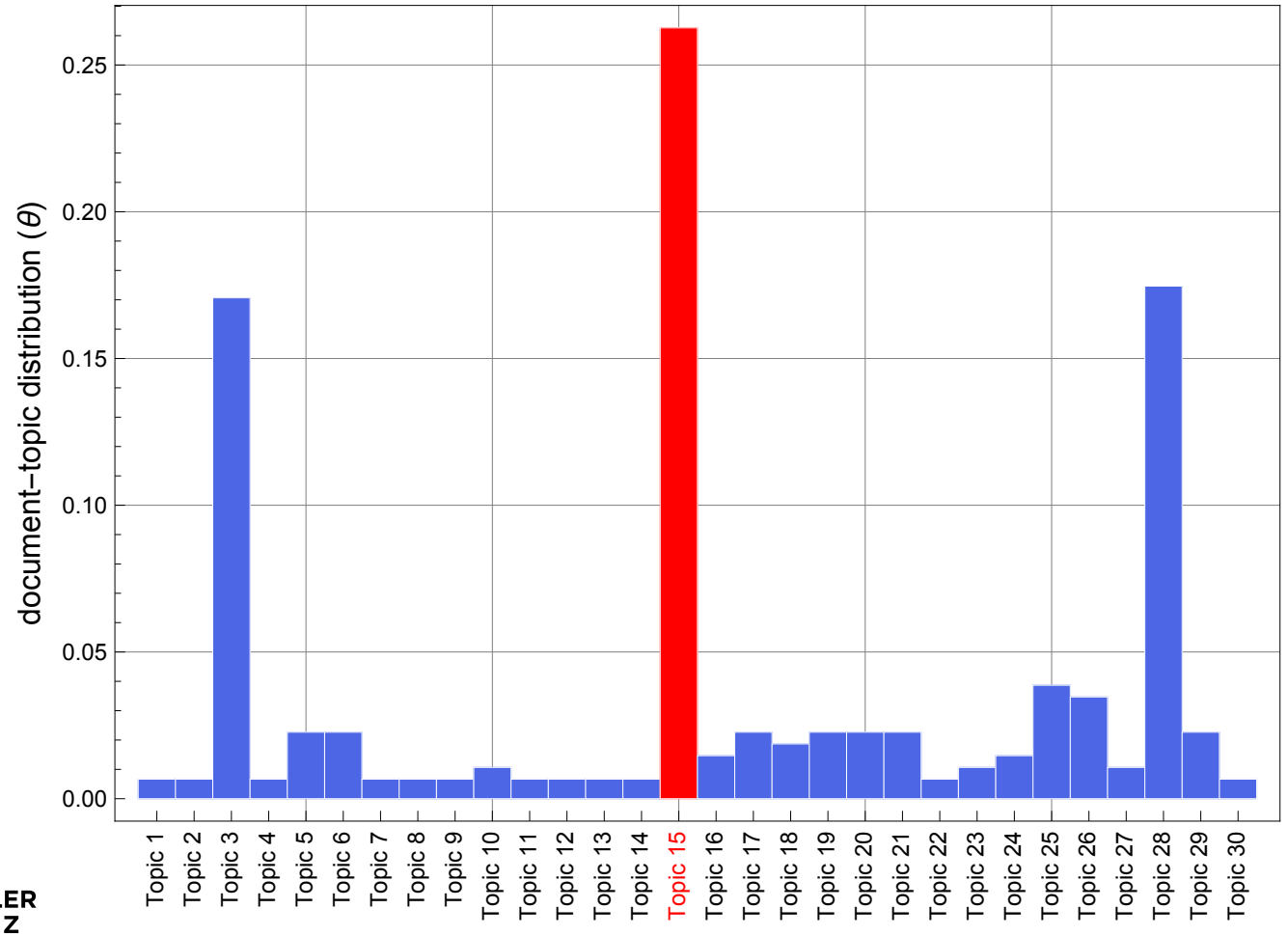


(preliminary) results II: some examples

- Topic15: health care and insurance

Topic 15	Weight (ϕ)
health	0.0199
insurance	0.0166
care	0.0164
services	0.0146
service	0.0109
cost	0.0108
quality	0.0101
medical	0.01
coverage	0.0096
premium	0.0093
risk	0.0085
airline	0.0077
insurers	0.0077
airlines	0.0074
plans	0.0068

