

# Introduction to networks in Computational Social Science

Markus Strohmaier

University of Mannheim, Germany  
GESIS – Leibniz Institute for the Social Sciences, Germany  
Complexity Science Hub Vienna

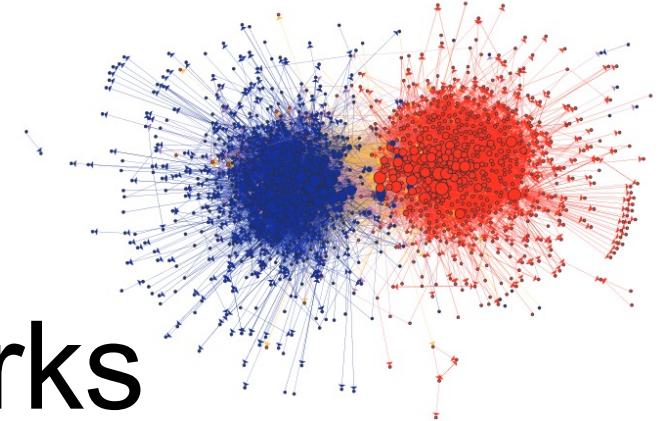
# The Social Sciences\*

\*a simplified view

... are interested in **understanding how people**

- **think/feel/behave** in social situations (social psychology),
- **relate** to each other (sociology),
- **govern** themselves (political science),
- **handle wealth** (socioeconomics), and
- **create culture** (anthropology).

Online networks as  
a universal  
abstraction level



# Online Networks

... any kind of online data that can be represented in a relational way

- users and social interactions between them
- users and activities
- users and their navigation behavior
- user behavior over time and place
- ....

# Online networks: a membrane for social phenomena

- Inequality
- Family
- Work
- Religion
- Migration
- Health
- Culture
- Elections
- Politics
- Polarization
- Environment
- Personality
- Demographics
- Crime

Many „real world“ social phenomena  
spill over to online networks.

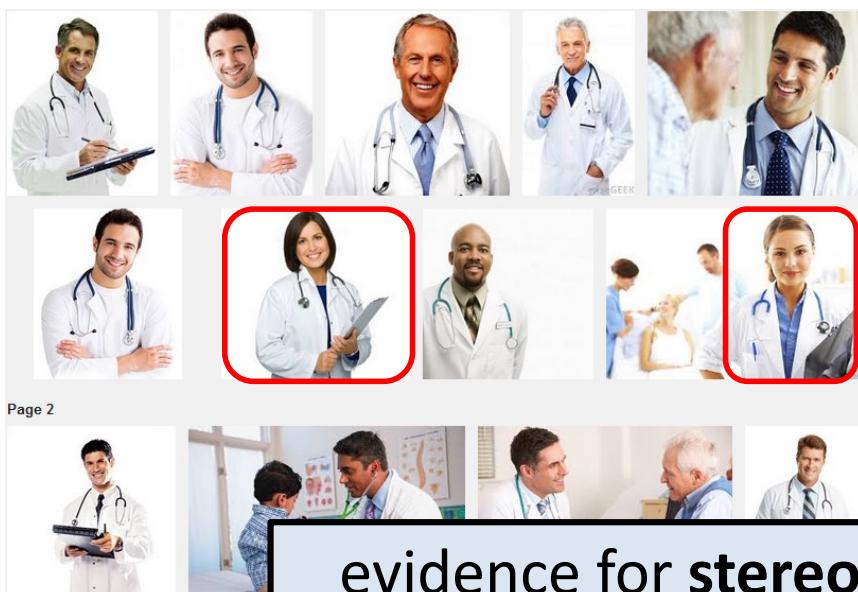
# Social Behavior online

Hypotheses:

- Online behavior is different from offline behavior (Bradley effect, Hawthorne effect, social desirability)
- The online world is equally part of our social reality as the offline world (Online tribes)
- Relational data is abundant, demographic data is sparse (IID assumptions do not hold)
- Often  $n = \text{all}$  or  $n > \text{all}$  (bots, spam, multiple IDs)
- Algorithms and online data shape opinions, social issues and social behavior (nudging)

# Gender inequality on the web

Google image query: „Doctor“



Google image query: „Nurse“

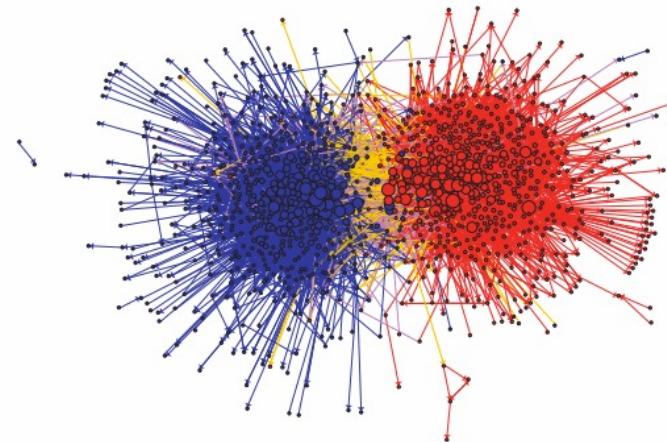


evidence for stereotype exaggeration and systematic underrepresentation of women

Kay, Matthew, Cynthia Matuszek, and Sean A. Munson. "Unequal Representation and Gender Stereotypes in Image Search Results for Occupations." *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2015.

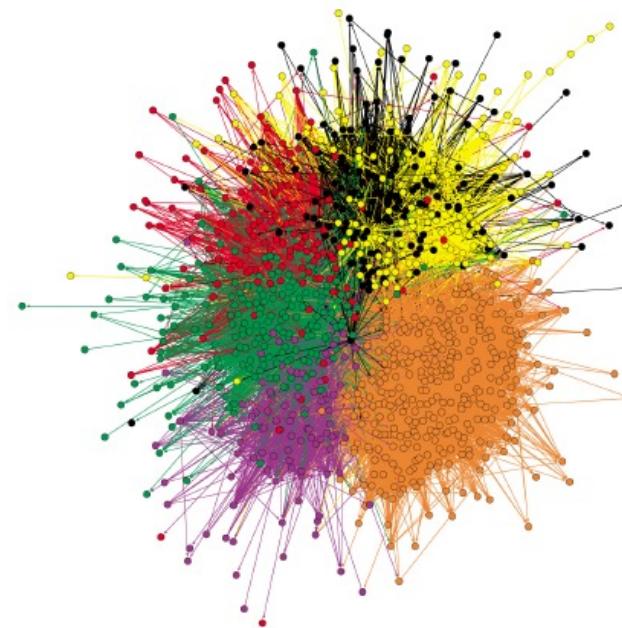
# Polarization on the web

## Weblogs



Adamic, Lada A., and Natalie Glance. "The political blogosphere and the 2004 US election: divided they blog." *Proceedings of the 3rd international workshop on Link discovery*. ACM, 2005.

## Twitter



H. Lietz, C. Wagner, A. Bleier, and M. Strohmaier. When politicians talk: Assessing online conversational practices of political parties on twitter. In International AAAI Conference on Weblogs and Social Media (ICWSM2014), Ann Arbor, MI, USA, June 2-4, 2014.

# *The social sciences have finally found their telescope ...*

Duncan Watts [2012-2014]

The Online world:

- acts as a membrane for social phenomena
- renders the unmeasurable measurable

Revolutions in science are often preceded  
by revolutions in measurements.

Networks as new kinds of data!

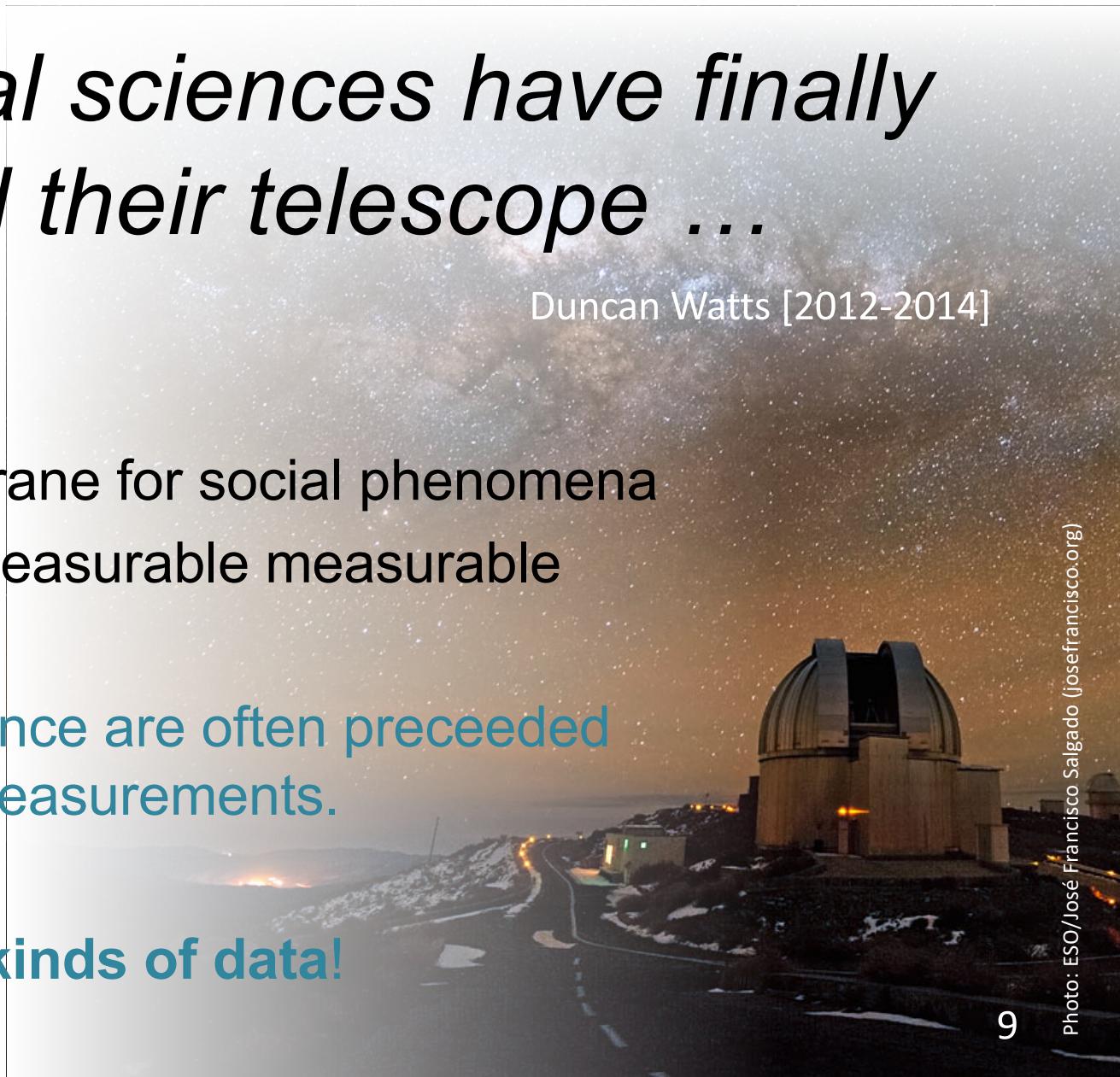


Photo: ESO/José Francisco Salgado (josefrancisco.org)

# This talk

measuring

- (i) sociological phenomena
  - Gender inequality on Wikipedia
- (ii) political phenomena
  - Emergence of a political movement in Germany
  - Political communication during the German national elections 2013

via networks and computational approaches.

Gender inequality on Wikipedia

## **MEASURING SOCIOLOGICAL PHENOMENA**

C. Wagner, D. Garcia, J. Mohsen, and M. Strohmaier. It's a man's Wikipedia? assessing gender inequality in an online encyclopedia. In International AAAI Conference on Web and Social Media (ICWSM2015), Oxford, UK, May 26-29, 2015.

# Gender Inequality on Wikipedia?

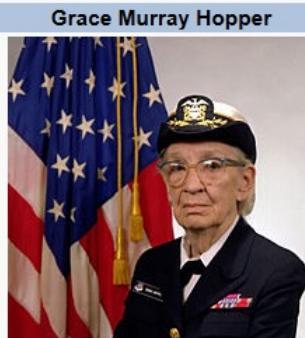
The Wikipedia Editor Community [2010]



## Grace Hopper

From Wikipedia, the free encyclopedia

**Grace Murray Hopper** (December 9, 1906 – January 1, 1992) was an American computer scientist and United States Navy rear admiral.<sup>[1]</sup> She was one of the first programmers of the [Harvard Mark I](#) computer in 1944,<sup>[2]</sup> and invented the first [compiler](#) for a computer programming language,<sup>[3][4][5][6][7]</sup> and was one of those who popularized the idea of machine-independent programming languages, which led to the development of [COBOL](#), one of the first [high-level](#)



## Donald Knuth

From Wikipedia, the free encyclopedia  
(Redirected from [Donald knuth](#))

**Donald Ervin Knuth** ([/kəˈnuθ/](#) *ku-NOOTH*; born January 10, 1938) is an American computer scientist, mathematician, and professor emeritus at Stanford University.<sup>[3]</sup> He is the author of the multi-volume work *The Art of Computer Programming*.<sup>[4]</sup> Knuth has been called the "father of the analysis of algorithms".<sup>[5]</sup> He contributed to the development of the rigorous analysis of the computational complexity of algorithms and systematized formal mathematical techniques for it. In the process he also popularized



Are notable men/women represented differently on Wikipedia?

guided-missile destroyer USS *Hopper*

Died

January 1, 1992 (aged 85)  
Arlington, Virginia, U.S.

