

UNDERSTANDING WIKIPEDIA

social collaborative knowledge building



niki kittur
nkittur@cs.cmu.edu

social
computing
lab @ Carnegie Mellon

Large scale social collaboration

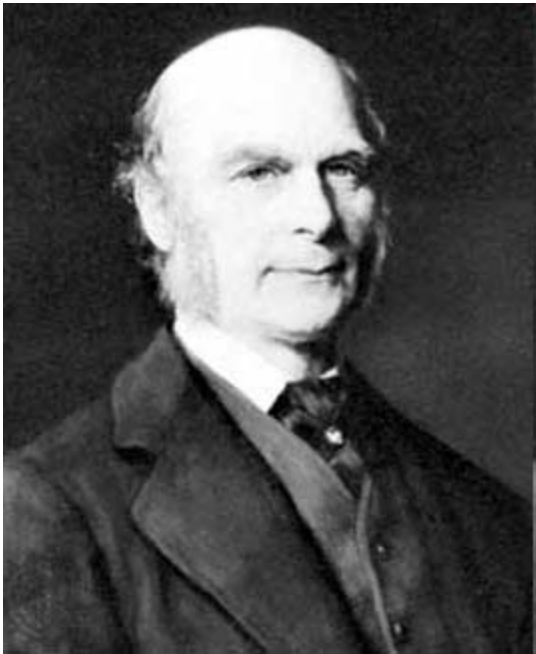
Advantages

- Solve problems too large for individual cognition
- Work of individuals benefit group
- Aggregating decisions -> better outcomes



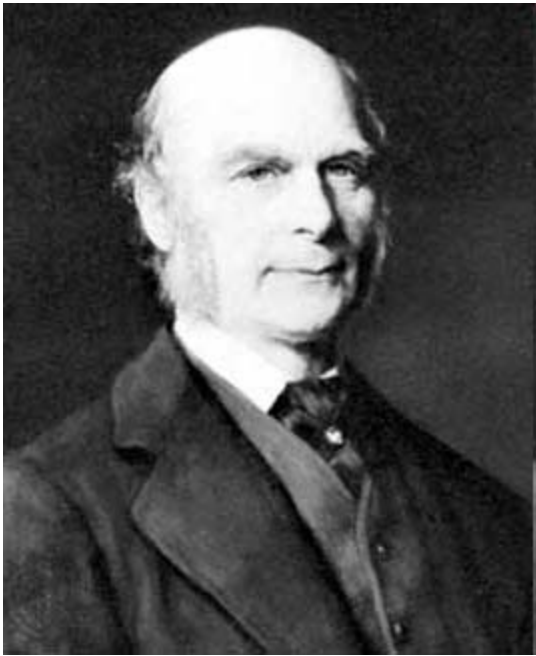
(Benkler, 2002; Golder & Huberman, 2006; Grudin, 1994; Raymond, 1999)