Karen, Eggleston (2000):
Risk Selection and Optimal Health Insurance–Provider Payment Systems
The Journal of Risk and Insurance

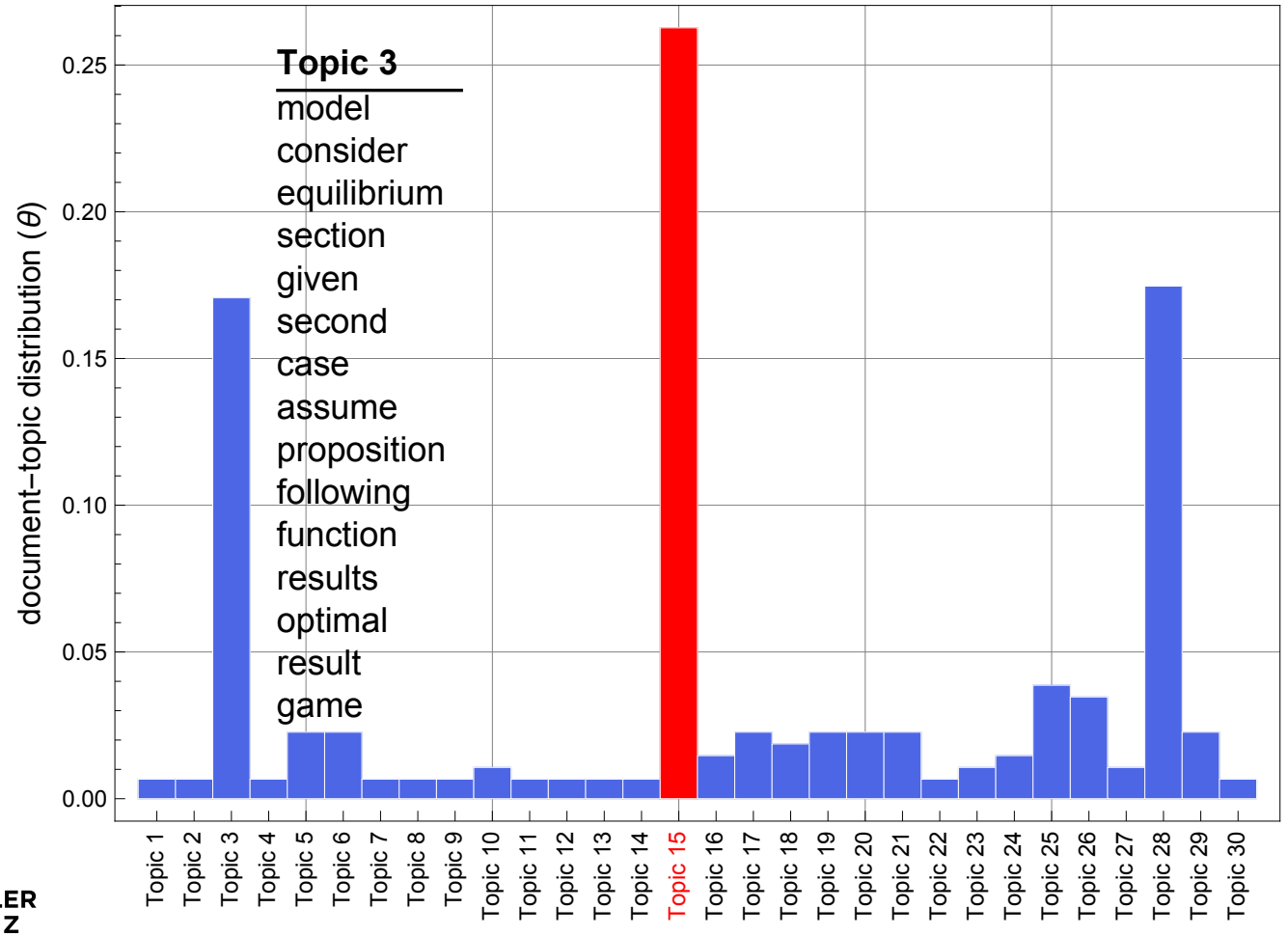


(preliminary) results II: some examples

- Topic15: health care and insurance

Topic 15	Weight (ϕ)
health	0.0199
insurance	0.0166
care	0.0164
services	0.0146
service	0.0109
cost	0.0108
quality	0.0101
medical	0.01
coverage	0.0096
premium	0.0093
risk	0.0085
airline	0.0077
insurers	0.0077
airlines	0.0074
plans	0.0068