Computer Science

Coverage, Visibility, Lexical, Structural



**WIKIPEDIA**  
The Free Encyclopedia

Main page  
Contents  
Featured content  
Current events  
Random article  
Donate to Wikipedia  
Wikipedia store

Interaction  
Help  
About Wikipedia  
Community portal  
Recent changes  
Contact page

Tools  
What links here  
Related changes  
Upload file  
Special pages  
Permanent link

Not logged in Talk Contributions Create account Log in

Article Talk

Read Edit View history Search

# Gender bias on Wikipedia

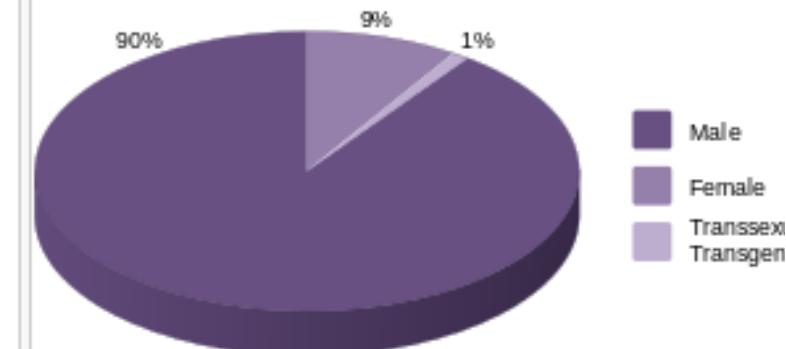
From Wikipedia, the free encyclopedia



This article is **outdated**. Please update this article to reflect recent events or newly available information. (December 2015)

**Gender bias on Wikipedia** refers to criticism of the online encyclopedia Wikipedia, and especially its English-language site, arguing that the nature and quantity of its content are biased due to the fact that a dominant majority of Wikipedia editors are male. It is among the most frequent criticisms of Wikipedia, and part of a more general criticism about systemic bias in Wikipedia. The Wikimedia Foundation, which runs Wikipedia, agrees with this criticism and has made an ongoing attempt to

Wikipedia editors are predominantly male



Surveyed Wikipedia editors were predominantly male. (Data from a 2011 Wikimedia Foundation survey of Wikipedia editors.)

**The Gender Gap on Wikipedia**

Contents [hide]

1 Research findings

# The Gender Gap Task Force

Not logged in [Talk](#) [Contributions](#) [Create account](#) [Log in](#)

Project page [Talk](#) Read Edit View history Search

## Wikipedia:WikiProject Countering systemic bias/Gender gap task force

From Wikipedia, the free encyclopedia  
< Wikipedia:WikiProject Countering systemic bias

[Talk page](#) [Participants](#) [Gender gap articles](#) [Gender gap mailing list](#) [WikiWomen's User Group](#) [Related WikiProjects](#) [Gender studies portal](#)

*Further information: Gender bias on Wikipedia and Meta:Address the gender gap*

Welcome to the **gender gap task force (GGTF)**, one of the task forces of [WikiProject Countering systemic bias](#). If you'd like to help, please add your name to the [list of participants](#).

According to *New York Magazine*, "Wikipedia famously bears one of the steepest gender gaps in contemporary culture."<sup>[1]</sup> Estimates of female Wikipedians range from 8.5 to 16.1 percent.<sup>[2][3]</sup>

Wikipedia "shares many characteristics with the hard-driving hacker crowd," according to [Joseph Reagle](#) of Harvard's [Berkman Center for Internet and Society](#), and is "open to very difficult, high-conflict people, even misogynists."<sup>[4]</sup> Issues discouraging women include a perception of Wikipedia as sexist; a perception of it as sexual, particularly because of the inclusion of pornography; lack of confidence; less spare time; problems with the editing interface; the belief that editing is tedious; and fewer opportunities to develop social relationships than on other websites.<sup>[1][5][6]</sup> Despite significant evidence to the contrary, certain commentators have denied that the gender gap is a problem, citing efforts to highlight it as "feminist ideology."<sup>[7]</sup>

The GGTF was started in 2013 to address some of the problems women face on Wikipedia, whether as editors or article subjects. Issues include editor interaction, dispute resolution, how admins and the Arbitration Committee are chosen, how policies are written and enforced, which articles about women are created and deleted, and [how those articles are written](#). In 2014 the [National Science Foundation](#) awarded a \$200,000 grant to two researchers to study the democratization of academic knowledge on Wikipedia and the effect on that of systemic gender bias.<sup>[8]</sup> That year the Wikimedia Foundation created the [Inspire Campaign](#) to fund projects aimed at closing the gender gap, and a new page on meta, [Address the gender gap](#). In 2015 the Foundation recognized a new user group, [WikiWomen's User Group](#).

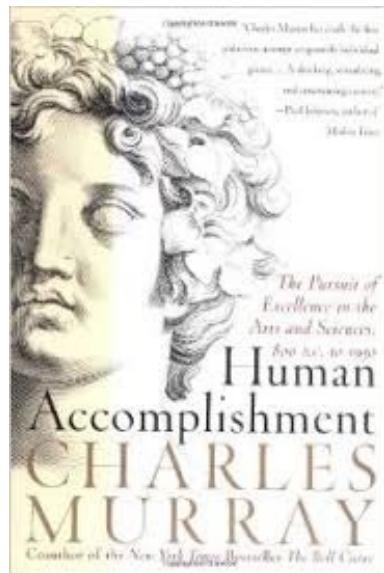


Contents [hide]

1 Participants

# Identifying Notable People

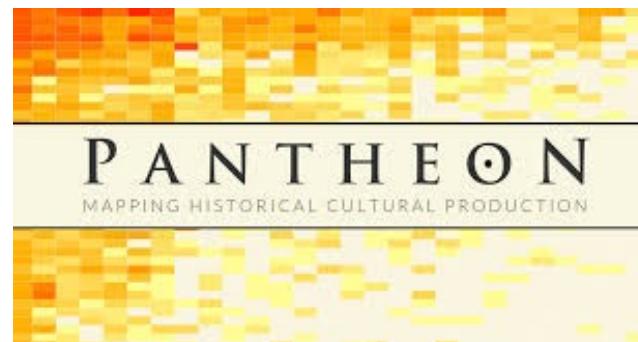
4k individuals (3% women)



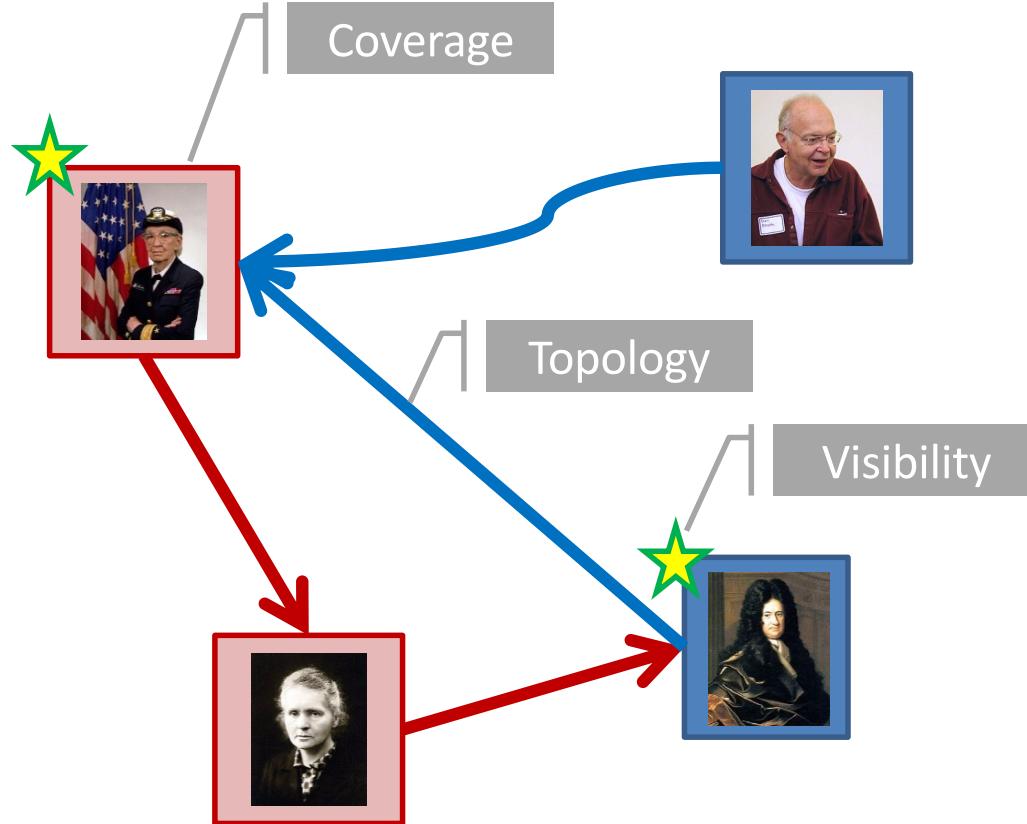
110k individuals (11% women)



11k individuals (13% women)



# Study Setup



Style

**Donald Knuth**

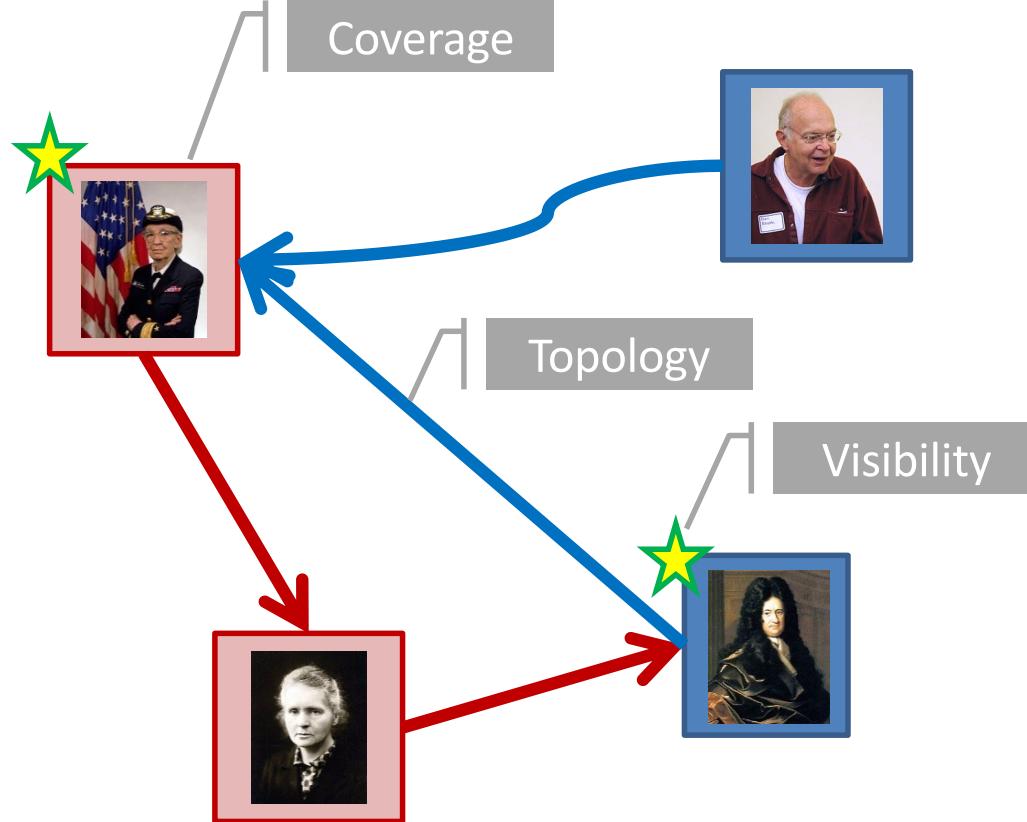
From Wikipedia, the free encyclopedia  
(Redirected from Donald knuth)

Donald Ervin Knuth (/kə'nuθ/[2] kə-NUUTH, born January 10, 1938) is an American computer scientist, mathematician, and professor emeritus at Stanford University.<sup>[3]</sup>

He is the author of the multi-volume work *The Art of Computer Programming*.<sup>[4]</sup> Knuth has been called the "father of the analysis of algorithms".<sup>[5]</sup> He contributed to the development of the rigorous analysis of the computational complexity of algorithms and systematized formal mathematical techniques for it. In the process he also popularized the asymptotic notation. In addition to fundamental contributions in several branches of theoretical computer science, Knuth is the creator of the TeX computer typesetting system, the related METAFONT font definition language and rendering system, and the Computer Modern family of typefaces.

