

# When the Wikipedians talk

Network and tree structure of Wikipedia discussion pages

David Laniado<sup>\*,†</sup>, Riccardo Tasso<sup>†</sup>, Yana Volkovich<sup>\*</sup> &  
Andreas Kaltenbrunner<sup>\*</sup>

<sup>\*</sup>Barcelona Media - Innovation Center  
Information, Technology & Society Research Group

<sup>†</sup>Politecnico di Milano  
Dipartimento di Elettronica e Informazione

ICWSM 2011, Barcelona, July 20<sup>th</sup>



- 1 Introduction
- 2 Data extraction
- 3 Social interaction networks
  - Basic network parameters
  - Directed assortativity
- 4 Discussion trees
  - Statistics about discussions
  - Differences between macro-categories
- 5 Conclusions

- 1 Introduction
- 2 Data extraction
- 3 Social interaction networks
  - Basic network parameters
  - Directed assortativity
- 4 Discussion trees
  - Statistics about discussions
  - Differences between macro-categories
- 5 Conclusions





WIKIPEDIA  
The Free Encyclopedia

- Main page
- Contents
- Featured content
- Current events
- Random article
- Donate to Wikipedia
- Interaction
  - Help
  - About Wikipedia
  - Community portal
  - Recent changes
  - Contact Wikipedia
- Toolbox
- Print/export
- Languages
  - Česky
  - Español
  - Français
  - Български
  - Kiswahili
  - Norsk (bokmål)
  - Русский
  - Svenska
  - ไทย

Article [Discussion](#)

[Read](#) [Edit](#) [View history](#)

[Log in / create account](#)

## Presidency of Barack Obama

From Wikipedia, the free encyclopedia

See also: *Timeline of the Presidency of Barack Obama*

The **Presidency of Barack Obama** began at noon EST on January 20, 2009 when he became the 44th President of the United States. Obama was a United States Senator from Illinois at the time of his victory over Arizona Senator John McCain in the 2008 presidential election. Obama became the first African-American president of the United States.

His policy decisions have addressed a global financial crisis and have included changes in tax policies, legislation to reform the United States health care industry, foreign policy initiatives and the phasing out of detention of prisoners at the Guantanamo Bay detention camp in Cuba. He attended the G-20 London summit and later visited U.S. troops in Iraq. On the tour of various European countries following the G-20 summit, he announced in Prague that he intended to negotiate substantial reduction in the world's nuclear arsenals, en-route to their eventual extinction. In October 2009, Obama was awarded the Nobel Peace Prize for "his extraordinary efforts to strengthen international diplomacy and cooperation between peoples."

### Contents [hide]

- Transition period
- Inauguration
- First 100 days
  - Expectations
  - Legislation and executive orders
- Approval ratings and opinion
  - 2009
    - Handling of the economy
    - 2010
  - Major legislation
    - Legislation signed
      - 2009
      - 2010
      - 2011
- Personnel
  - Cabinet appointees
  - Notable non-Cabinet positions
  - Judicial nominees
    - Supreme Court
    - Other courts
- Policies
  - Economy
  - Ethics
    - Lobbying reform
    - Transparency
  - Foreign policy
    - Guantanamo Bay detention camp
    - Overseas Contingency Operation
  - Gun control
  - Science and technology
    - Cybersecurity
    - Environment
    - NASA
    - Stem cell research

### Presidency of Barack Obama



**44th President of the United States**

**Incumbent**

**Assumed office**

January 20, 2009

**Vice President** Joe Biden

**Preceded by** George W. Bush

**Born** August 4, 1961 (age 49)  
Honolulu, Hawaii, United States<sup>11</sup>

**Birth name** Barack Hussein Obama II<sup>11</sup>

**Nationality** American

**Political party** Democratic

**Spouse(s)** Michelle Obama (m. 1992)

**Children** Malia Ann (b. 1998)  
Natasha (Sasha) (b. 2001)

**Residence** The White House

**Alma mater** Occidental College  
Columbia University (B.A.)



WIKIPEDIA  
The Free Encyclopedia

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

Interaction  
Help  
About Wikipedia  
Community portal  
Recent changes  
Contact Wikipedia

Toolbox  
Print/export

Log in / create account

Article Discussion

Read Edit New section View history

Search

## Talk:Presidency of Barack Obama

From Wikipedia, the free encyclopedia

[Skip to table of contents](#)

This is the **talk page** for discussing improvements to the **Presidency of Barack Obama** article.