Karen, Eggleston (2000):
Risk Selection and Optimal Health Insurance–Provider Payment Systems
The Journal of Risk and Insurance

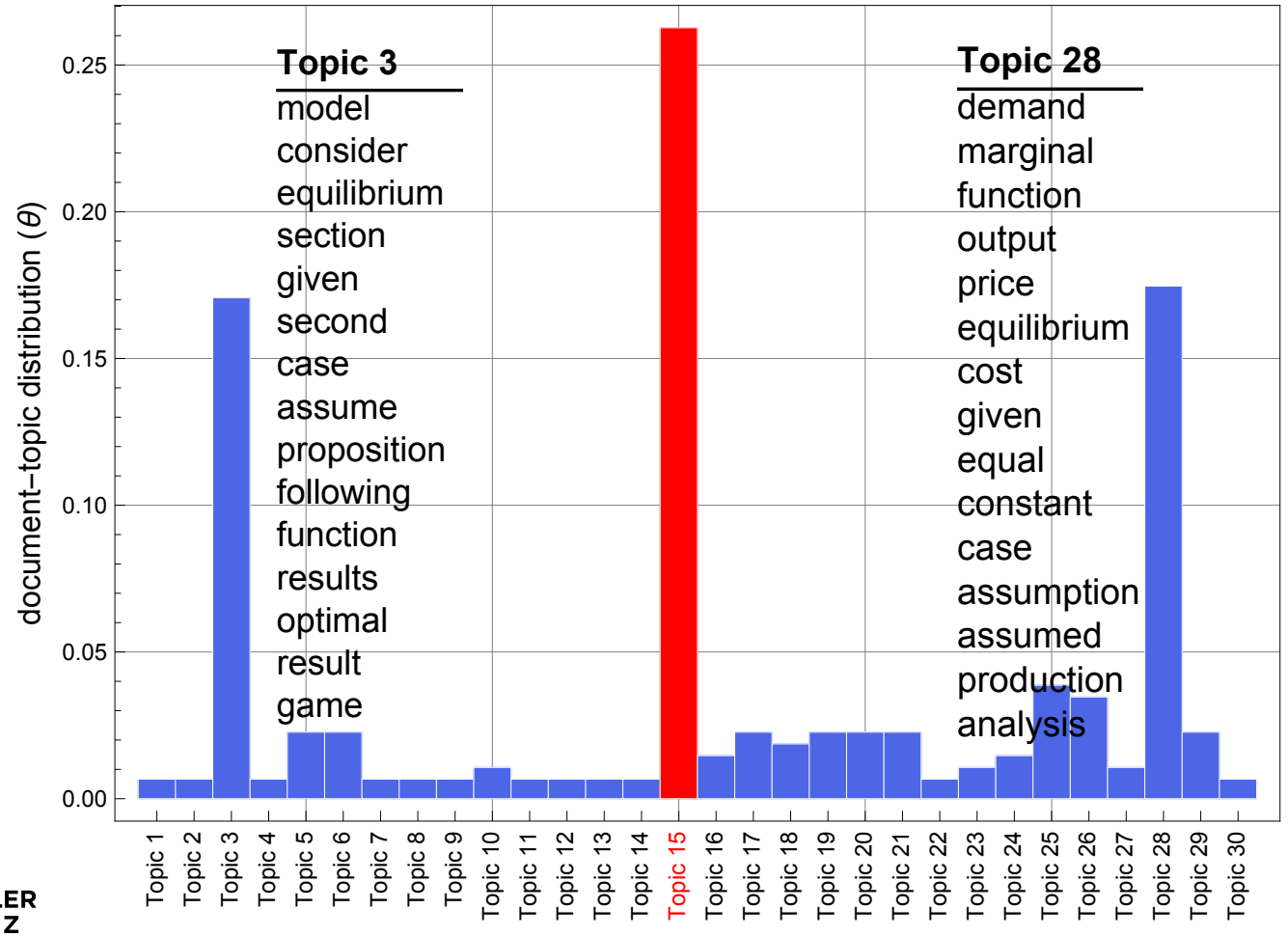


(preliminary) results II: some examples

- Topic15: health care and insurance

Topic 15	Weight (ϕ)
health	0.0199
insurance	0.0166
care	0.0164
services	0.0146
service	0.0109
cost	0.0108
quality	0.0101
medical	0.01
coverage	0.0096
premium	0.0093
risk	0.0085
airline	0.0077
insurers	0.0077
airlines	0.0074
plans	0.0068