History



Sir Francis Galton

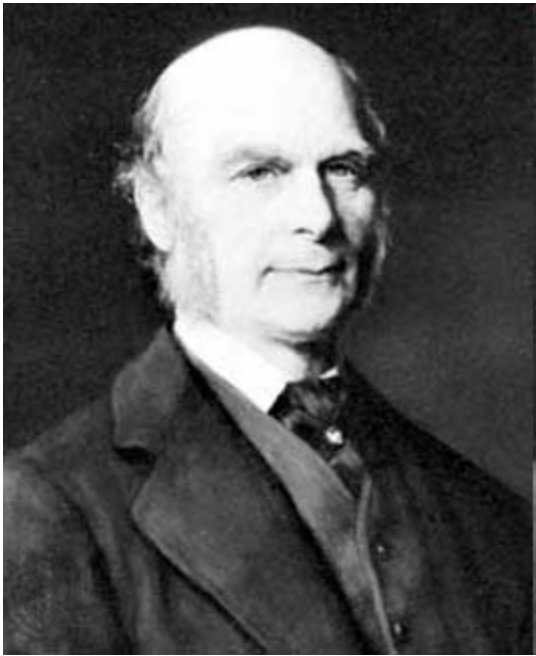
History



Sir Francis Galton



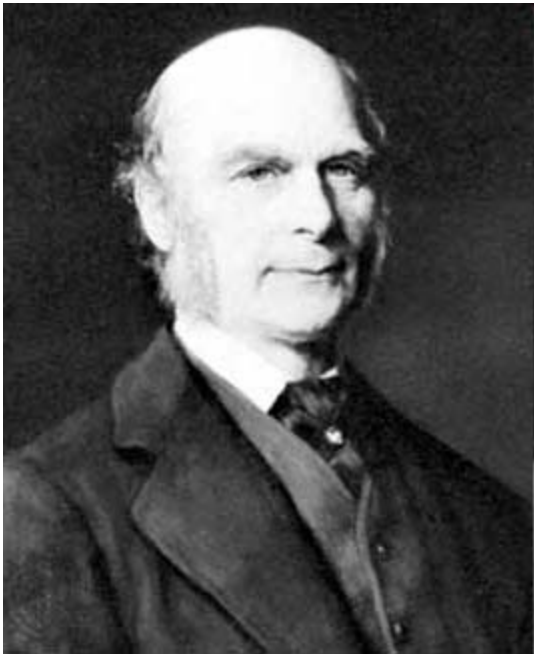
History



Sir Francis Galton



History



Sir Francis Galton

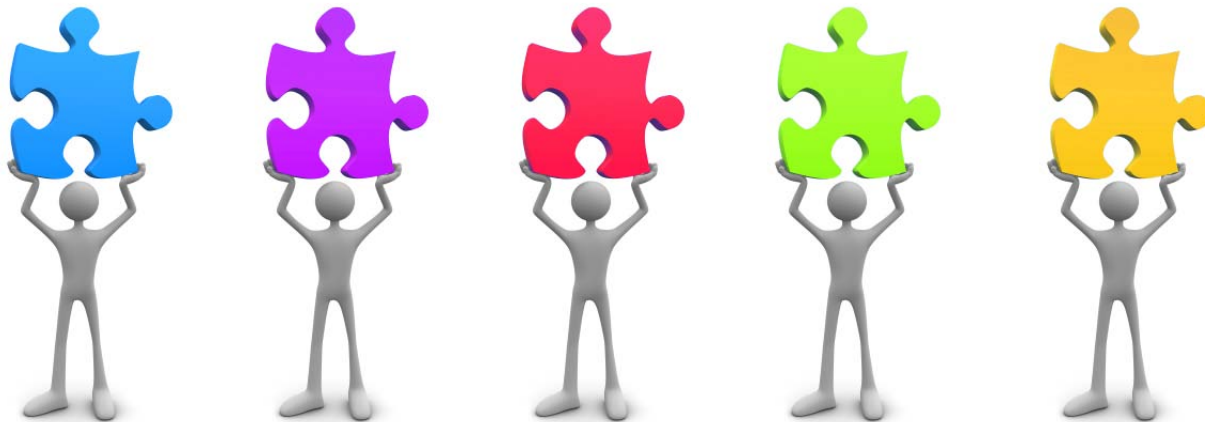


Online collective intelligence

- Predicting: Iowa Electronic Market
- Filtering: Digg, Reddit
- Organizing: del.icio.us
- Recommending: netflix, amazon product reviews

Common assumptions

1. Independent judgments
2. Automatic aggregation



Complex information processing

- Independent judgments and automatic aggregation are not enough
 - Scientists collaborating on a new discovery
 - Detectives cooperating to track serial killer
 - Volunteers writing encyclopedia
- Need to coordinate, build consensus
- Coordination is the norm, not the exception



Research question

How do we harness the power of the crowd for complex tasks that involve coordination?