- This is not a forum for general discussion of the article's subject.**
- Put new text under old text.** [Click here to start a new topic.](#)
- Please sign and date your posts** by typing four tildes (~~~~).
- New to Wikipedia?** Welcome! Ask questions, get answers.

- Be polite
- Assume good faith
- Avoid personal attacks
- Be welcoming

#### Article policies

- No original research
- Neutral point of view
- Verifiability

[Archives](#): 1, 2, 3, 4



**This article has been placed on article probation.** Editors making disruptive edits may be **blocked** temporarily from editing the encyclopedia, or subject to other administrative remedies, according to standards that may be higher than elsewhere on Wikipedia. Please see [Talk:Barack Obama/Article probation](#) for full information and to review the decision.

**Administrators:** when sanctioning an editor for disruption to an article under probation, please be sure to record the action in the appropriate log. The log is linked [here](#), under "decision and log" on the sanction's row in the table.



This article is within the scope of **WikiProject Barack Obama**, a collaborative effort to improve the coverage of Barack Obama on Wikipedia. If you would like to participate, please visit the project page, where you can join the [discussion](#) and see a list of open tasks.



[Barack Obama portal](#)

**B**

This article has been rated as **B-Class** on the [quality scale](#).

**Top**

This article has been rated as **Top-importance** on the [importance scale](#).

#### Contents [hide]

- 2010 election results
- WikiLeaks release of secret documents
- 2010 elections
- Filibuster, etc.
- New source - Terrible, Horrible, etc.
- Job approval ratings

## 2010 election results

[\[edit\]](#)

I feel we should add something on here about the Republican victories in the 2010 elections and how that will be a turning point for Obama similar to the Presidency of Bill Clinton article. Where should we put that? [Politics2012 \(talk\)](#) 03:39, 09 November 2010 (UTC)

How will it be a "turning point"? Do you have a *crystal ball* or something? -- [Scjessey \(talk\)](#) 21:53, 9 November 2010 (UTC)

I meant that we should put somewhere in here something about Republican victories in general. It says something in the Bill Clinton Presidency article about the 94 Republican victories and in the George W. Bush presidency article about the Democratic victories in the 2006 elections. We need to put something in here about the 2010 Republican victories. But where? [Politics2012 \(talk\)](#) 03:39, 09 November 2010 (UTC)

I disagree. The information in those other articles has been written with the benefit of an historical perspective, but we do not have that luxury here. We must wait and see what sources say about how Republican victories have significantly impacted the Obama presidency, if and when they appear in the future. -- [Scjessey \(talk\)](#) 17:26, 10 November 2010 (UTC)

[Archives](#)  
1, 2, 3, 4

# Introduction

## The hidden side of Wikipedia

- Since 2007, growth of Wikipedia has notably slowed down [B. Suh et al.; "The singularity is not near: slowing growth of Wikipedia;" in *Proc. of WikiSym '09*]
- The *hidden side* of Wikipedia is gaining importance
  - article talk pages → explicit coordination and discussion
  - user talk pages → personal communications (sort of *public inbox*)
- Article "Barack Obama":
  - discussion split into 72 pages
  - 22 000 comments in the article talk pages
  - (17 500 edits done to the article)

## Goal

Detect patterns of interaction in the communications between the Wikipedians

- Unlike in other online discussion spaces, in Wikipedia the users discuss to reach consensus and to coordinate their activity with each other

## Contribution

- extract the discussions' structure
- analyze interaction networks
- study discussion trees

- 1 Introduction
- 2 Data extraction
- 3 Social interaction networks
  - Basic network parameters
  - Directed assortativity
- 4 Discussion trees
  - Statistics about discussions
  - Differences between macro-categories
- 5 Conclusions



I think the bot is archiving this talk page too soon and too frequently.