Karen, Eggleston (2000):
Risk Selection and Optimal Health Insurance–Provider Payment Systems
The Journal of Risk and Insurance

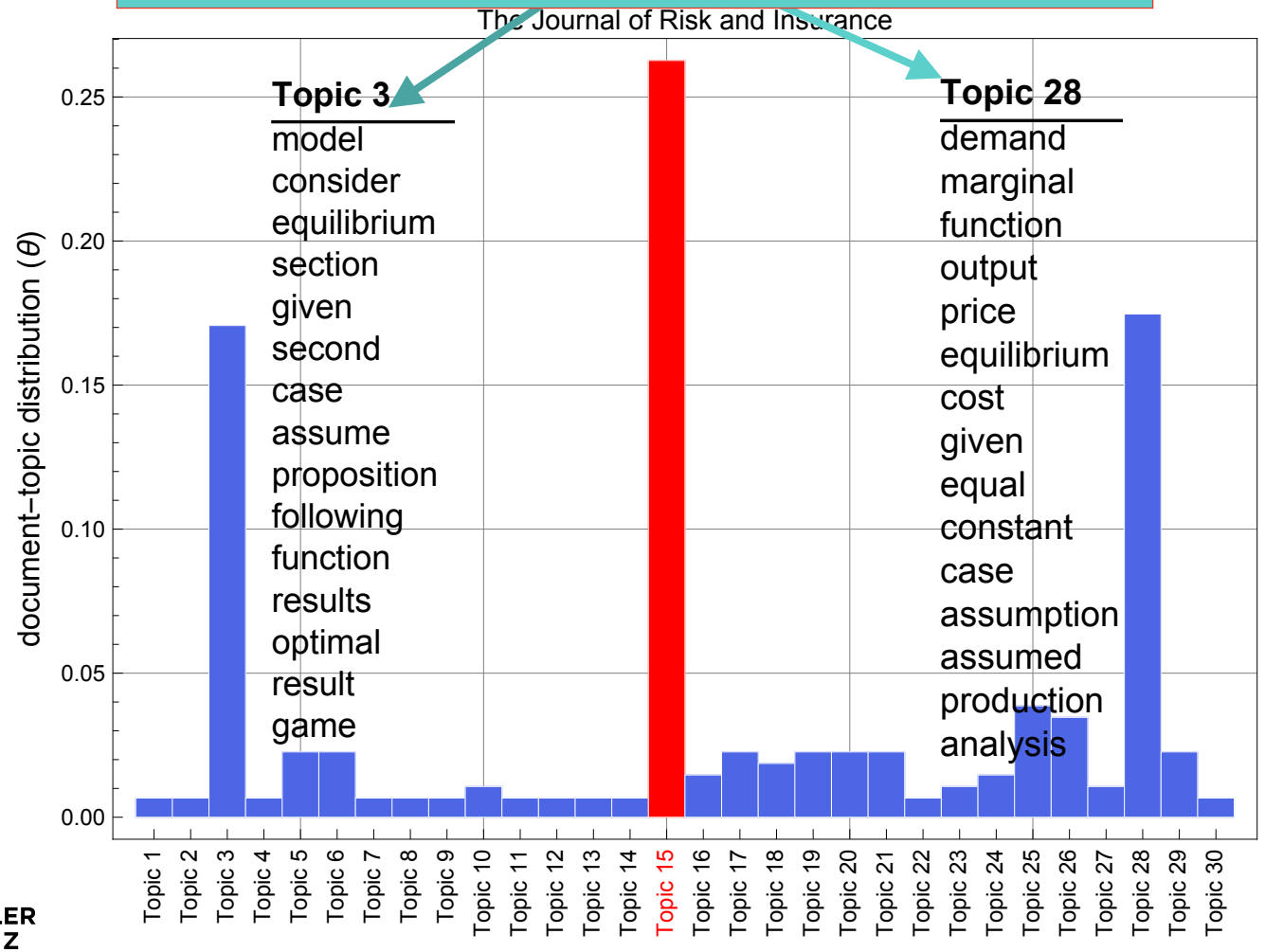


(preliminary) results II: some examples

- Topic15: health car

Question to the audience: Whats the difference between these two topics? Some suggestions?