Wiki

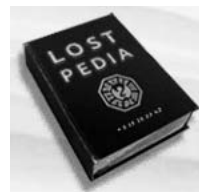
- Started by Ward Cunningham in 1994
- Enables anyone to edit or create pages using web browser
- Saves history of all edits
- Easy to revert an edit to an older version



MediaWiki syntax	Equivalent HTML
"Take some more <code>[[tea]]</code> ," the March Hare said to Alice, very earnestly.	<code><p>"Take some more tea ," the March Hare said to Alice, very earnestly.</p></code>
"I've had nothing yet," Alice replied in an offended tone: "so I can't take more."	<code><p>"I've had nothing yet," Alice replied in an offended tone: "so I can't take more."</p></code>
"You mean you can't take ' <code>less</code> '," said the Hatter: "it's very easy to take ' <code>more</code> ' than nothing."	<code><p>"You mean you can't take <i>less</i>," said the Hatter: "it's very easy to take <i>more</i> than nothing."</p></code>

Wikis applied to many domains

- Corporate memory (Socialtext)
- Intelligence gathering (Intellipedia)
- Scholarly research (Scholarpedia)
- Writing army doctrine
- Collecting knowledge about *anything*



WIKIPEDIA

English

The Free Encyclopedia

2 683 000+ articles

日本語

フリー百科事典

550 000+ 記事

Deutsch

Die freie Enzyklopädie

847 000+ Artikel

Français

L'encyclopédie libre

746 000+ articles

Polski

Wolna encyklopedia

566 000+ haseł

Русский

Свободная энциклопедия

344 000+ статей

Español

La enciclopedia libre

431 000+ artículos

Italiano

L'enciclopedia libera

527 000+ voci

Português

A enciclopédia livre

449 000+ artigos

Nederlands

De vrije encyclopedie

508 000+ artikelen



Why study Wikipedia?

- May have thousands of individuals involved in a single sensemaking task
 - Integrating many conflicting sources into an article
- Many tasks require high coordination
 - Planning an article
 - Building consensus on what should be included
 - Organizing and structuring
 - Resolving conflicts
 - Achieving neutral point of view
- Full history available (200+ million edits, 2.5+TB)

History

- Founded in Jan 15, 2001 as a feeder to Nupedia, an expert-reviewed free encyclopedia
- Co-founders Jimmy Wales & Larry Sanger
- Heavily subsidized by bomis.com (1996), Wales' search portal, ad and porn company