[edit]

What do you think? [Grundle2600 \(talk\)](#) 22:53, 23 June 2009 (UTC)

I think this page is very short, so no harm in keeping the discussions a bit longer. As an incremental step I'll increased the archive time from 7 days to 10 days, per your comment. [Wikidemon \(talk\)](#) 01:19, 24 June 2009 (UTC)

Thank you. I didn't even know you could change it like that! [Grundle2600 \(talk\)](#) 02:40, 24 June 2009 (UTC)

### Coup d'etat in Honduras

[edit]

Why does this article not mention Obama's first coup d'etat. Just go to <http://www.globalresearch.ca> to find out more information about this current event. —Preceding [unsigned](#) comment added by [99.255.173.155 \(talk\)](#) 03:26, 5 July 2009 (UTC)

Because there needs to be better sourcing for something as controversial as this. [QueenofBattle \(talk\)](#) 19:39, 5 July 2009 (UTC)

[Better sourcing](#) [Grundle2600 \(talk\)](#) 01:22, 8 July 2009 (UTC)

Typing in Obama's name, constitution, and the guys name does not mean better sourcing. Simply put, there has to be a very reliable source that says exactly that this was Obama's first coup d'etat. There can be no [WP:OR](#) or [WP:SYNTH](#).

[Brothejr \(talk\)](#) 01:30, 8 July 2009 (UTC)

OK. That's a good point. [ABC News](#) says that Honduras' President Manuel Zelaya tried to give himself another term, despite the fact that Honduras' constitution prohibits such a thing, and that Obama said he supports Zelaya in this action. [Grundle2600 \(talk\)](#) 01:37, 8 July 2009 (UTC)

And of course it says much more than that - you leave out a great deal of context, not to mention leaving to the side more in-depth coverage in any number of other sources. Also you should explain exactly how you envision the [Honduran coup](#) being mentioned in this article. Personally I think it's rather important in a global history sense, but I'm not sure it deserves mention in the article on the [Presidency of Barack Obama](#). He has talked about it, of course, but then again so has most every major leader and foreign policy NGO in the Western Hemisphere.

—[Bigtimepeace](#) | [talk](#) | [contribs](#) 07:35, 8 July 2009 (UTC)

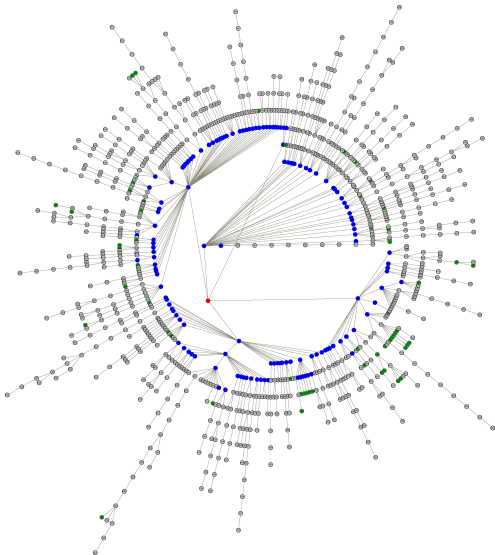
I was just helping out with sources. I don't necessarily think that it should, or shouldn't, be cited in this article. Perhaps it could be cited in the article about the event itself, instead. [Grundle2600 \(talk\)](#) 08:54, 8 July 2009 (UTC)

[2009 Honduran coup d'état](#) already says, "President Barack Obama of the United States said "We believe that the coup was not legal and that President Zelaya remains the President of Honduras." [55][157]" I think that's enough, and it doesn't necessarily have to be added to this article. It could be cited in this article to, but I don't give it high priority in this article. [Grundle2600 \(talk\)](#) 09:04, 8 July 2009 (UTC)



# Data extraction

Discussion tree for article "Presidency of Barack Obama"



- red → root (the article)
- blue → structural nodes
- green → anonymous comments
- grey → registered comments

# Dataset

## Basic quantities of the data analysed

Dump of the English Wikipedia, dated March 12th, 2010.