Topic 15	Weight (ϕ)
health	0.0199
insurance	0.0166
care	0.0164
services	0.0146
service	0.0109
cost	0.0108
quality	0.0101
medical	0.01
coverage	0.0096
premium	0.0093
risk	0.0085
airline	0.0077
insurers	0.0077
airlines	0.0074
plans	0.0068



(preliminary) results II: some examples

(preliminary) results II: some examples

Topic 22	Weight (ϕ)
anthony	0.0087
morris	0.0066
occurring	0.0064
klaus	0.0053
capturing	0.004
observers	0.0036
tighter	0.0036
chowdhury	0.0034
tai	0.003
phillip	0.0025
stayed	0.0025
quiet	0.0023
vic	0.0023
fication	0.0023
stacking	0.0021

(preliminary) results II: some examples

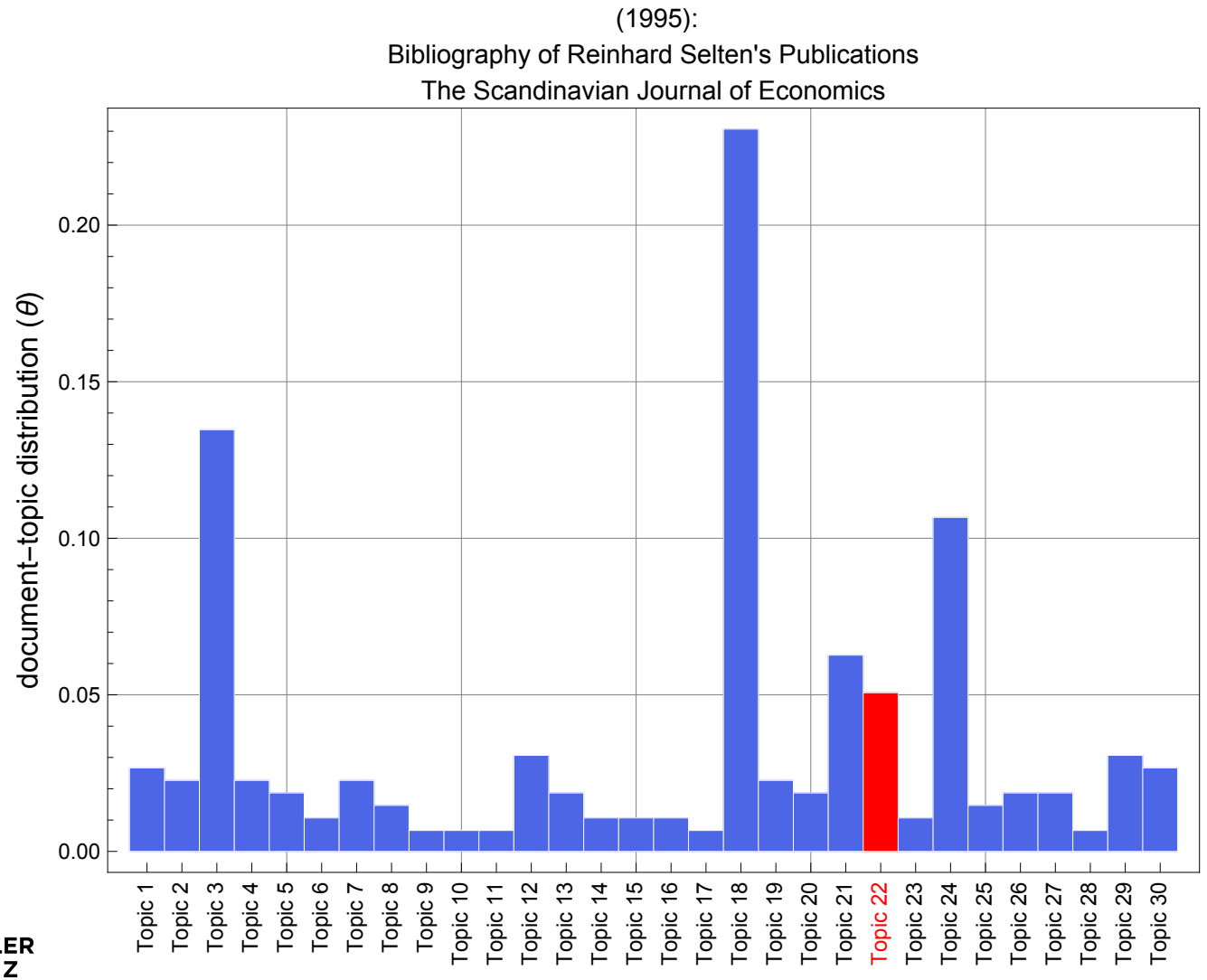
- Topic22:???