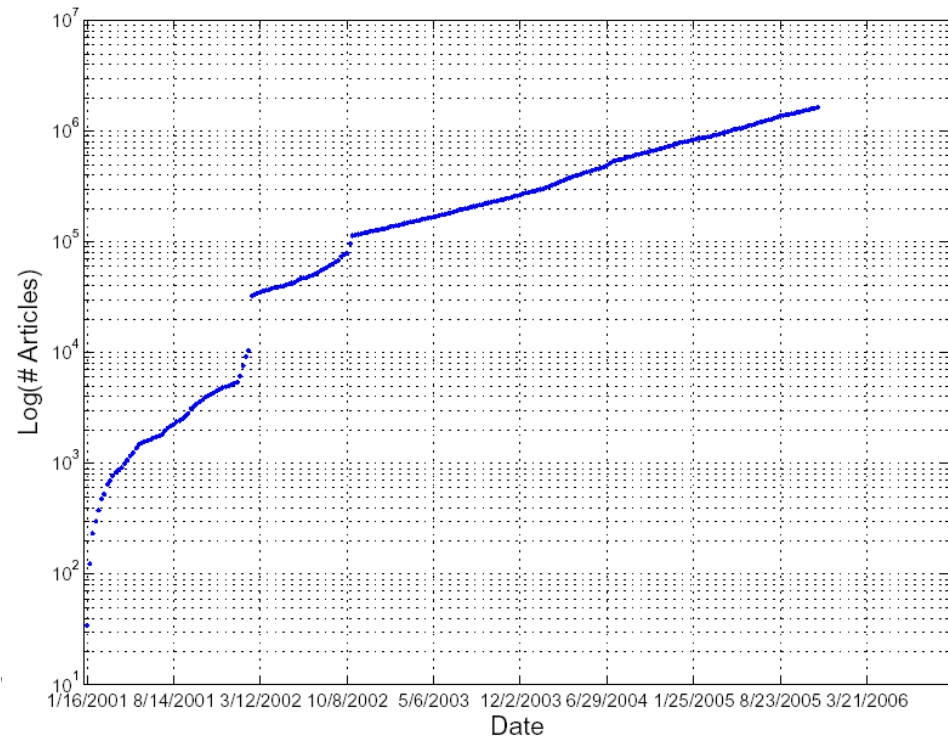
Jimmy Wales



Larry Sanger

Growth

- As of 7/2007
- English version
 - ~2.5M articles in English language Wikipedia
 - 151,934 unique editors
 - Doubling number of articles every 40 weeks
 - Average 435 words/article (vs. 650 in Britannica)
- All versions
 - 250 languages
 - 11 language versions > 100K article
 - 282,875 unique editors
- Spin-offs
 - Wiktionary, wikibooks, wikiversity, wikinews, wikispecies, wikiquote, wikisource



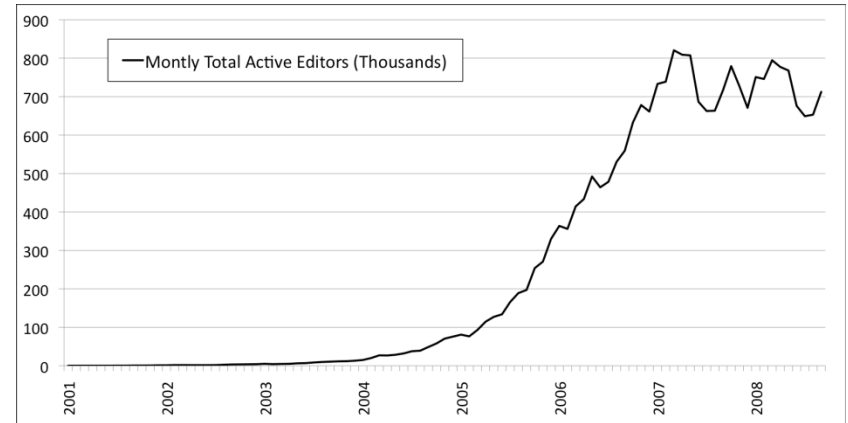
Growth in articles, English language Wikipedia (Zhang, 2007)

Slowing growth

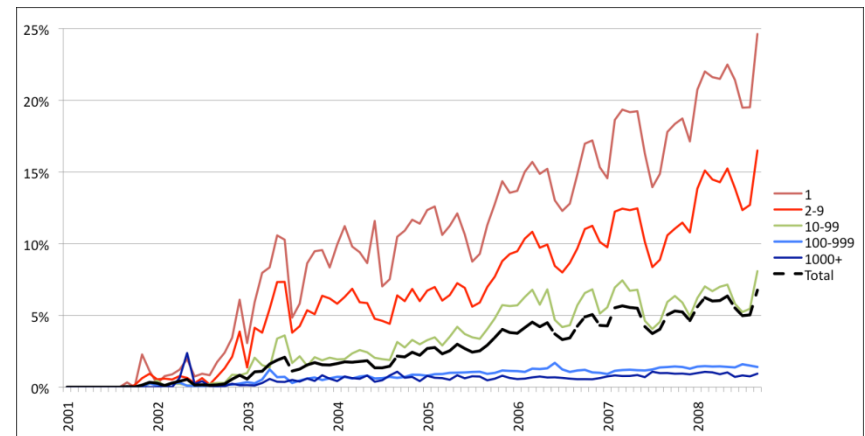
- Since 2007, slowing growth

Why?