List of notable people from: Pantheon, Human Accomplishment, Freebase

# Study Setup



Style

**Donald Knuth**

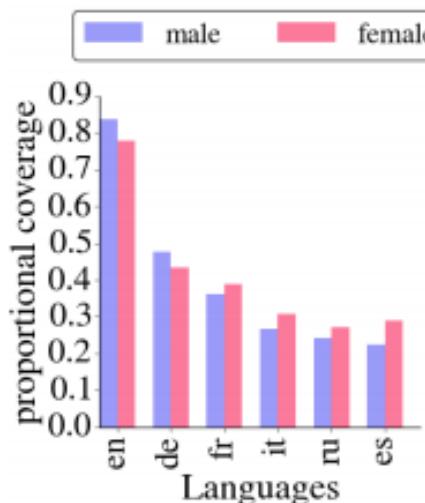
From Wikipedia, the free encyclopedia  
(Redirected from Donald knuth)

**Donald Ervin Knuth** (/kəˈnuθ/<sup>[2]</sup> *ku-Nooth*, born January 10, 1938) is an American computer scientist, mathematician, and professor emeritus at Stanford University.<sup>[3]</sup>

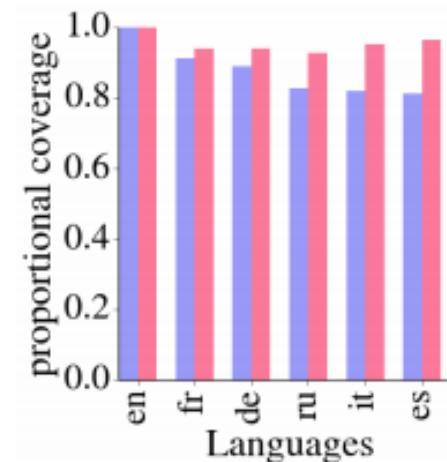
He is the author of the multi-volume work *The Art of Computer Programming*.<sup>[4]</sup> Knuth has been called the "father of the analysis of algorithms".<sup>[5]</sup> He contributed to the development of the rigorous analysis of the computational complexity of algorithms and systematized formal mathematical techniques for it. In the process he also popularized the *asymptotic notation*. In addition to fundamental contributions in several branches of theoretical computer science, Knuth is the creator of the TeX computer typesetting system, the related METAFONT font definition language and rendering system, and the Computer Modern family of typefaces.

List of notable people from: Pantheon, Human Accomplishment, Freebase

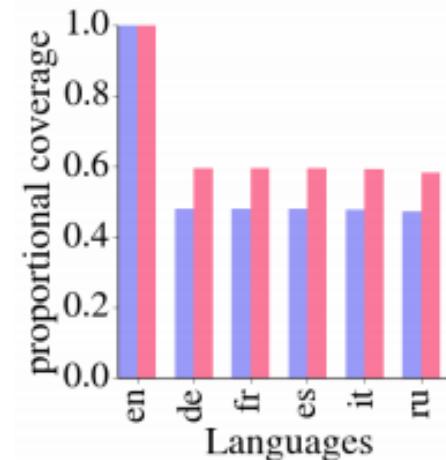
# Coverage: Women are well covered



(a) Freebase

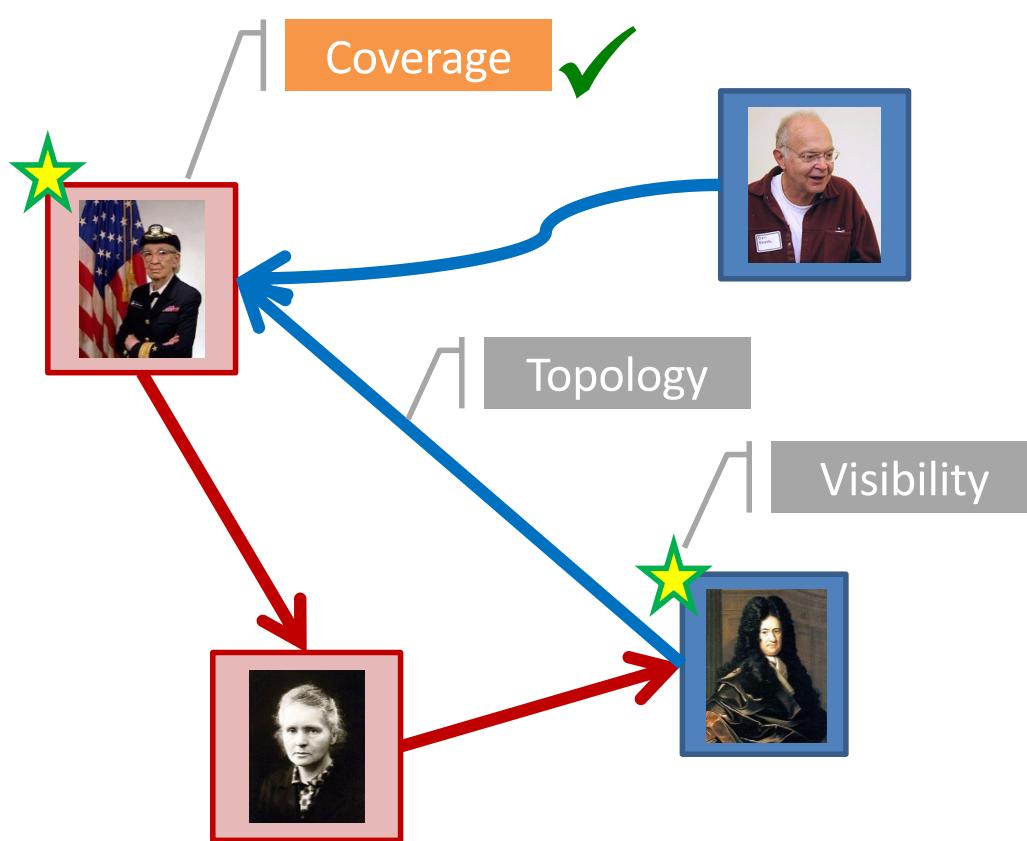


(b) HA



(c) Pantheon

# Study Setup



Style

**Donald Knuth**

From Wikipedia, the free encyclopedia  
(Redirected from Donald knuth)

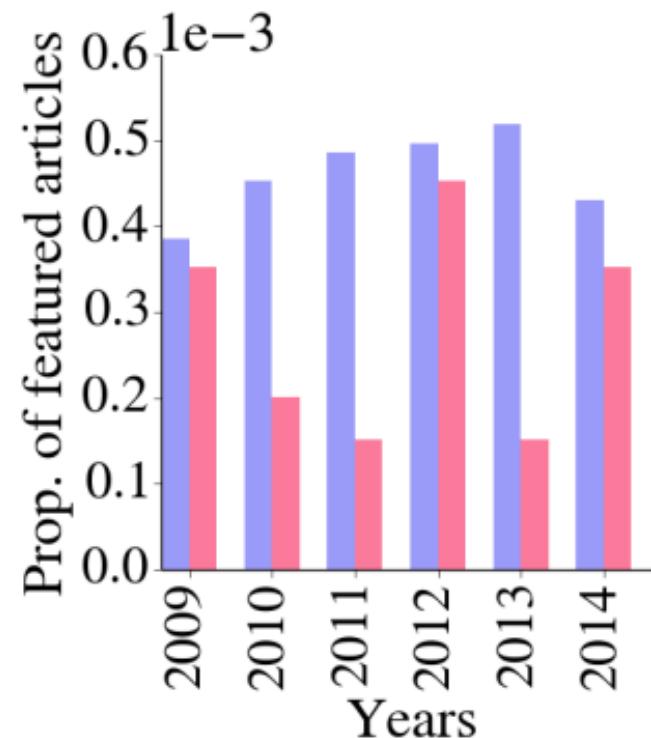
**Donald Ervin Knuth** (/kəˈnuθ/; born January 10, 1938) is an American computer scientist, mathematician, and professor emeritus at Stanford University.<sup>[3]</sup>

He is the author of the multi-volume work *The Art of Computer Programming*.<sup>[4]</sup> Knuth has been called the "father of the analysis of algorithms".<sup>[5]</sup> He contributed to the development of the rigorous analysis of the computational complexity of algorithms and systematized formal mathematical techniques for it. In the process he also popularized the asymptotic notation. In addition to fundamental contributions in several branches of theoretical computer science, Knuth is the creator of the TeX computer typesetting system, the related METAFONT font definition language and rendering system, and the Computer Modern family of typefaces.

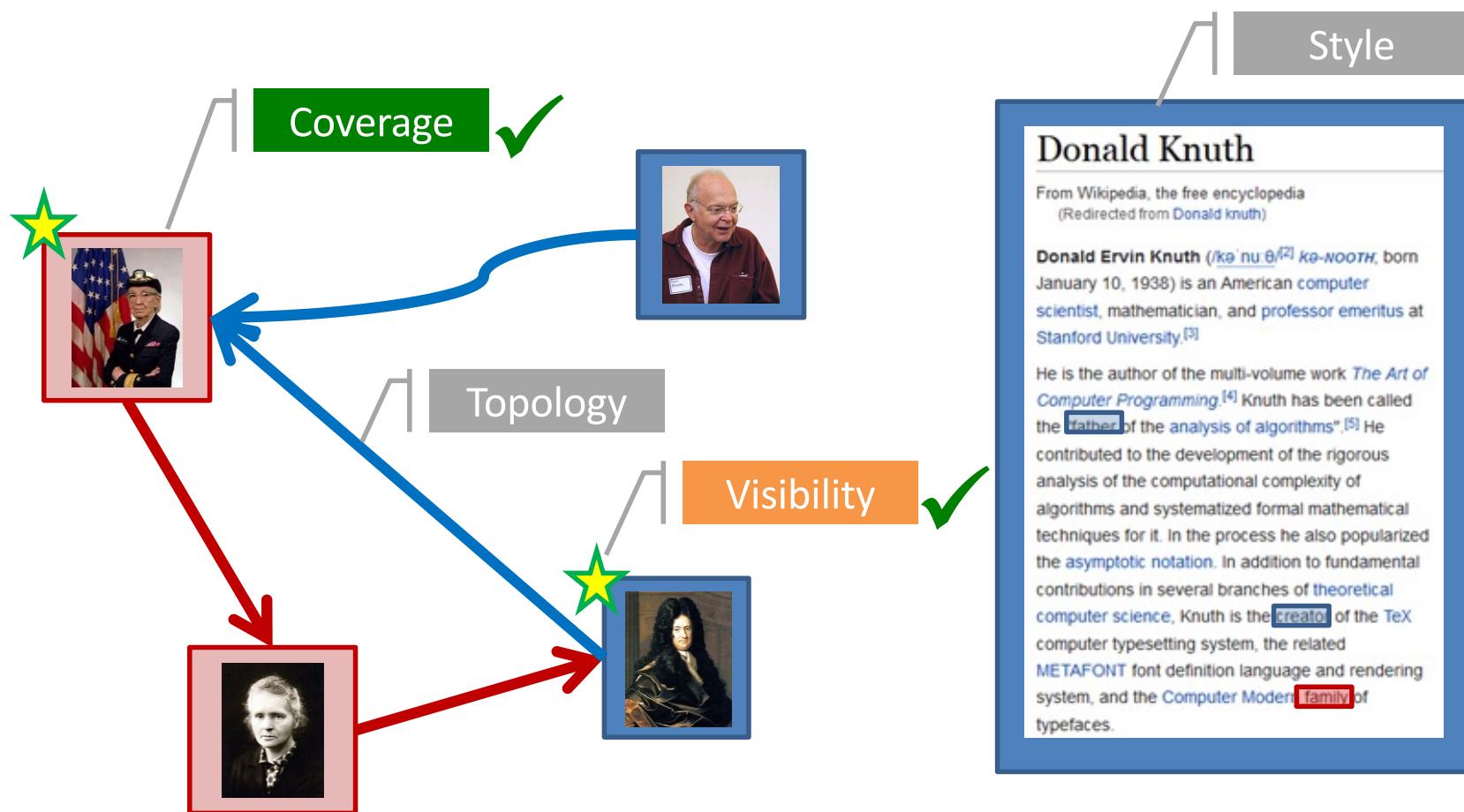
List of notable people from: Pantheon, Human Accomplishment, Freebase

# Visibility: both genders are equally visible

- m/f biographies  
**equally promoted to**  
the Wikipedia  
frontpage
- Differences not  
**significant**



# Study Setup



List of notable people from: Pantheon, Human Accomplishment, Freebase

# Style: Women are described differently

y



Fox, M. F.; Johnson, D. G.; and Rosser, S. V. 2006. Women, gender, and technology. University of Illinois Press.