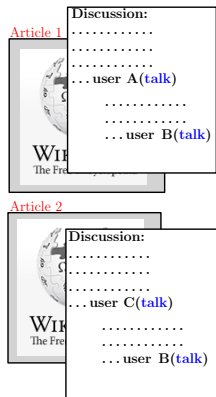
#articles	3 210 039	
#edits of article pages	402 851 686	
#articles with talk page (ATP)	871 485	(27.1%)
#comments in ATP	9 421 976	
#users who comment articles	350 958	(2.8%)
#registered users	12 651 636	
#user talk pages (UTP)	1 662 818	(13.1%)
#comments in UTP	13 670 980	

- 1 Introduction
- 2 Data extraction
- 3 Social interaction networks**
  - Basic network parameters
  - Directed assortativity
- 4 Discussion trees
  - Statistics about discussions
  - Differences between macro-categories
- 5 Conclusions

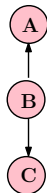
# Social interaction networks

## Network construction

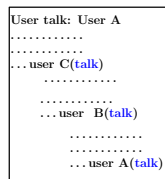
- **article reply network** → direct replies in articles discussion pages.



article reply network



- **user reply network** → direct replies in user talk pages.
- **wall network** → personal messages posted on another user's talk page.



wall network

usertalk network



# Basic network parameters

Global measures of the Wikipedia discussion and talk network

variable	reply-NW	talk-NW	wall-NW
#nodes with edges	204 017	114 258	1 861 702
w. in-degree $\geq 1$	121 682	103 147	1 832 168
w. out-degree $\geq 1$	182 881	63 334	177 331
#edges	1 489 734	852 065	4 412 212
mean distance	4.10	3.86	4.06
maximal distance	15	11	12
mean in/out-degree	7.30	7.46	2.37
clustering coeff.	0.083	0.053	0.035
reciprocity	0.44	0.45	0.15



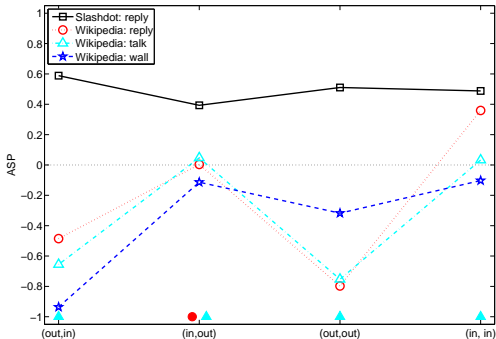
# Directed assortativity

## Approach

- Assortativity: measure of *diversity* in networks
  - do nodes having many connections preferentially interact with one another or with poorly connected nodes?
- Directed assortativity: correlation between *in-* and *out-* degree of *source* and *target* nodes [J. Foster et al.; “Edge direction and the structure of networks ”; in *PNAS*, 2010]
- Results depend on the network structure
  - To assess statistical significance, contrast the results with an ensemble of  $N=100$  randomized networks having the same degree sequence
- Normalization of the four assortativity measures for a network (Assortativity Significance Profile)

# Directed assortativity profiles

## Comparison of the directed Assortativity Significance Profile



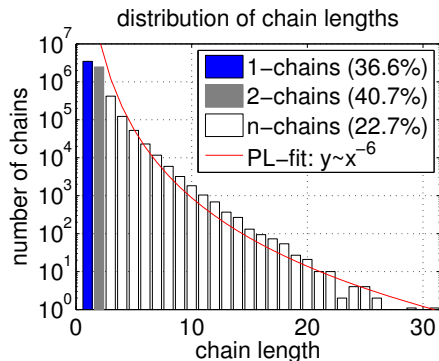
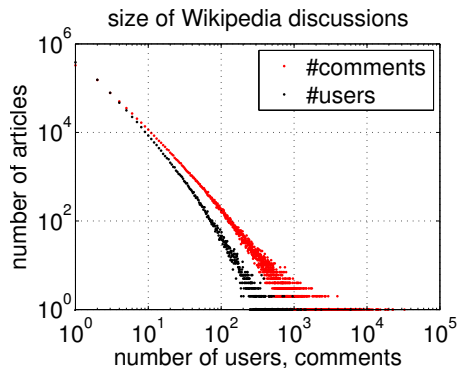
Where ASP score is not significant ( $|Z| < 2$ ), it is marked with the corresponding symbol at the figure bottoms.