Topic 22	Weight (ϕ)
anthony	0.0087
morris	0.0066
occurring	0.0064
klaus	0.0053
capturing	0.004
observers	0.0036
tighter	0.0036
chowdhury	0.0034
tai	0.003
phillip	0.0025
stayed	0.0025
quiet	0.0023
vic	0.0023
fication	0.0023
stacking	0.0021

(preliminary) results II: some examples

- Topic22:???

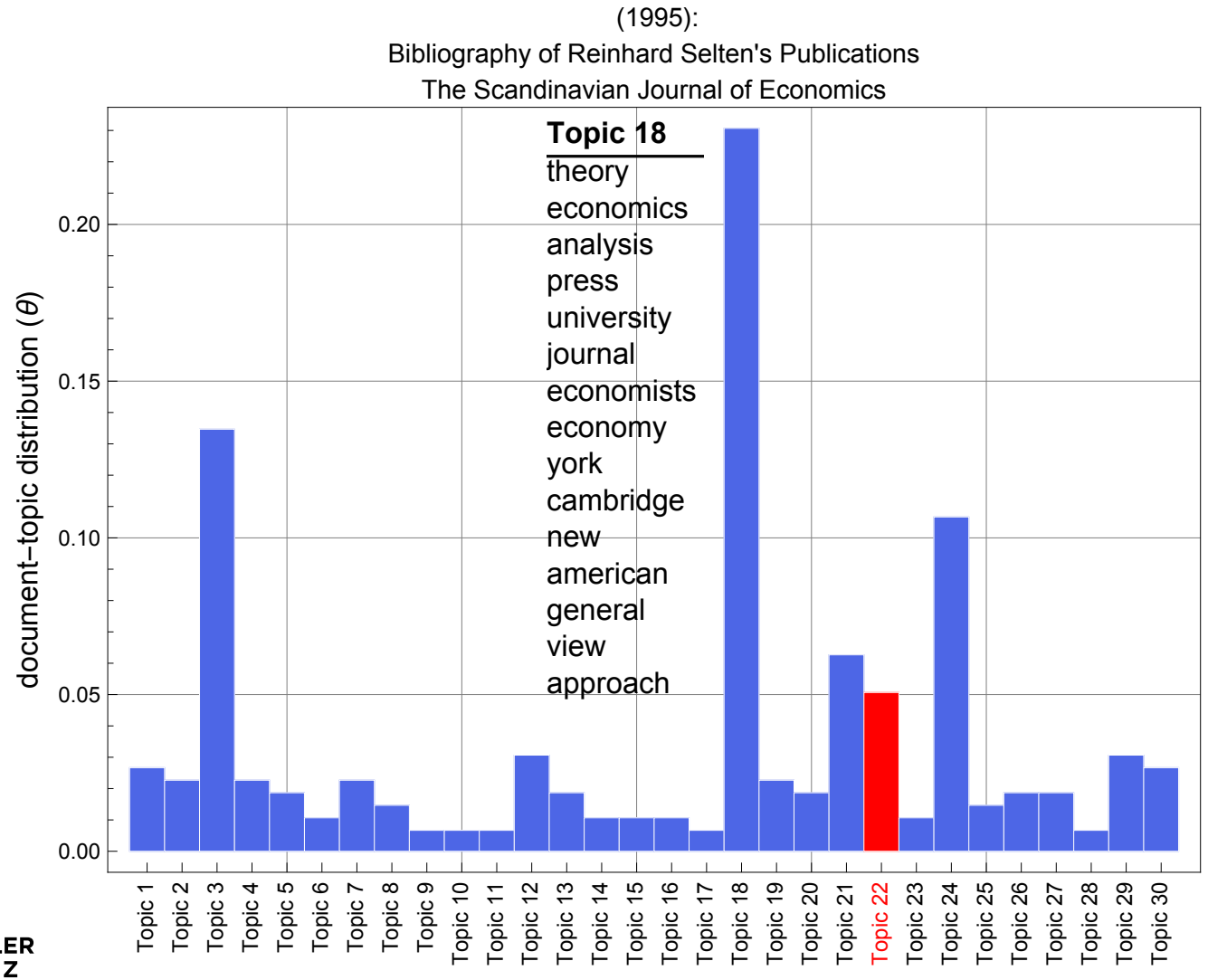
Topic 22	Weight (ϕ)
anthony	0.0087
morris	0.0066
occurring	0.0064
klaus	0.0053
capturing	0.004
observers	0.0036
tighter	0.0036
chowdhury	0.0034
tai	0.003
phillip	0.0025
stayed	0.0025
quiet	0.0023
vic	0.0023
fication	0.0023
stacking	0.0021