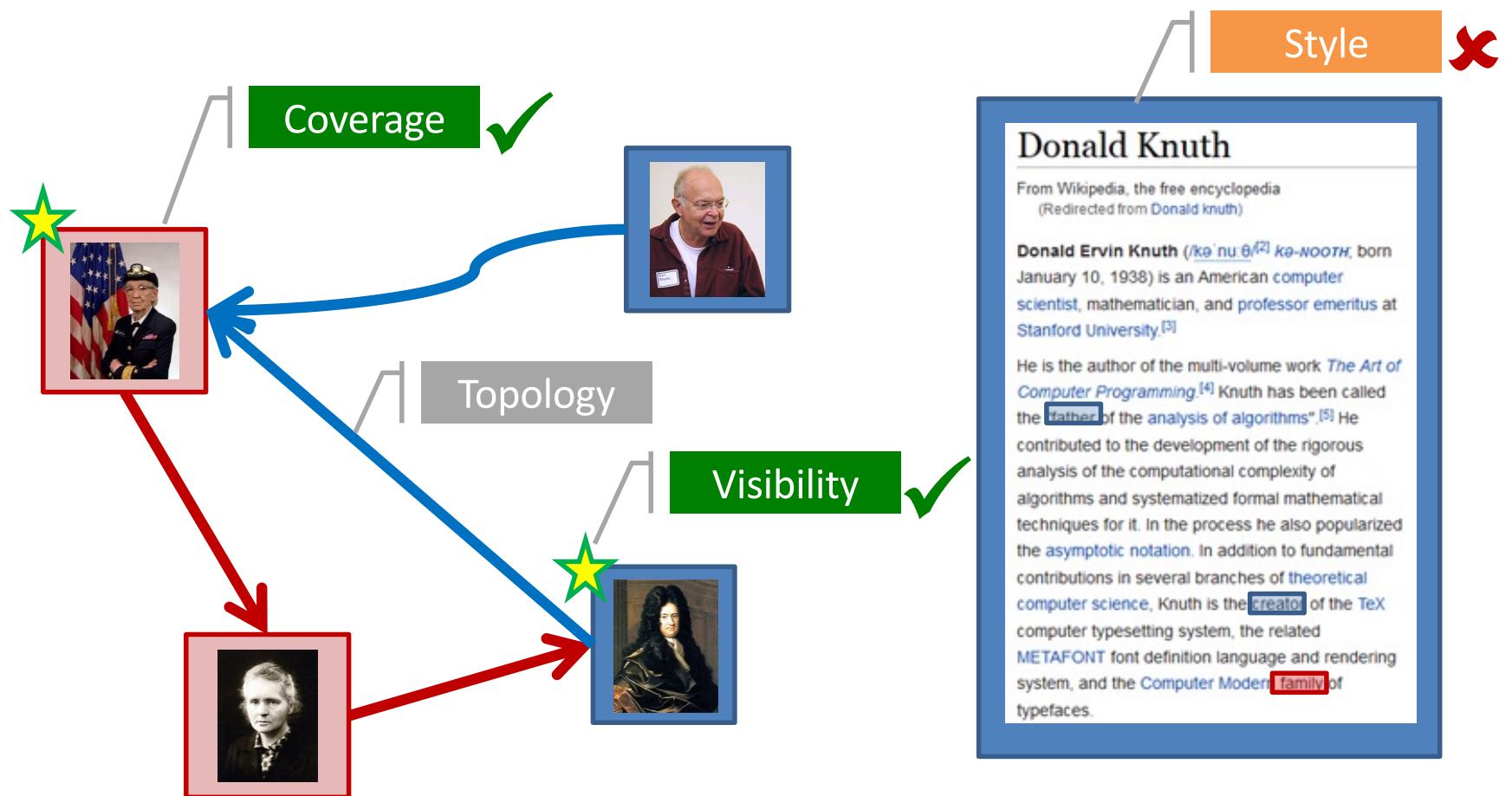
# Style:

## Discriminative words for women

- **train a NB-classifier**
- **TFIDF scores** of word stems as features
- **Log likelihood ratios** for feature assessment
- Words for men **do not emphasize gender or relations**

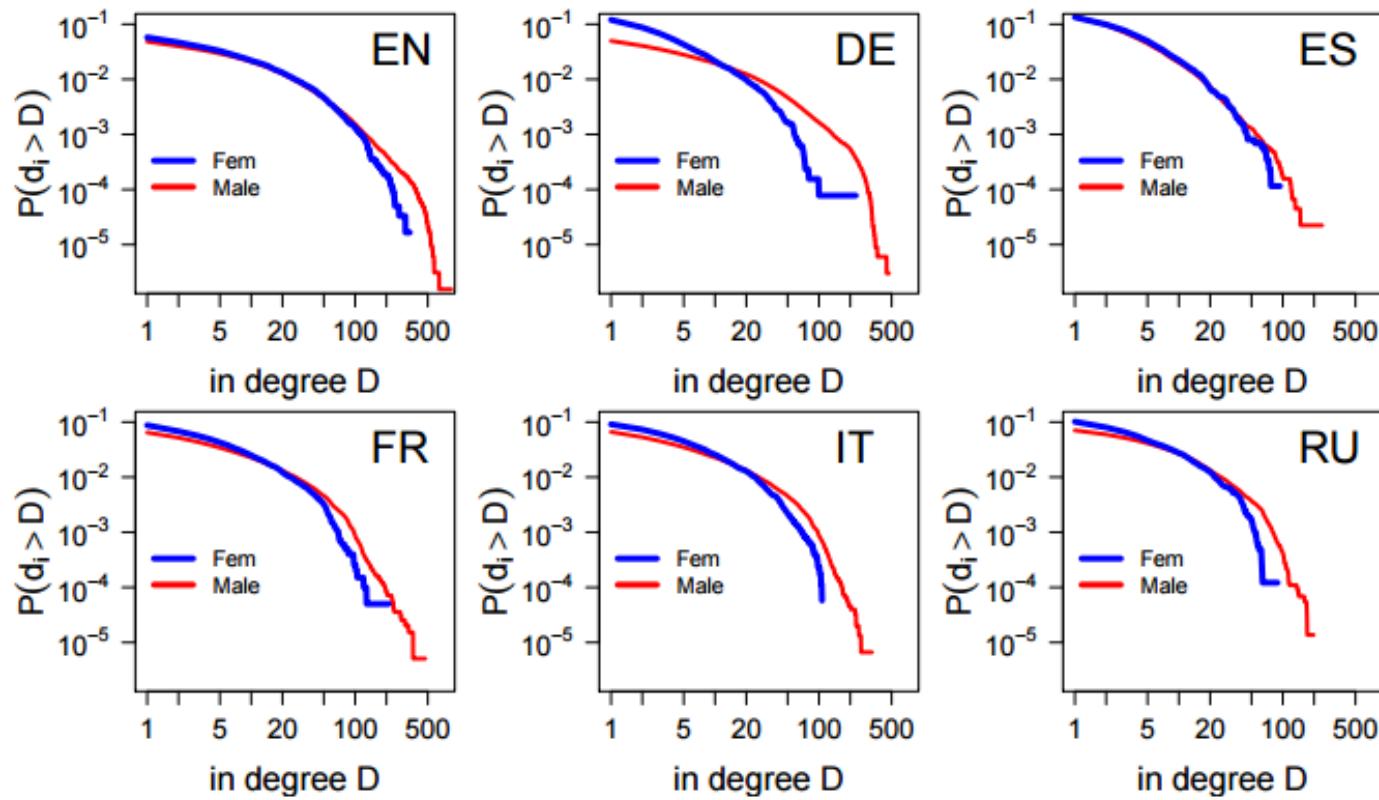
Category	Term	Female	Male
Relationship	husband	9.2	1.0
Gender	female	8.2	1.0
Family	aunt	6.5	1.0
Gender	women	6.4	1.0
Gender	madam	6.1	1.0
Gender	woman	5.6	1.0
Family	grandmoth	5.5	1.0
Gender	girl	5.3	1.0
Gender	mrs	4.9	1.0
Relationship	divorc	4.4	1.0
Gender	ladi	4.4	1.0
Relationship	wed	4.3	1.0
Relationship	marriag	3.8	1.0
Relationship	lover	3.8	1.0
Family	babi	3.7	1.0
Family	sister	3.5	1.0
Family	child	3.0	1.0
Family	mother	3.0	1.0