- Fewer new topics to write about
- Growing resistance to new contributions
- Greater coordination overhead



Number of active editors per month



Proportion reverted edits (by editor class)

Wisdom of crowds poll

What proportion of Wikipedia (in words) is made up of articles?

0-25% | 25-50% | 50-75% | 75-100%

Wisdom of crowds poll

Article



WIKIPEDIA
The Free Encyclopedia

navigation

- [Main page](#)
- [Contents](#)
- [Featured content](#)
- [Current events](#)
- [Random article](#)

search

[Go](#) [Search](#)

interaction

- [About Wikipedia](#)
- [Community portal](#)
- [Recent changes](#)
- [Contact Wikipedia](#)
- [Donate to Wikipedia](#)
- [Help](#)

toolbox

- [What links here](#)
- [Related changes](#)
- [Upload file](#)

[article](#) [discussion](#) [edit this page](#) [history](#)

 [Log in / create account](#)

Music of Italy

From Wikipedia, the free encyclopedia

See also: *Music history of Italy*

The **music of Italy** ranges across a broad spectrum of [opera](#) and instrumental [classical music](#), the traditional styles of the country's different regions, and a body of popular music drawn from both native and imported sources. [Music](#) has traditionally been one of the cultural markers of Italian national and ethnic identity and holds an important position in society and in [politics](#). Italian innovation in [musical scales](#), [harmony](#), [notation](#), and [theatre](#) enabled the development of [opera](#) in the late 16th century, and much of modern [European classical music](#), such as the [symphony](#) and [concerto](#).

Instrumental and vocal classical music is an iconic part of Italian identity, spanning experimental art music and international fusions to symphonic music and opera. Opera is integral to Italian musical culture, and has become a major segment of [popular music](#). The Neapolitan song, *canzone Napoletana*, and the *cantautori* singer-songwriter traditions are also popular domestic styles that form an important part of the Italian music industry, alongside imported genres like [jazz](#), [rock](#) and [hip hop](#). Italian folk music is an important part of the country's musical heritage, and spans a diverse array of regional styles, instruments and dances.

Contents [\[hide\]](#)

1 Characteristics

1.1 Social identity

Music of Italy	
Genres:	<div>Classical: Opera</div> <div>Pop: Rock (Hardcore) - Hip hop - Folk - jazz - Progressive rock</div>
History and Timeline	
Awards	Italian Music Awards
Charts	Federation of the Italian Music Industry
Festivals	<div>Sanremo Festival - Umbria Jazz Festival - Ravello Festival - Festival dei Due Mondi - Festivalbar</div>
Media	Music media in Italy
National anthem	Il Canto degli Italiani
Regional scenes	

Discussion



WIKIPEDIA
The Free Encyclopedia

navigation

- [Main page](#)
- [Contents](#)
- [Featured content](#)
- [Current events](#)
- [Random article](#)

search

interaction

- [About Wikipedia](#)
- [Community portal](#)
- [Recent changes](#)
- [Contact Wikipedia](#)
- [Donate to Wikipedia](#)
- [Help](#)

toolbox

- [What links here](#)
- [Related changes](#)
- [Upload file](#)

[article](#) [discussion](#) [edit this page](#) [new section](#) [history](#)

 [Log in / create account](#)

Talk:Music of Italy

From Wikipedia, the free encyclopedia



Music of Italy has been listed as one of the *Arts good articles* under the [good article criteria](#). If you can improve it further, [please do so](#). *If it no longer meets these criteria, you can delist it, or ask for a [reassessment](#).*

Article milestones [show]

This article has been reviewed by the **Version 1.0 Editorial Team**.



This article has been selected for **Version 0.7** and subsequent [release versions](#) of Wikipedia.