(preliminary) results II: some examples

- Topic22:???

Topic 22	Weight (ϕ)
anthony	0.0087
morris	0.0066
occurring	0.0064
klaus	0.0053
capturing	0.004
observers	0.0036
tighter	0.0036
chowdhury	0.0034
tai	0.003
phillip	0.0025
stayed	0.0025
quiet	0.0023
vic	0.0023
fication	0.0023
stacking	0.0021



Summary & Outlook

Summary & Outlook

- (Preliminary) results of the descriptive statistics show that the discourse on competition

Summary & Outlook

- (Preliminary) results of the descriptive statistics show that the discourse on competition
  - Starts with a time lag of about 10-15 years to intensify compared to the overall development of economic research output

Summary & Outlook

- (Preliminary) results of the descriptive statistics show that the discourse on competition
  - Starts with a time lag of about 10-15 years to intensify compared to the overall development of economic research output
  - Is highly concentrated in terms of publication outlets: a third of all articles in the sample are published in 10 journals

Summary & Outlook

- (Preliminary) results of the descriptive statistics show that the discourse on competition
  - Starts with a time lag of about 10-15 years to intensify compared to the overall development of economic research output
  - Is highly concentrated in terms of publication outlets: a third of all articles in the sample are published in 10 journals
  - Is – given the marginalization of heterodox approaches within economics – somewhat balanced: a substantial part of the literature related to competition is published in (a few) heterodox journals

Summary & Outlook

Summary & Outlook

- Results of the first topic modeling exercise are promising

Summary & Outlook

- Results of the first topic modeling exercise are promising
  - A bird’s eye view on the overall corpus of research articles delivers a topic structure which links competition to various meaningful topics that emerge within the literature.

Summary & Outlook

- Results of the first topic modeling exercise are promising
  - A bird's eye view on the overall corpus of research articles delivers a topic structure which links competition to various meaningful topics that emerge within the literature.
  - Analysis of the topic-document distributions enables

Summary & Outlook

- Results of the first topic modeling exercise are promising
  - A bird's eye view on the overall corpus of research articles delivers a topic structure which links competition to various meaningful topics that emerge within the literature.
  - Analysis of the topic-document distributions enables
    - a reconstruction of the genesis of specific topics (and its disappearance)

Summary & Outlook

- Results of the first topic modeling exercise are promising
  - A bird's eye view on the overall corpus of research articles delivers a topic structure which links competition to various meaningful topics that emerge within the literature.
  - Analysis of the topic-document distributions enables
    - a reconstruction of the genesis of specific topics (and its disappearance)
    - an analysis of co-occurrences between specific topics that relate to competition

Summary & Outlook

Summary & Outlook

- However, model parameters still need to be optimized

Summary & Outlook

- However, model parameters still need to be optimized
  - Number of Topics (more vs. less)

Summary & Outlook

- However, model parameters still need to be optimized
  - Number of Topics (more vs. less)
  - Different Dirichlet priors (see also Tang et al. 2014, Wallach et al. 2009)

Summary & Outlook

- However, model parameters still need to be optimized
  - Number of Topics (more vs. less)
  - Different Dirichlet priors (see also Tang et al. 2014, Wallach et al. 2009)
    - α = small: each document is associated with few topics

Summary & Outlook

- However, model parameters still need to be optimized
  - Number of Topics (more vs. less)
  - Different Dirichlet priors (see also Tang et al. 2014, Wallach et al. 2009)
    - α = small: each document is associated with few topics
    - β = small: topics are word sparse (large β means more word-diffused and similar topics)