# Study Setup

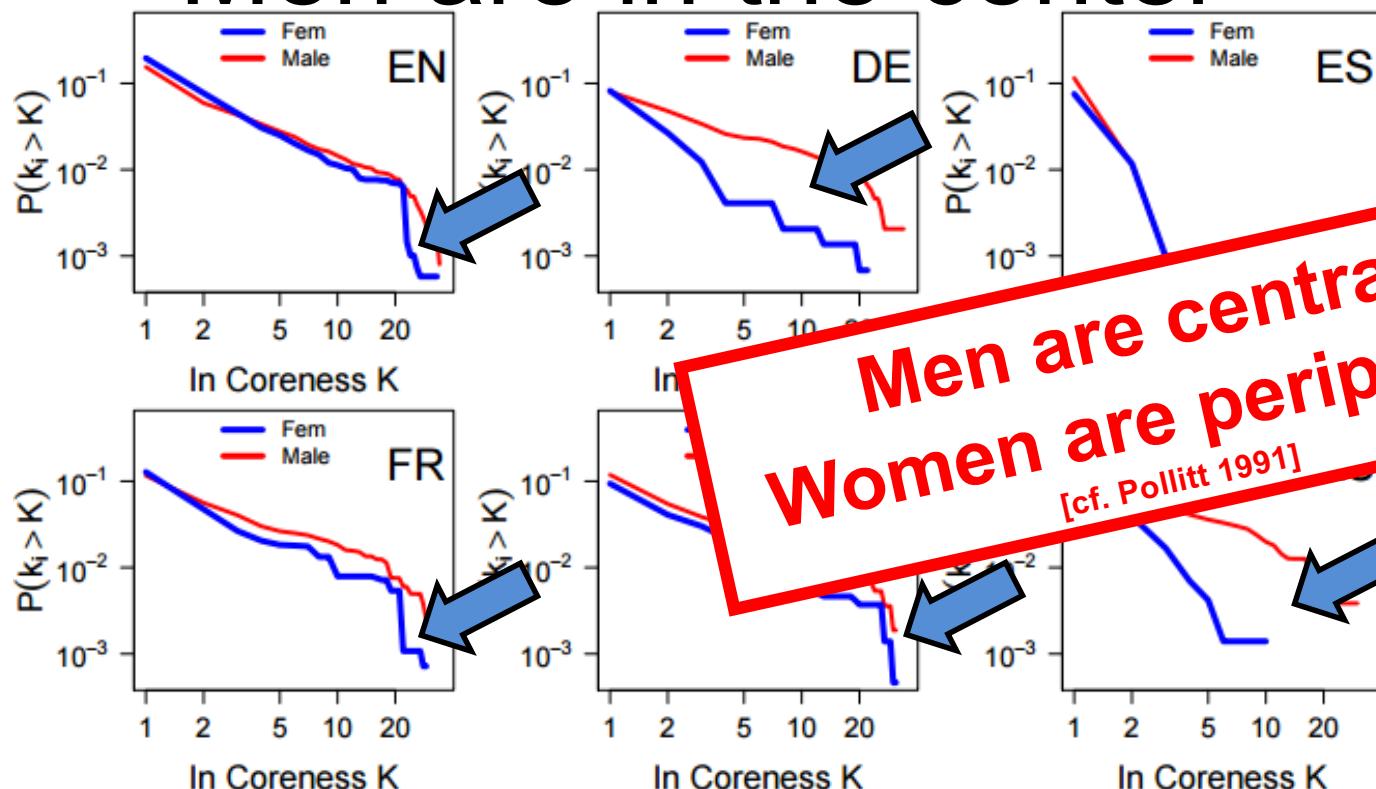


List of notable people from: Pantheon, Human Accomplishment, Freebase

# Topology: Men are better connected



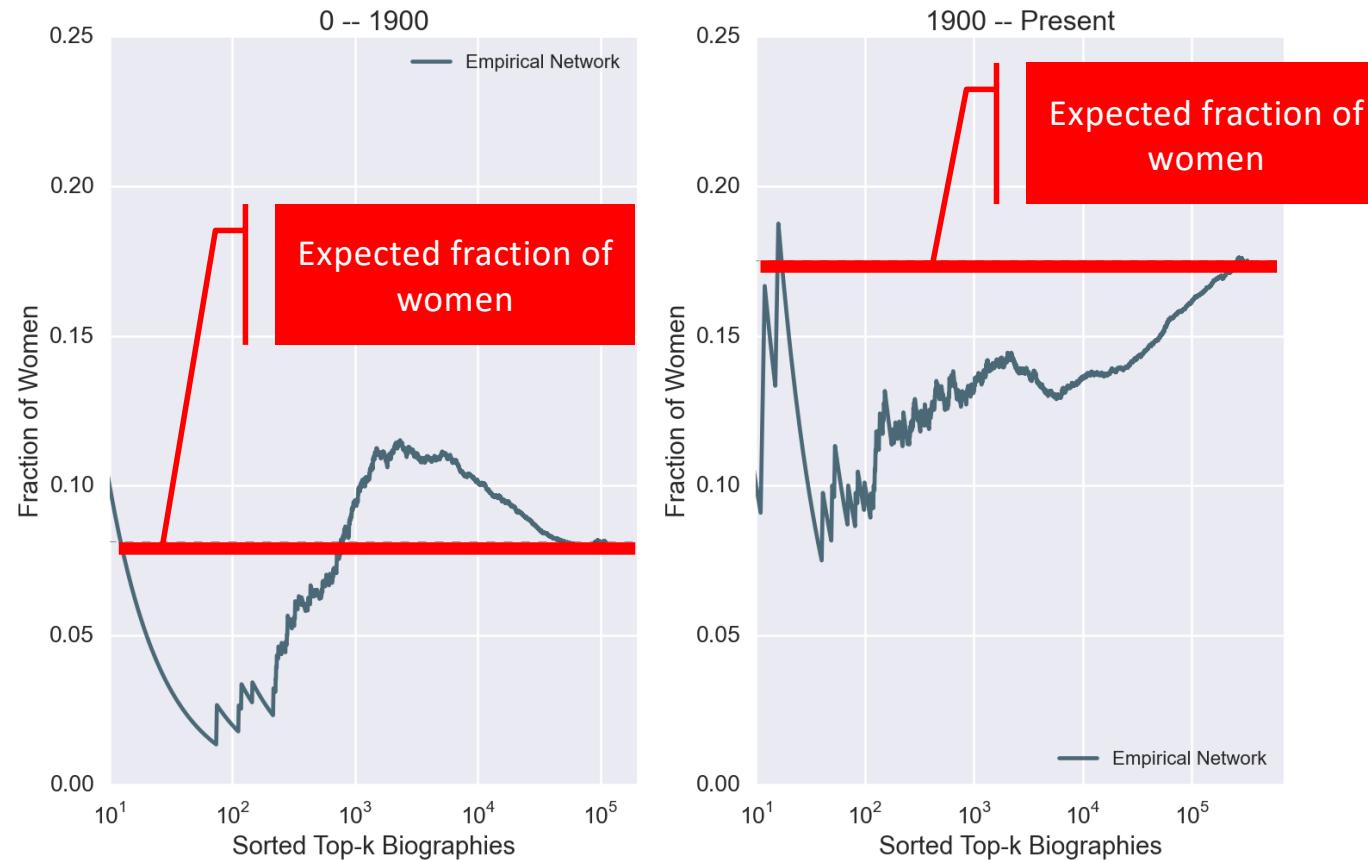
# Topology: Men are in the center



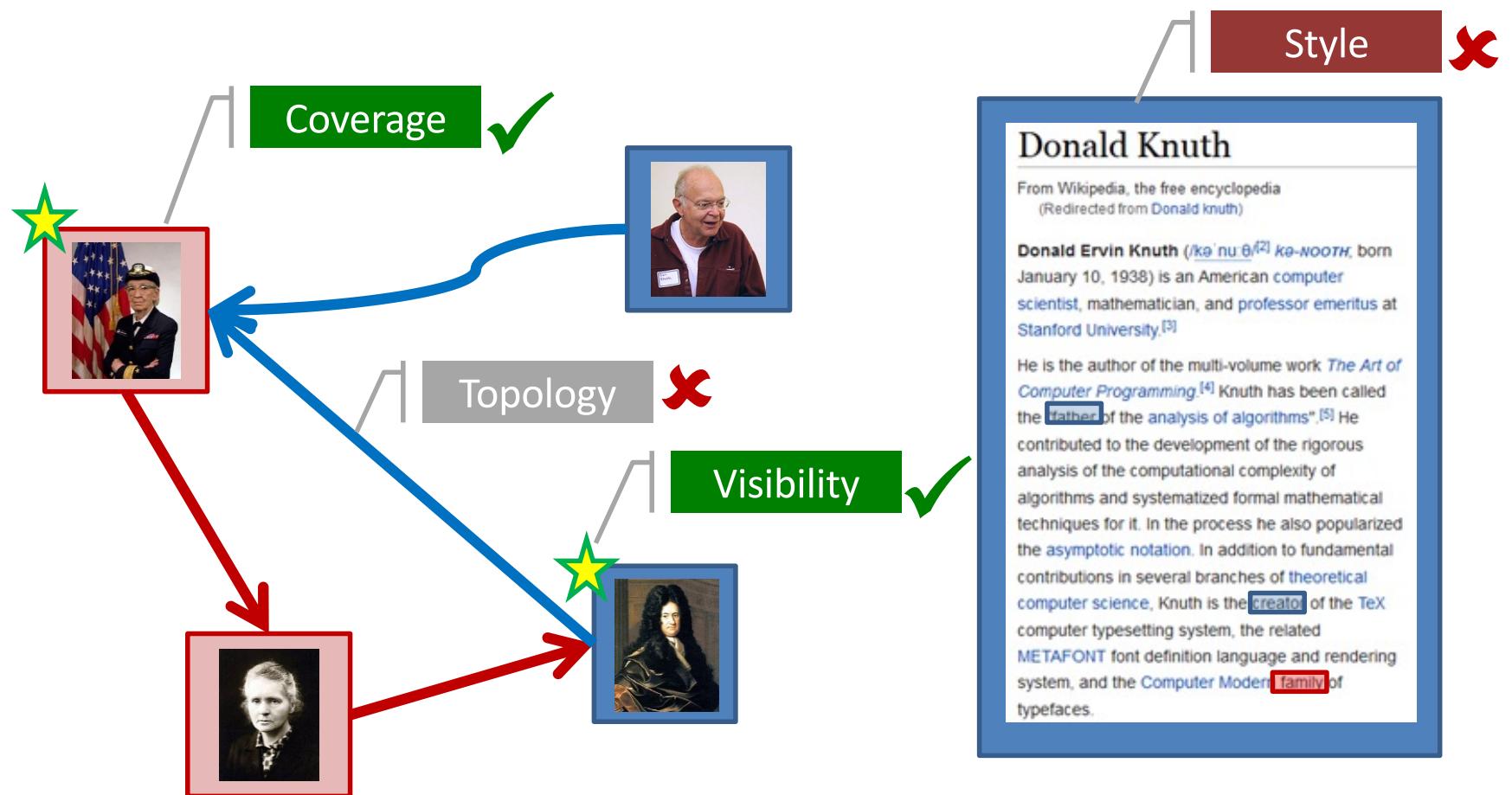
The k-core is the largest subnetwork comprising only nodes of degree at least  $k$ .

# Topology: Women rank worse

Claudia Wagner, Eduardo Graells-Garrido, David Garcia and Filippo Menczer, **Women through the glass ceiling: gender asymmetries in Wikipedia**, EPJ Data Science Journal, 2016



# Summary of Findings



# This talk

measuring

(i) sociological phenomena

- Gender inequality on Wikipedia

**(ii) political phenomena**

- Emergence of a political movement in Germany

- Political communication during the German national elections 2013

via networks and computational approaches.

Emergence of the Pirate Party in Germany

## **MEASURING POLITICAL PHENOMENA**

C. Kling, J. Kunegis, H. Hartmann, M. Strohmaier and S. Staab. Voting behaviour and power in online democracy: A study of LiquidFeedback in Germany's pirate party. In International AAAI Conference on Web and Social Media (ICWSM2015), Oxford, UK, May 26-29, 2015.

# The Pirate Party – an emerging and global political movement



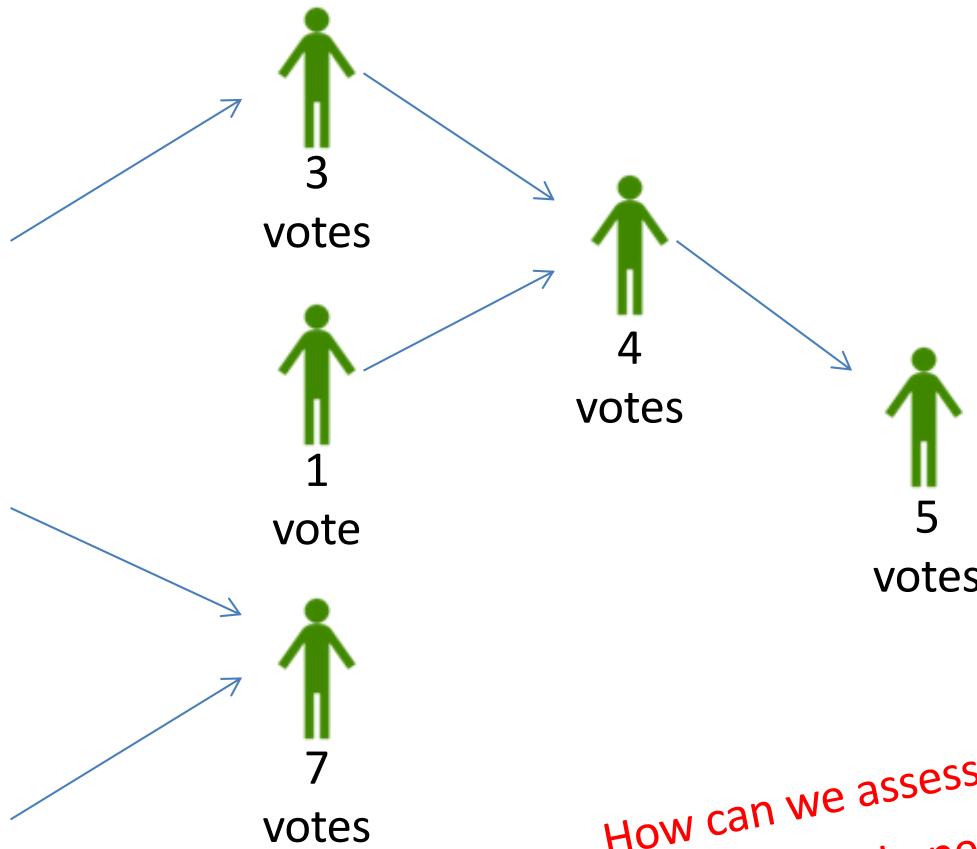
# The German Pirate Party

- 35.000 members
- LiquidFeedback as a web platform (2010-13)
  - 499,009 votes
  - 13,836 users
  - 14,964 delegations

# Structure of LiquidFeedback

- Users create *initiatives*, which are grouped by *issues* and belong to *areas*
  - Area: Environmental issues
  - Issue: CO2 output has to be reduced.
  - Initiative: Subsidise wind turbines!
- Delegations on *global*, *initiative*, *issue* and *area level*
- “Back-delegations” possible

# LiquidFeedback: An Online Delegative Democracy Platform

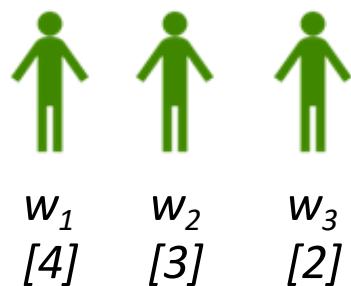


Delegations are:

- transitive
- general/specific
- ephemeral

How can we assess power of people in such a system?  
How do people vote and use power?

# Assessing power: Banzhaf power index



- q: quora
- $w_i$  weight of player i
- $\{q: w_1, w_2, \dots, w_n\}$

{6: 4,3,2}

{p1}

{p2}

{p3}

{p1, p2}

{p1, p3}

{p1, p2, p3}

{p2, p3}

Not a good power indicator

$$BPI^*(i) = \frac{\text{\#times } p_i = \text{critical}}{\text{\#times all critical}}$$

$$BPI^*(1) = 3/5$$

$$BPI^*(2) = 1/5$$

$$BPI^*(3) = 1/5$$

Key:

Loosing coalition

Winning coalition

p<sub>i</sub>... critical player

*Banzhaf power index:*

*Votes are independent*

*Shapley power index:*

*Votes are homogeneous*

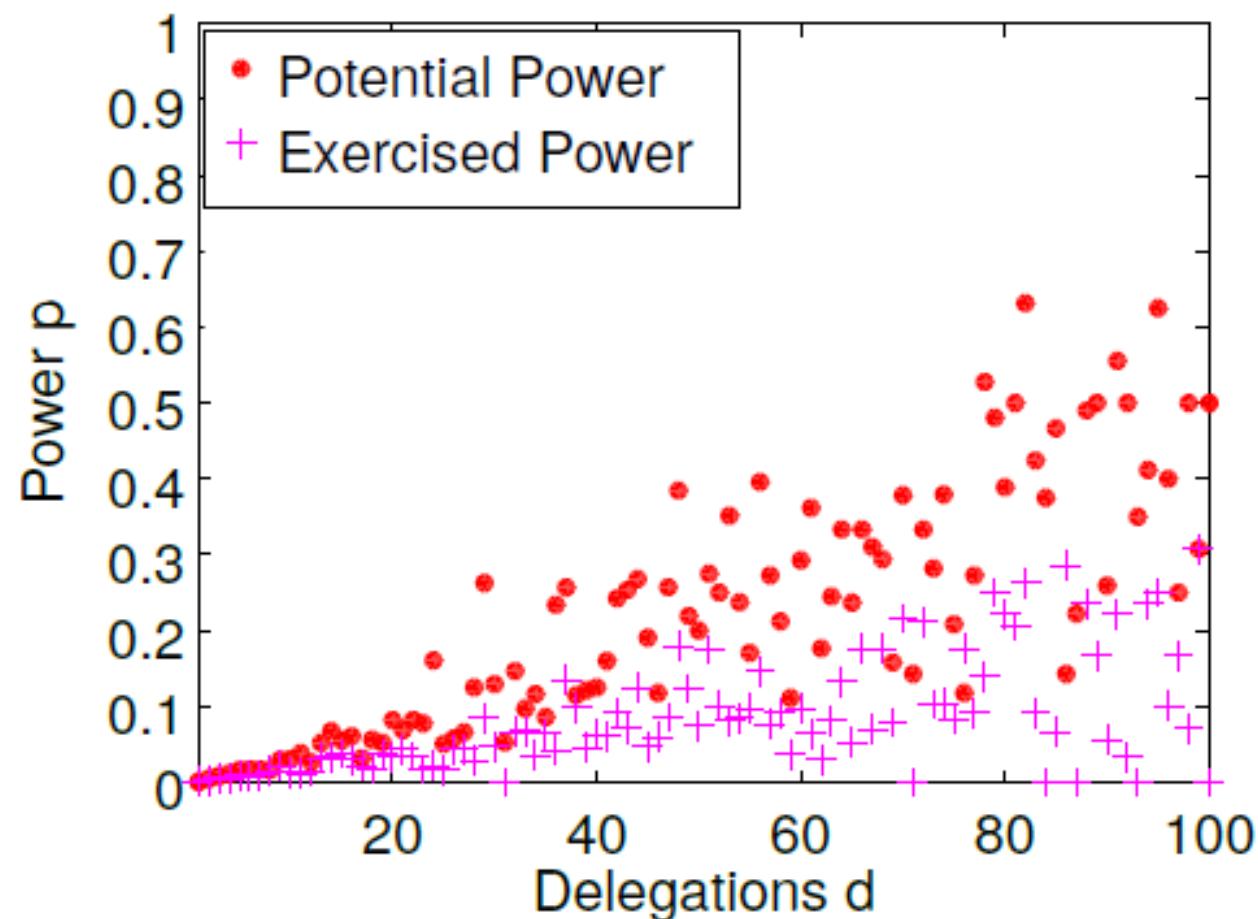
*Potential Power:*

*Measured power in the dataset*

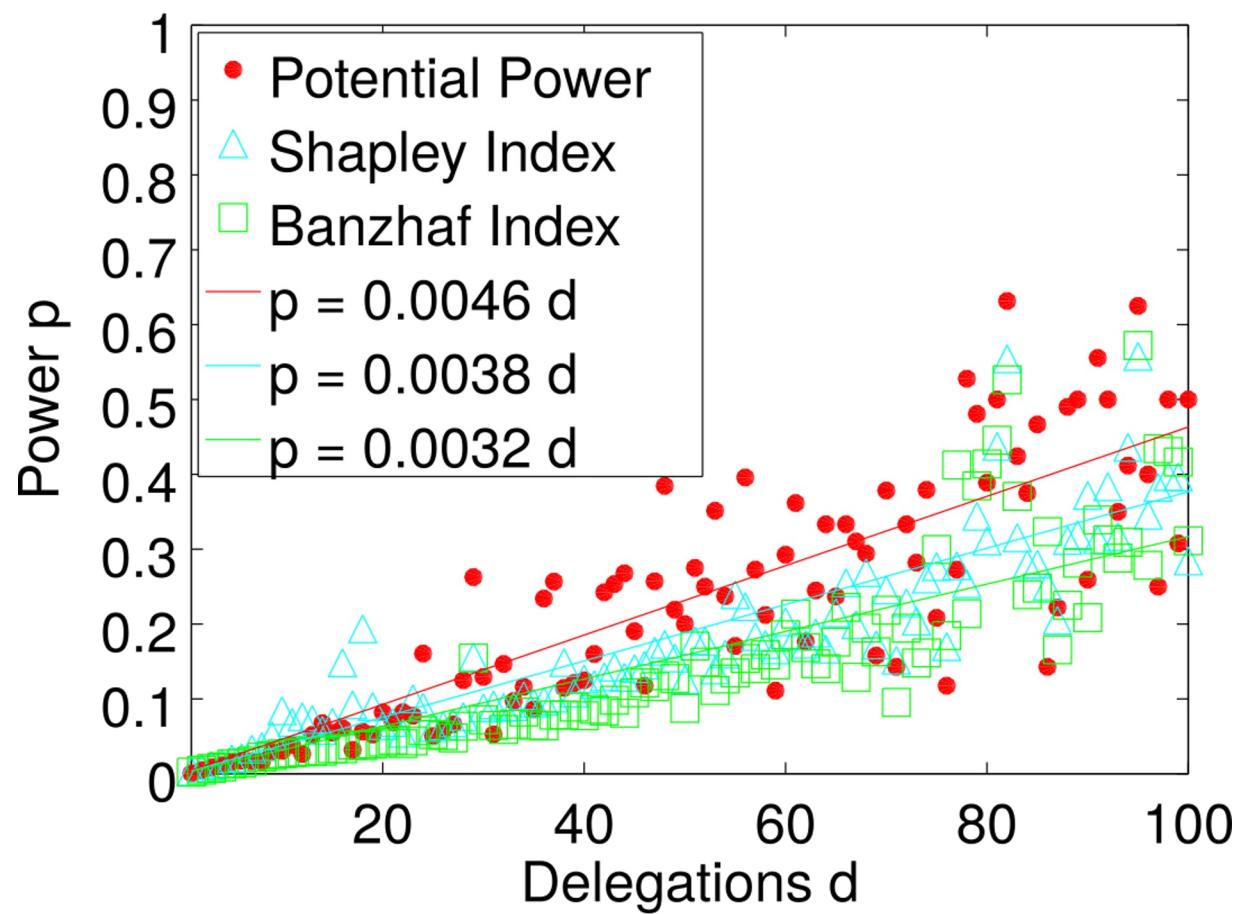
*Exercised Power:*

*Power used to actually turn votes*

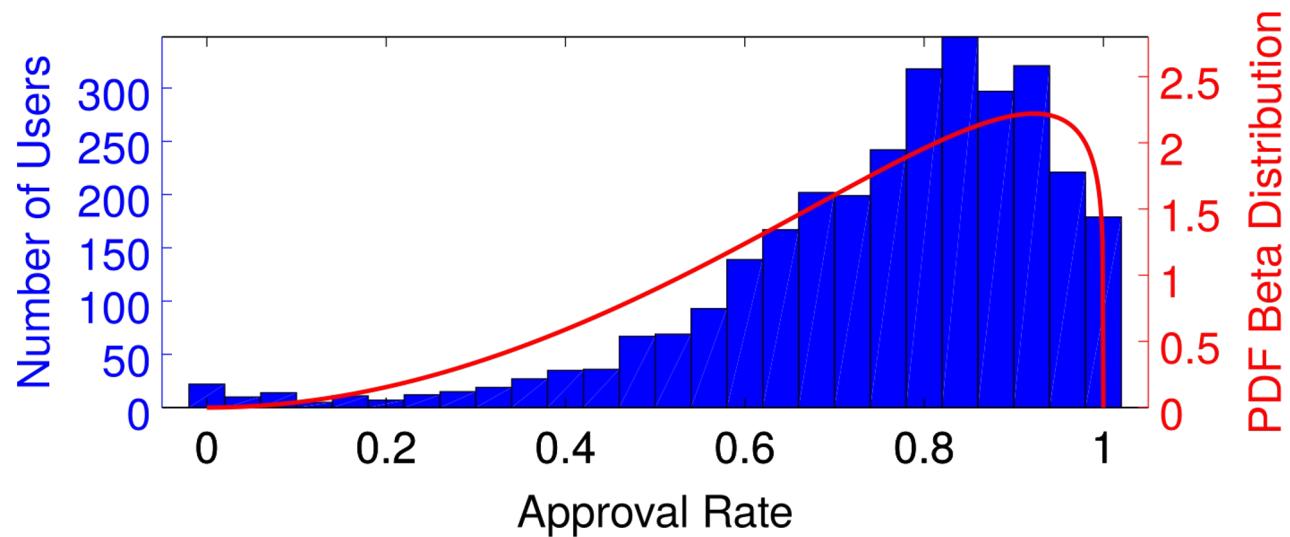
# Empirical Power



# Power Indices

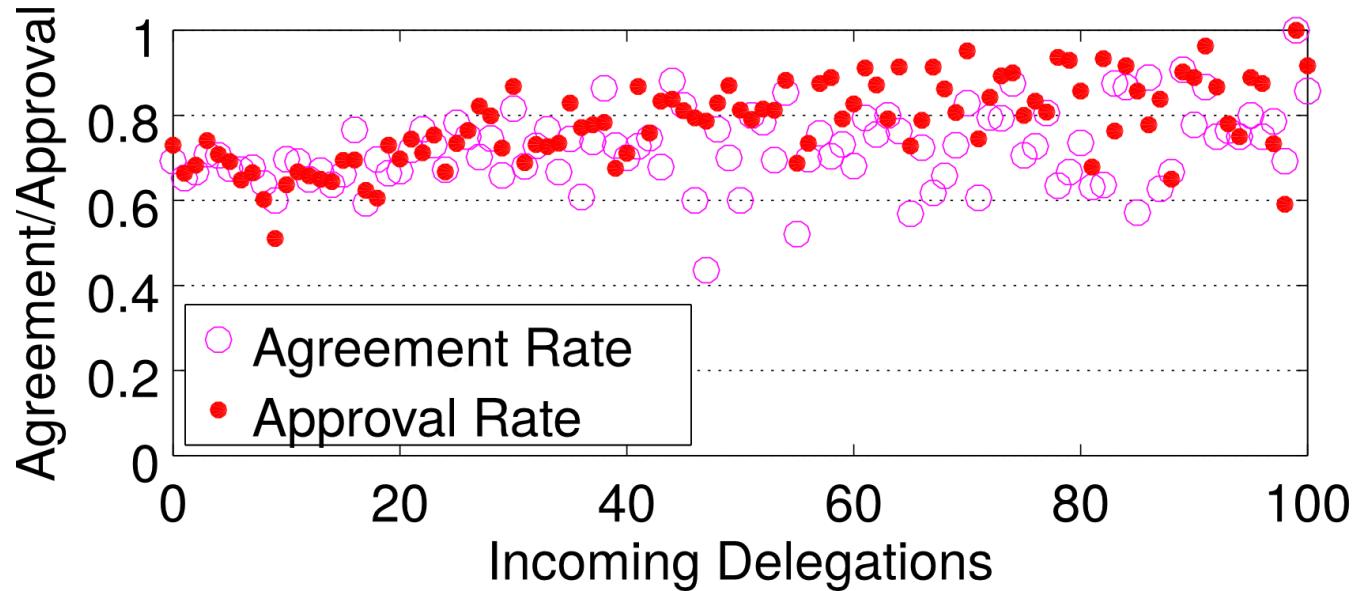


# Average Approval Rate



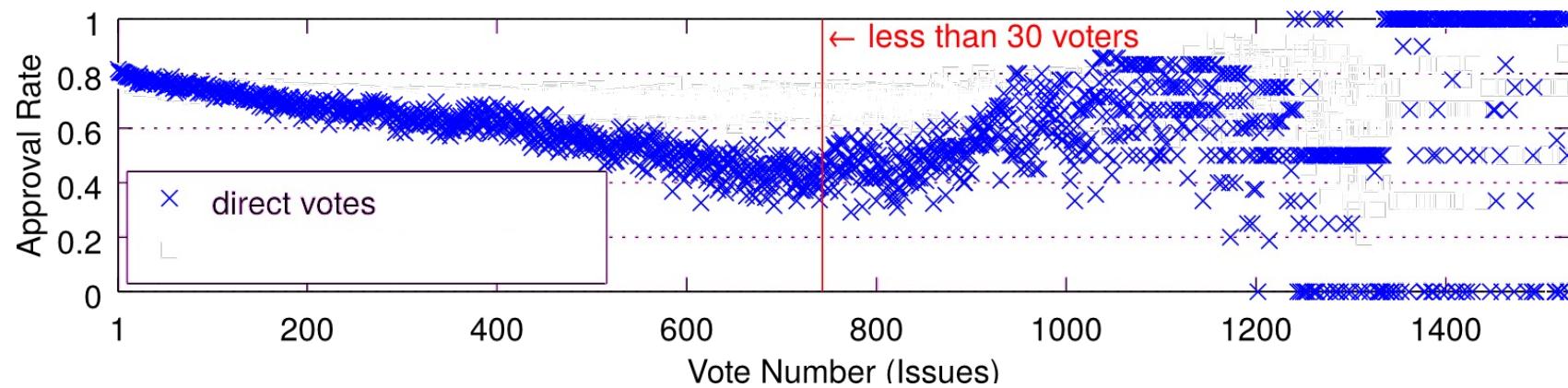
*(How many users agree with x% of all voted proposals?)*

# Agreement & Approval



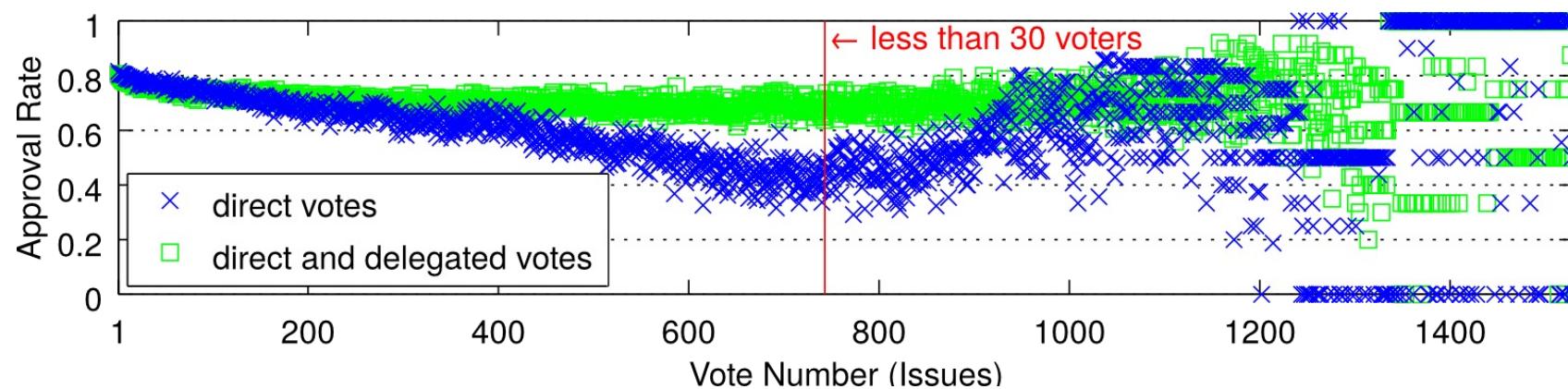
Voters with many delegations tend to vote more positive and tend to agree more with the majority

# *Approval Rate*



→ Approval rate decreases with voting experience

# *Approval Rate*



→ Delegates stabilise the approval rate?