- 1 Introduction
- 2 Data extraction
- 3 Social interaction networks
  - Basic network parameters
  - Directed assortativity
- 4 Discussion trees
  - Statistics about discussions
  - Differences between macro-categories
- 5 Conclusions

# Statistics about discussions

Distributions of users and comments by article discussion

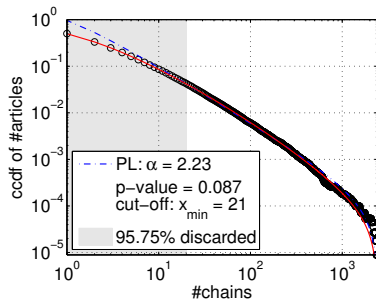
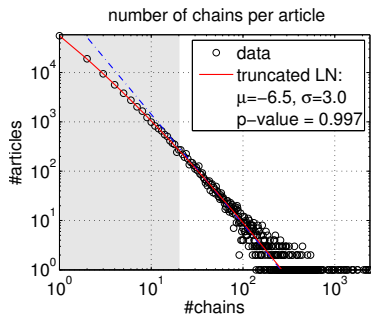


- 85% of articles have  $\leq 10$  comments
- 15 000 articles have  $\geq 100$  comments

# Discussion trees

## Size of the discussions

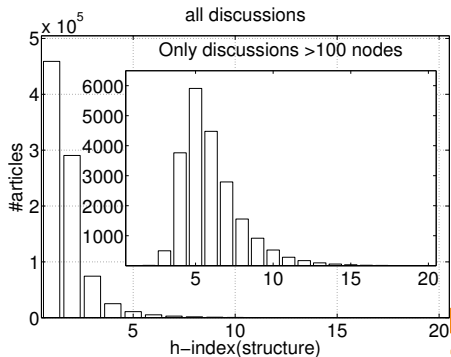
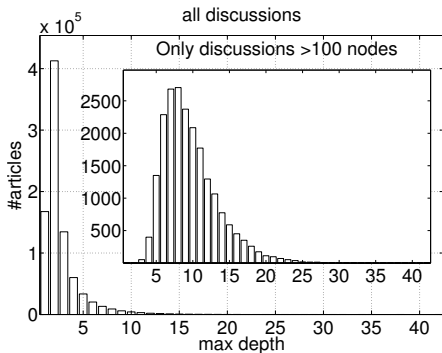
- Number of users involved
- Number of *chains* of length  $\geq 3$ , or consecutive replies between two users
  - example chain of length 3:  $A \leftarrow B \leftarrow A$
  - good indicator of *conflictive* discussions



# Discussion trees

## Length of the discussions

- Maximum *depth*
  - sensitive to the presence of individual threads
- *h-index* of the discussion tree
  - maximum number  $h$  such that there are:  
 $\geq h$  comments at depth  $\geq h$