Additional information: [show]



Music of Italy is within the scope of the **WikiProject Regional and national music**, an attempt at building a resource on the music of all the peoples and places of the world. Please visit the project's [listing](#) to see the [article's assessment](#) and to help us improve the article as we [push to 1.0](#).



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



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



Music of Italy is within the scope of **WikiProject Italy**, an attempt to build a comprehensive and detailed guide to articles on Italy on Wikipedia. If you would like to participate, you can edit the article attached to this page, or visit the [project](#)

Discussion

and see a list of open tasks.



High

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

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

This article has been rated but has no comments. If appropriate, please review the article and leave comments [here](#) to identify the strengths and weaknesses of the article and how best to improve it.

Contents [\[hide\]](#)

- [1 regions](#)
- [2 sound files](#)
- [3 Steps forward](#)
- [4 citation needed?](#)
- [5 small change to lead](#)
- [6 another change to lead?](#)
- [7 imported styles](#)
- [8 Vasco Rossi](#)
- [9 GA Pass](#)
- [10 Fair use rationale for Image:Sanremofest.jpg](#)
- [11 Music history of Italy](#)



Archives

- [1. December 2005 – August 2006](#)

regions

[\[edit\]](#)

I adjusted that section by shifting a few sentences so that we get a smoother flow from lead to northern to southern.
[Jeffmatt](#) 05:35, 24 August 2006 (UTC)

Looks good - I'm also fine with adding a paragraph on film music, though I'm not entirely sure I agree with the placement. I think I'll make some small edits tonight and hopefully make a couple sound samples (may not have time to upload them, but I'll make them). [Tuf-Kat](#) 00:20, 25 August 2006 (UTC)

Edit history

- (cur) (prev) ○ 06:46, 31 May 2004 TUF-KAT (Talk | contribs) **m** (*Category:World music*) (undo)
- (cur) (prev) ○ 19:10, 23 April 2004 TUF-KAT (Talk | contribs) **m** (*Reverted edits by 207.193.136.3 to last version by TUF-KAT*) (undo)
- (cur) (prev) ○ 18:51, 23 April 2004 207.193.136.3 (Talk) (undo)
- (cur) (prev) ○ 18:09, 12 March 2004 TUF-KAT (Talk | contribs) (undo)
- (cur) (prev) ○ 16:53, 11 March 2004 Richard cocks (Talk | contribs) **m** (*seperate -> separate*) (undo)
- (cur) (prev) ○ 05:14, 5 March 2004 TUF-KAT (Talk | contribs) (*chanting*) (undo)
- (cur) (prev) ○ 06:38, 4 March 2004 Jiang (Talk | contribs) **m** (*Reverted edits by 24.69.255.203 to last version by TUF-KAT*) (undo)
- (cur) (prev) ○ 02:00, 26 February 2004 24.69.255.203 (Talk) (undo)
- (cur) (prev) ○ 07:23, 17 February 2004 TUF-KAT (Talk | contribs) (*samples*) (undo)
- (cur) (prev) ○ 02:25, 27 January 2004 TUF-KAT (Talk | contribs) (undo)
- (cur) (prev) ○ 01:53, 27 January 2004 TUF-KAT (Talk | contribs) (*some expanding*) (undo)
- (cur) (prev) ○ 02:40, 13 January 2004 Snobot (Talk | contribs) **m** (*snobot - Robot-assisted disambiguation: Dante*) (undo)
- (cur) (prev) ○ 15:01, 4 January 2004 81.98.77.229 (Talk) (undo)
- (cur) (prev) ○ 08:58, 12 December 2003 Uriber (Talk | contribs) **m** (*Palestinian*) (undo)
- (cur) (prev) ○ 02:14, 2 December 2003 TUF-KAT (Talk | contribs) **m** (undo)
- (cur) (prev) ○ 02:05, 2 December 2003 TUF-KAT (Talk | contribs) (*expanding*) (undo)
- (cur) (prev) ○ 22:17, 30 November 2003 TUF-KAT (Talk | contribs) **m** (undo)
- (cur) (prev) ○ 08:42, 30 November 2003 TUF-KAT (Talk | contribs)