# Implications

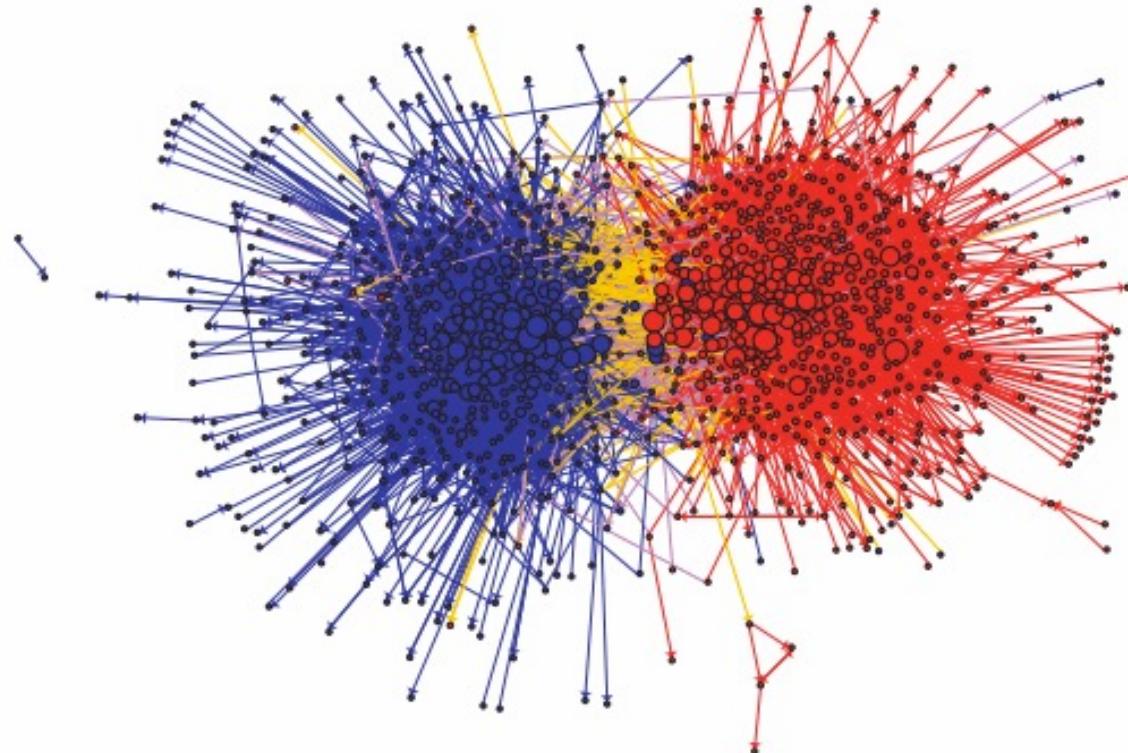
- Previously unobservable political processes become visible
- Existing measures of political power need to be reevaluated in the light of new data
- Networked voting data provides unique insights into how political movements shape up and evolve

The 2013 German National Elections on Twitter

## **MEASURING POLITICAL PHENOMENA**

H. Lietz, C. Wagner, A. Bleier, and M. Strohmaier. When politicians talk: Assessing online conversational practices of political parties on twitter. In *International AAAI Conference on Weblogs and Social Media (ICWSM2014)*, Ann Arbor, MI, USA, June 2-4, 2014.

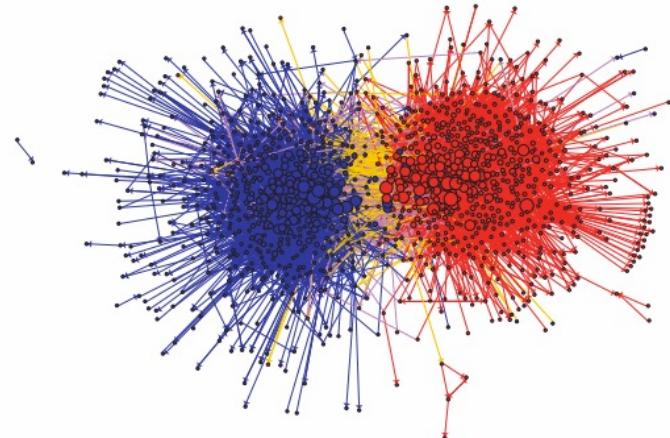
# The blogosphere in the 2004 US presidential elections



Adamic, Lada A., and Natalie Glance. "The political blogosphere and the 2004 US election: divided they blog." *Proceedings of the 3rd international workshop on Link discovery*. ACM, 2005.

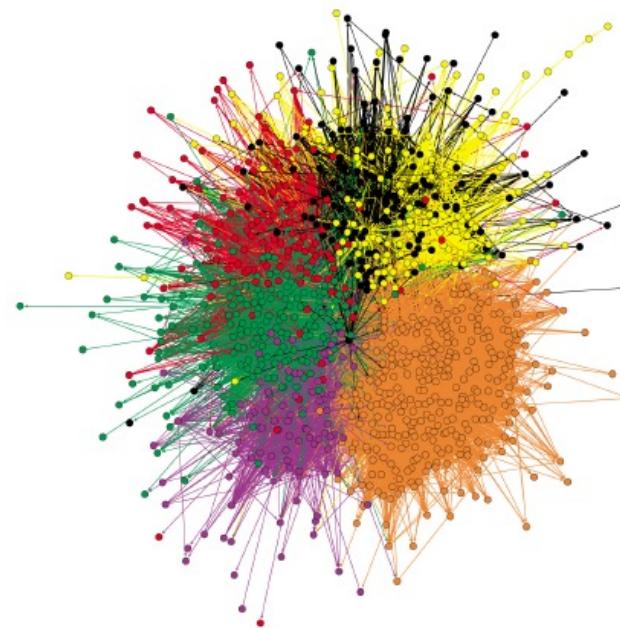
# Political processes on the WWW

**Weblogs (US 2004)**



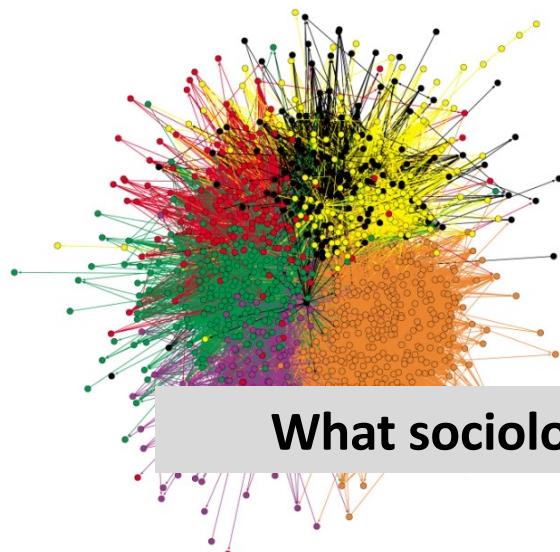
Adamic, Lada A., and Natalie Glance. "The political blogosphere and the 2004 US election: divided they blog." *Proceedings of the 3rd international workshop on Link discovery*. ACM, 2005.

**Twitter (Germany 2013)**

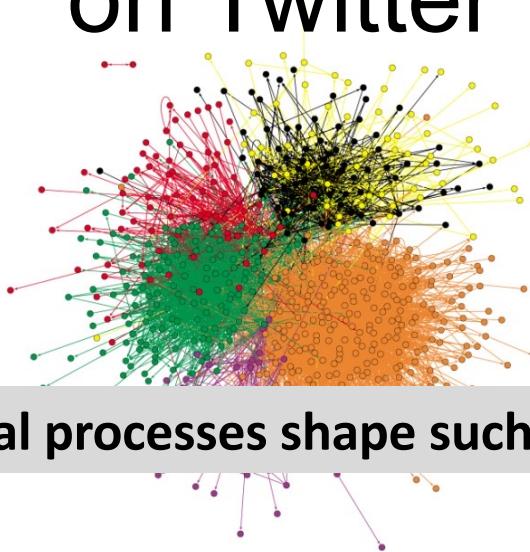


H. Lietz, C. Wagner, A. Bleier, and M. Strohmaier. When politicians talk: Assessing online conversational practices of political parties on twitter. In International AAAI Conference on Weblogs and Social Media (ICWSM2014), Ann Arbor, MI, USA, June 2-4, 2014.

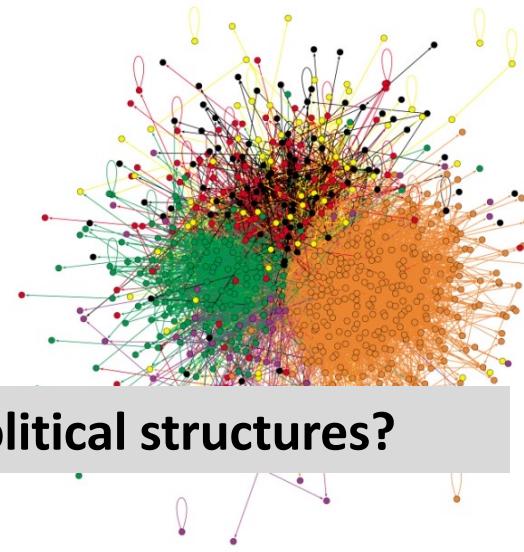
# German Politicians (n=1031) on Twitter



Politicians following  
( $H=0.83$ )



Politicians retweeting  
( $H=0.90$ )

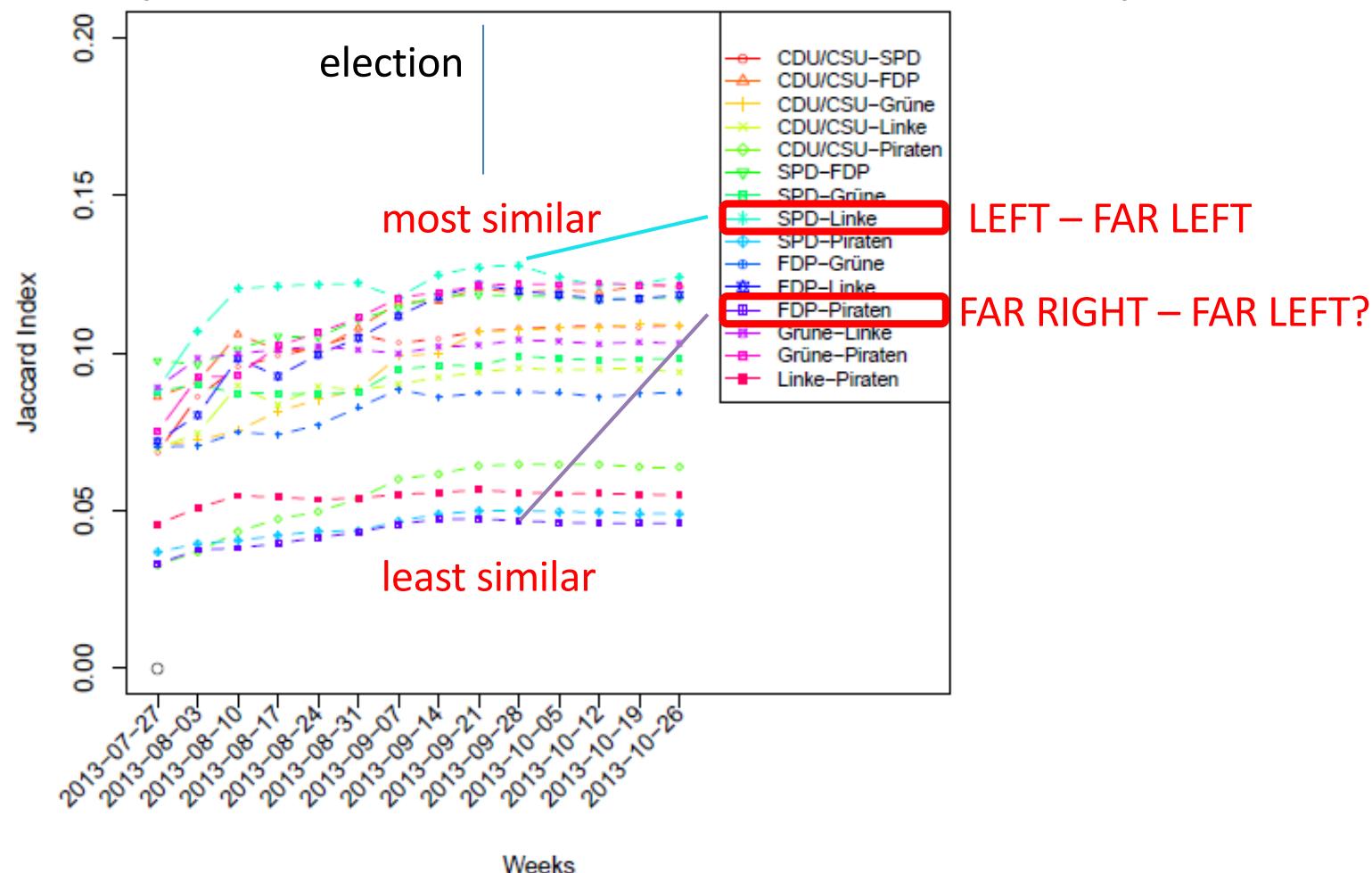


Politicians mentioning  
( $H=0.79$ )