Summary & Outlook

- However, model parameters still need to be optimized
  - Number of Topics (more vs. less)
  - Different Dirichlet priors (see also Tang et al. 2014, Wallach et al. 2009)
    - α = small: each document is associated with few topics
    - β = small: topics are word sparse (large β means more word-diffused and similar topics)
    - Symmetrical vs. asymmetrical priors

Summary & Outlook

- However, model parameters still need to be optimized
  - Number of Topics (more vs. less)
  - Different Dirichlet priors (see also Tang et al. 2014, Wallach et al. 2009)
    - α = small: each document is associated with few topics
    - β = small: topics are word sparse (large β means more word-diffused and similar topics)
    - Symmetrical vs. asymmetrical priors
  - Text preprocessing (see e.g. Schofield et al. 2017)

Summary & Outlook

- However, model parameters still need to be optimized
  - Number of Topics (more vs. less)
  - Different Dirichlet priors (see also Tang et al. 2014, Wallach et al. 2009)
    - α = small: each document is associated with few topics
    - β = small: topics are word sparse (large β means more word-diffused and similar topics)
    - Symmetrical vs. asymmetrical priors
  - Text preprocessing (see e.g. Schofield et al. 2017)
    - “Pseudo-abstracts” vs. full-text analysis?

Summary & Outlook

- However, model parameters still need to be optimized
  - Number of Topics (more vs. less)
  - Different Dirichlet priors (see also Tang et al. 2014, Wallach et al. 2009)
    - α = small: each document is associated with few topics
    - β = small: topics are word sparse (large β means more word-diffused and similar topics)
    - Symmetrical vs. asymmetrical priors
  - Text preprocessing (see e.g. Schofield et al. 2017)
    - “Pseudo-abstracts” vs. full-text analysis?
    - Remove standard stop-words vs. topic-specific stop-words?

Summary & Outlook

- However, model parameters still need to be optimized
  - Number of Topics (more vs. less)
  - Different Dirichlet priors (see also Tang et al. 2014, Wallach et al. 2009)
    - α = small: each document is associated with few topics
    - β = small: topics are word sparse (large β means more word-diffused and similar topics)
    - Symmetrical vs. asymmetrical priors
  - Text preprocessing (see e.g. Schofield et al. 2017)
    - “Pseudo-abstracts” vs. full-text analysis?
    - Remove standard stop-words vs. topic-specific stop-words?
    - Word stemming or not?

Summary & Outlook

Summary & Outlook

- Possible next steps are

Summary & Outlook

- Possible next steps are
  - Analyzing the corresponding bibliometric meta-data

Summary & Outlook

- Possible next steps are
  - Analyzing the corresponding bibliometric meta-data
    - via publication year: when do these topics emerge/disappear?

Summary & Outlook

- Possible next steps are
  - Analyzing the corresponding bibliometric meta-data
    - via publication year: when do these topics emerge/disappear?
    - via publication outlet: where do these topics emerge/disappear?

Summary & Outlook

- Possible next steps are
  - Analyzing the corresponding bibliometric meta-data
    - via publication year: when do these topics emerge/disappear?
    - via publication outlet: where do these topics emerge/disappear?
    - via cited references: what are the conceptual foundations of a topic?

Summary & Outlook

- Possible next steps are
  - Analyzing the corresponding bibliometric meta-data
    - via publication year: when do these topics emerge/disappear?
    - via publication outlet: where do these topics emerge/disappear?
    - via cited references: what are the conceptual foundations of a topic?
  - A more nuanced analysis / Extensions

Summary & Outlook

- Possible next steps are
  - Analyzing the corresponding bibliometric meta-data
    - via publication year: when do these topics emerge/disappear?
    - via publication outlet: where do these topics emerge/disappear?
    - via cited references: what are the conceptual foundations of a topic?
  - A more nuanced analysis / Extensions
    - Extracting topics from period/decade subsamples (e.g. 1960s, 1970s,… see also Ambrosino et al. 2018)

Summary & Outlook

- Possible next steps are
  - Analyzing the corresponding bibliometric meta-data
    - via publication year: when do these topics emerge/disappear?
    - via publication outlet: where do these topics emerge/disappear?
    - via cited references: what are the conceptual foundations of a topic?
  - A more nuanced analysis / Extensions
    - Extracting topics from period/decade subsamples (e.g. 1960s, 1970s,… see also Ambrosino et al. 2018)
    - Replication for other fields (e.g. Sociology, Political Science, Anthropology)

Many thanks for your attention