Compare selected versions

(Latest | Earliest) View (newer 50) (older 50) (20 | 50 | 100 | 250 | 500)

Edit history

- (cur) (prev) 06:46, 31 May 2004 TUF-KAT (Talk | contribs) **m** (*Category:World music*) (undo)
- (cur) (prev) 19:10, 23 April 2004 TUF-KAT (Talk | contribs) **m** (*Reverted edits by 207.193.136.3 to last version by TUF-KAT*) (undo)
- (cur) (prev) 18:51, 23 April 2004 207.193.136.3 (Talk) (undo)
- (cur) (prev) 18:09, 12 March 2004 TUF-KAT (Talk | contribs) (undo)
- (cur) (prev) 16:53, 11 March 2004 Richard cocks (Talk | contribs) **m** (*seperate -> separate*) (undo)
- (cur) (prev) 05:14, 5 March 2004 TUF-KAT (Talk | contribs) (*chanting*) (undo)
- (cur) (prev) 06:38, 4 March 2004 Jiang (Talk | contribs) **m** (*Reverted edits by 24.69.255.203 to last version by TUF-KAT*) (undo)
- (cur) (prev) 02:00, 26 February 2004 24.69.255.203 (Talk) (undo)
- (cur) (prev) 07:23, 17 February 2004 TUF-KAT (Talk | contribs) (*samples*) (undo)
- (cur) (prev) 02:25, 27 January 2004 TUF-KAT (Talk | contribs) (undo)
- (cur) (prev) 01:53, 27 January 2004 TUF-KAT (Talk | contribs) (*some expanding*) (undo)
- (cur) (prev) 02:40, 13 January 2004 Snobot (Talk | contribs) **m** (*snobot - Robot-assisted disambiguation: Dante*) (undo)
- (cur) (prev) 15:01, 4 January 2004 81.98.77.229 (Talk) (undo)
- (cur) (prev) 08:58, 12 December 2003 Uriber (Talk | contribs) **m** (*Palestinian*) (undo)
- (cur) (prev) 02:14, 2 December 2003 TUF-KAT (Talk | contribs) **m** (undo)
- (cur) (prev) 02:05, 2 December 2003 TUF-KAT (Talk | contribs) (*expanding*) (undo)
- (cur) (prev) 22:17, 30 November 2003 TUF-KAT (Talk | contribs) **m** (undo)
- (cur) (prev) 08:42, 30 November 2003 TUF-KAT (Talk | contribs)

Compare selected versions

(Latest | Earliest) View (newer 50) (older 50) (20 | 50 | 100 | 250 | 500)

Policies + Procedures

Wikipedia:List of policies and guidelines

From Wikipedia, the free encyclopedia

This page lists Wikipedia **policies and guidelines**. Policies are shown in bold type. For lists of key pages with summaries of their content, see [List of policies](#) and [List of guidelines](#).

General

- [*The five pillars of Wikipedia*](#) (fundamental principles)
- [Policies and guidelines](#)
- [Ignore all rules](#)
- [Privacy policy](#) (Wikimedia Foundation policy)

Copyright and licensing

- [Terms of Use](#) (Wikimedia Foundation policy)
- Texts of licenses: [CC-BY-SA 3.0](#) and [GFDL](#)
- [Copyrights](#) (see also [Copyright violations](#))
- [Reusing Wikipedia content](#)
- [Non-free content criteria](#)
- [Non-free content](#)
- [Non-free use rationale guideline](#)
- [Non-U.S. copyrights](#)
- [Public domain](#)
- [Plagiarism](#)
- [Copying within Wikipedia](#)

Project roles

- [Administrators](#)
- [Oversight](#)
- [CheckUser](#) (see also [Arbitration Committee policy on CheckUser and oversight](#))
- [Volunteer response team](#)
- [Global rights policy](#)
- [Bot policy](#)

How good it is? Nature article

- 50 unattributed science articles from Wikipedia + Britannica
- Sent to academic experts to identify factual errors, critical omissions and misleading statements
 - Attempted a cross-section of science: Biology, Physics, Chemistry, etc.
 - Articles matched for length
 - 84% response rate (42 articles)
- Conclusions: Errors in both encyclopedia are roughly comparable
 - 2 serious errors in each
 - 165 less serious errors in Wikipedia vs. 123 in EB

How does it work?