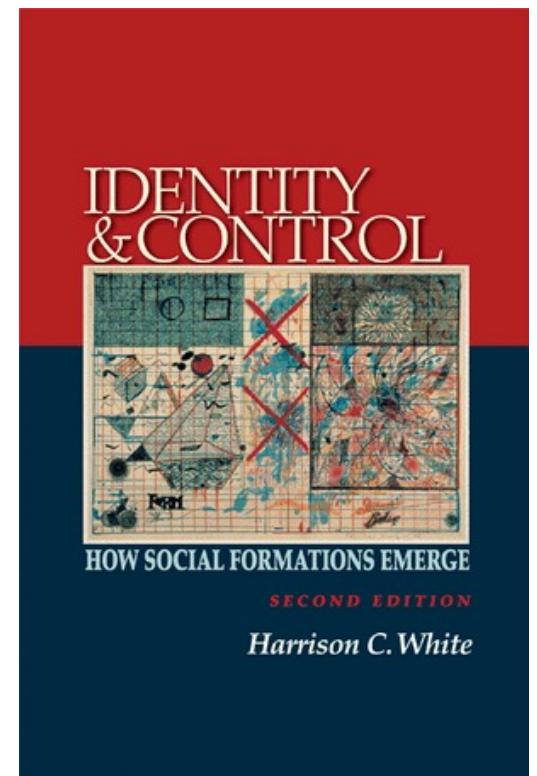
# German National Elections 2013

## Similarity between parties as measured by hashtags



# Theory & Research Design

- We measure **identity** and **control** of parties over time
- Each political party is an **entity**
- **Identity is established** through
  - semantic focus and
  - (dis-)similarity to other parties
- **Control is established** through
  - reproduction

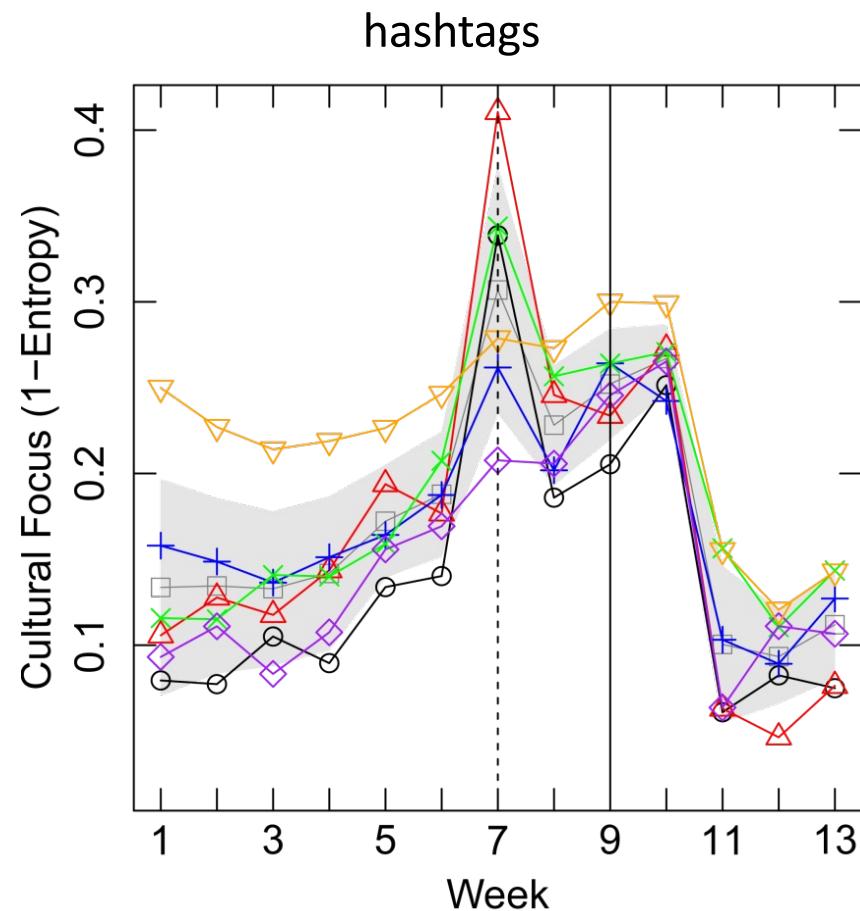


# Identity through Semantic Focus

- How strongly **does a party focus on a few facts?**
- Operationalization of control in a semantic „space“
- Measure: Based on entropy
- Ranges from 0 (no focus) to 1 (maximum focus)

$$F(\sigma_i) = 1 - \frac{-\sum_{j=1}^n p(a_j) * \log_2 p(a_j)}{\log_2(n)}$$

1 minus the normalized entropy of vector  $\sigma_i$

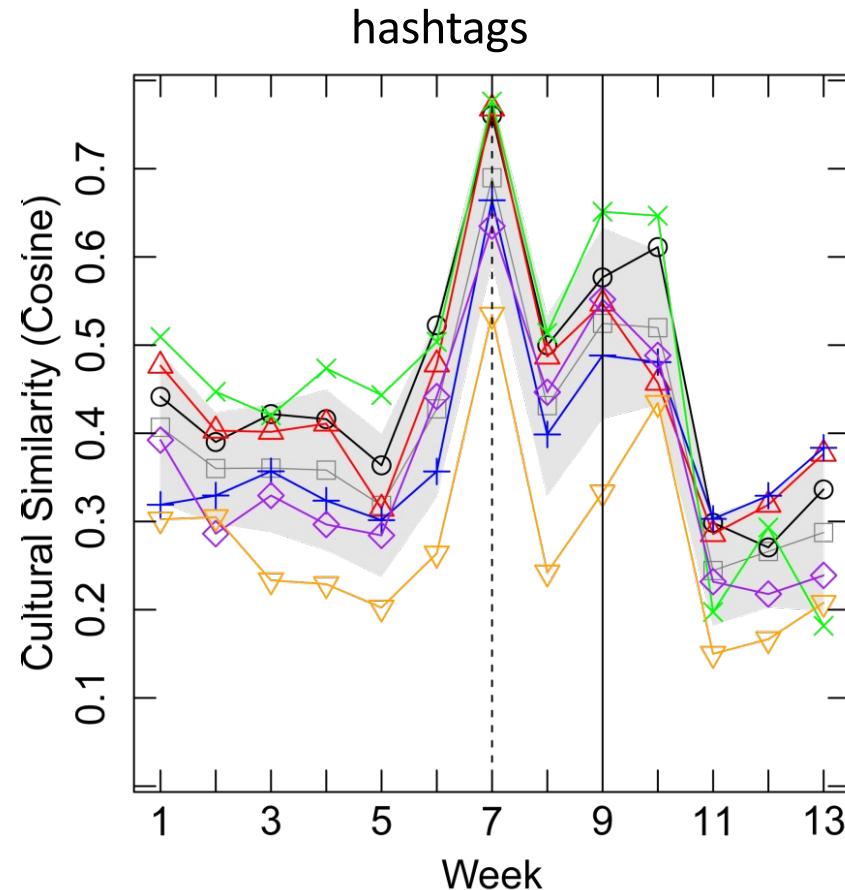


# Identity through Semantic Similarity

- How similar are parties in terms of their semantic vectors?
- Measure: Cosine similarity
- Ranges from 0 (no similarity) to 1 (maximum similarity)

$$S(\sigma_i, \sigma_j) = \frac{\sigma_i \cdot \sigma_j}{\|\sigma_i\| \|\sigma_j\|}$$

Cosine similarity of vectors  $\sigma_i$  and  $\sigma_j$

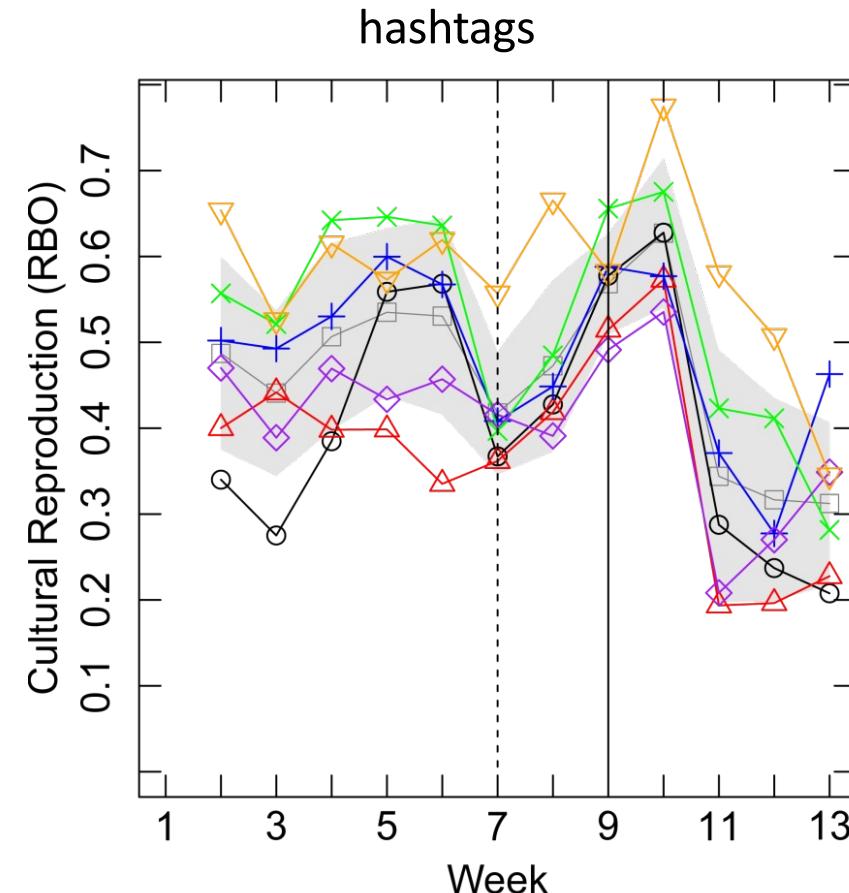


# Control through Semantic Reproduction

- How stable is a party's meaning vector over time?
- Characterizing changes between  $t$  and  $t-1$
- Measure: Rank Biased Overlap
- Ranges from 0 (no cultural stability) to 1 (maximum reproduction)

$$R_i(\sigma_i, p) = (1 - p) \sum_{d=1}^{\infty} \frac{2 \cdot \sigma_i^{1:d}(t1) \cap \sigma_i^{1:d}(t2)}{|\sigma_i^{1:d}(t1) + \sigma_i^{1:d}(t2)|} p^{(d-1)}$$

Fat-tail-weighted similarity of vectors at  $t1$  and  $t2$



# Implications

- Political parties compete for identity and control in a semantic space
- Stark differences between different conversational practices  
*(retweet, mention, hashtag)*
- Attempts at exercising control are pertubated by external events
- Initial measures for success of political communication

# This talk

measuring

- (i) sociological phenomena
  - Gender inequality on Wikipedia
- (ii) political phenomena
  - Emergence of a political movement in Germany
  - Political communication during the German national elections 2013

via networks and computational approaches.

# Social phenomena increasingly spill over to online networks

- Inequality
- Family
- Work
- Religion
- Migration
- Health
- Elections
- Politics
- Media
- Environment
- Personality
- Demographics

Networks as useful abstractions!  
Networks as useful models!

# Thank you.

*Vielen Dank!*

Markus Strohmaier