# Most discussed Wikipedia articles

Top 20 articles ordered by number of chains in the discussion

#	Title	chains	comments	users	h-index	max. depth	edits
1	Intelligent design	2413	22454 (3)	954 (13)	16 (20)	20 (358)	9179 (53)
2	Gaza War	2358	17961 (6)	607 (47)	19 (2)	27 (28)	11499 (29)
3	Barack Obama	2301	22756 (2)	2360 (2)	18 (6)	21 (245)	17453 (6)
4	Sarah Palin	2182	19634 (4)	1221 (9)	17 (10)	25 (56)	12093 (24)
5	Global warming	2178	19138 (5)	1382 (5)	17 (10)	20 (358)	14074 (15)
6	Main Page	2065	32664 (1)	5969 (1)	15 (34)	22 (169)	4003 (674)
7	Chiropractic	1772	13684(13)	243 (389)	18 (6)	29 (17)	6190 (204)
8	Race and intelligence	1764	13790(12)	410 (126)	17 (10)	24 (74)	7615 (100)
9	Anarchism	1589	14385 (9)	496 (76)	20 (1)	28 (22)	12589 (19)
10	British Isles	1556	12044(16)	576 (56)	17 (10)	23 (113)	4047 (658)
11	CRU <sup>1</sup> hacking incident	1551	11536(17)	474 (88)	17 (10)	20 (358)	2346 (2364)
12	Jesus	1397	17916 (7)	1239 (7)	13(119)	16 (1383)	17081 (7)
13	Circumcision	1356	10469(21)	436 (113)	17 (10)	26 (42)	7354 (117)
14	Homeopathy	1323	13509(14)	516 (68)	17 (10)	25 (56)	6902 (151)
15	George W. Bush	1281	15257 (8)	1969 (3)	14 (65)	18 (676)	32314 (1)
16	September 11 attacks	1250	13830(11)	1244 (6)	16 (20)	26 (42)	11086 (30)
17	Evolution	1165	13404(15)	942 (16)	13(119)	23 (113)	9780 (44)
18	Catholic Church	1162	14104(10)	620 (43)	15 (34)	18 (676)	14082 (14)
19	Cold fusion	1098	8354 (29)	359 (174)	15 (34)	20 (358)	4320 (557)
20	2008 South Ossetia war	1075	10596(20)	853 (20)	17 (10)	23 (113)	9930 (43)

In parenthesis: rank according to the corresponding variable

<sup>1</sup> Climatic Research Unit

# Assigning articles to macro-categories

Approach by [Kittur et al.; "What's in Wikipedia?: mapping topics and conflict using socially annotated category structure." In *Proc. of CHI '09*]

- 21 macro-categories from Wikipedia
  - ~ direct sub-categories of *Category:Main\_topic\_classifications*
- assign each category to the closest macro-category

for each article:

- consider all categories to which it is directly assigned
- combine macro-categories corresponding to these categories to assign a percentage
- improved effectiveness by adding a weight to penalise links in the "wrong direction" in the category graph.

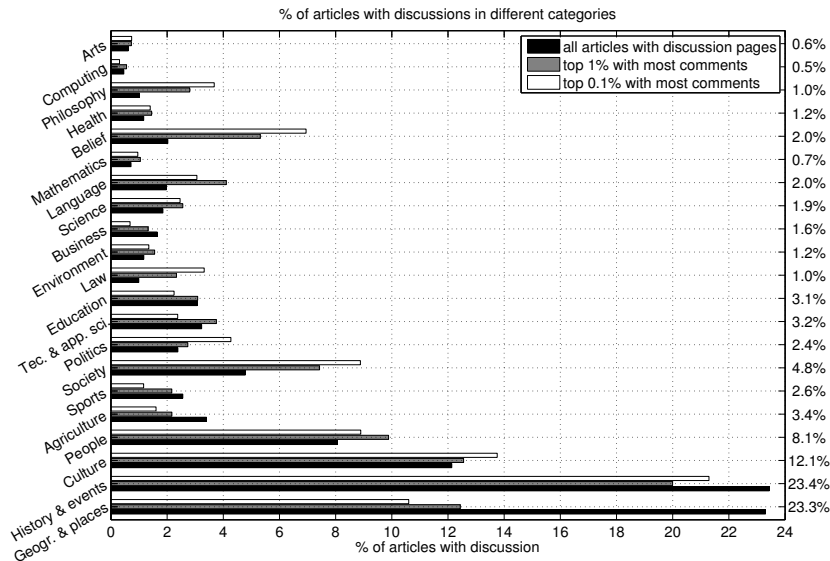
# Assigning articles to macro-categories

## Example

### Example:

- Article "Barcelona" belongs to 3 categories:
  - Geography of Catalonia → Geography
  - Phoenician colonies in Spain → History
  - 10s BC establishments → History
- Barcelona → History (66.7%), Geography (33.3%)

# Article discussions categorisation



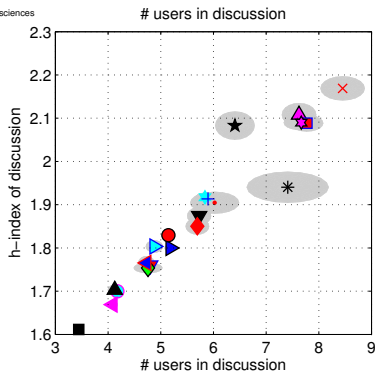
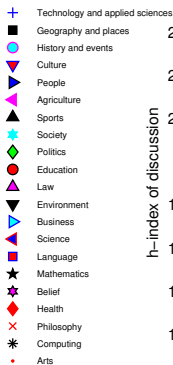
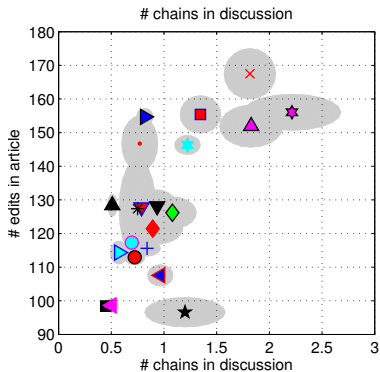


# Article discussions by semantic area

Differences between Categories

# of chains vs. # of edits

# of users vs. h-index



Gray areas → 95% confidence interval with bootstrap test

- 1 Introduction
- 2 Data extraction
- 3 Social interaction networks
  - Basic network parameters
  - Directed assortativity
- 4 Discussion trees
  - Statistics about discussions
  - Differences between macro-categories
- 5 Conclusions



# Conclusions

- Considerable effort for the data extraction
  - the parser will be released as free software
- Patterns in the interaction among the Wikipedians
  - strong relationship between the core and the periphery of the community
  - the existence of different roles could be further investigated
- Measures to characterize the discussions
  - h-index allows to study the depth of the discussion trees
  - reply chains help to detect conflictive topics
- Evidence of structural differences among discussions from different macro-categories

page discussion create + protect watch refresh

## Editing Talk:When the Wikipedians Talk: Network and Tree Structure of Wikipedia Discussion Pages

You have followed a link to a page that does not exist yet.

To create the page, start typing in the box below (see the [help page](#) for more info).

If you are here by mistake, click your browser's **back** button.



-- Questions? --

\* |

Please note that all contributions to AIRWiki are considered to be released under the Attribution-Noncommercial-Share Alike 3.0 Unported (see [AIRWiki:Copyrights](#) for details). If you do not want your writing to be edited mercilessly and redistributed at will, then do not submit it here.



You are also promising us that you wrote this yourself, or copied it from a public domain or similar free resource. **DO NOT SUBMIT COPYRIGHTED WORK WITHOUT PERMISSION!**

Summary:

This is a minor edit  Watch this page

[Cancel](#) | [Editing help](#) (opens in new window)

# References I

-  Crandall, D. J., Cosley, D., Huttenlocher, D. P., Kleinberg, J. M., and Suri, S. (2008). Feedback effects between similarity and social influence in online communities. *In Proc. of KDD '08.*
-  Foster, J. G., Foster, D. V., Grassberger, P., and Paczuski, M. (2010). Edge direction and the structure of networks. *PNAS*, 107(24):10815–10820.
-  Gómez, V., Kappen, H. J., and Kaltenbrunner, A. (2011). Modeling the structure and evolution of discussion cascades. *In Proc. of Hypertext 2011.*
-  Kittur, A., Chi, E. H., and Suh, B. (2009). What's in wikipedia?: mapping topics and conflict using socially annotated category structure. *In Proc. of CHI '09*, pages 1509–1512, New York, NY, USA. ACM.
-  Laniado, D. and Tasso, R. (2011). Co-authorship 2.0: Patterns of collaboration in Wikipedia. *In Proc. of Hypertext 2011.*
-  Szell, M., Lambiotte, R., and Thurner, S. (2010). Multirelational organization of large-scale social networks in an online world. *PNAS*, 107(31):13636–13641.