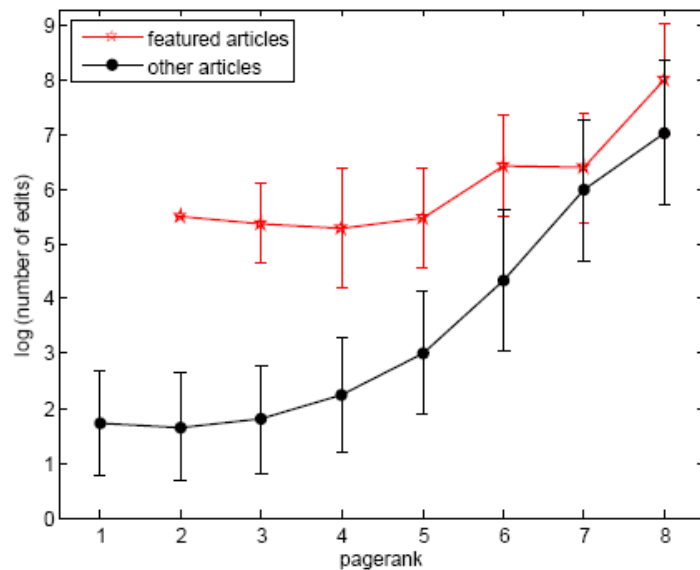
How does it work?

- “Wisdom of crowds” - Many independent judgments
 - “with enough eyeballs all bugs are shallow”
- More contributors ->
 - more information
 - fewer errors
 - less bias

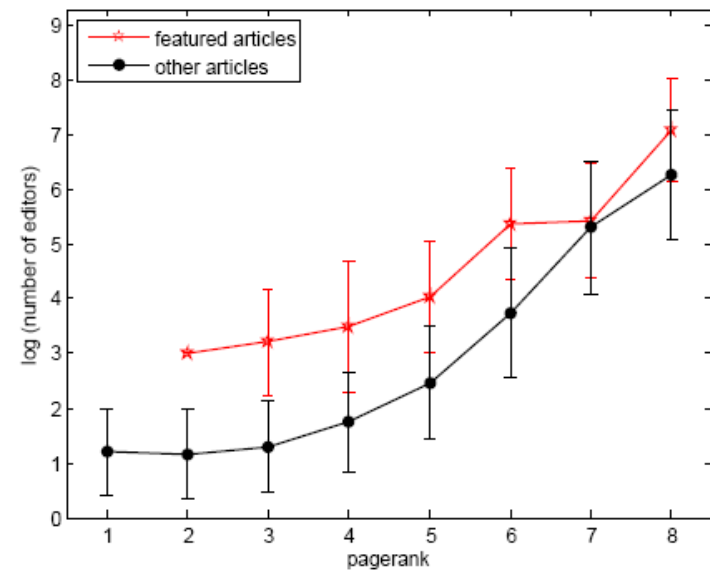
Wilkinson & Huberman, 2007

- Examined featured articles vs. non-featured articles
 - Controlling for PageRank (i.e., popularity)
- Featured articles = more edits, more editors
- More work, more people => better outcomes

Edits



Editors



Difficulties with generalizing results

- Cross-sectional analysis
 - Reverse causation: articles which become featured may subsequently attract more people
- Coarse quality metrics
 - Fewer than 2000 out of >2,000,000 articles are featured
- What about coordination?

Coordination costs

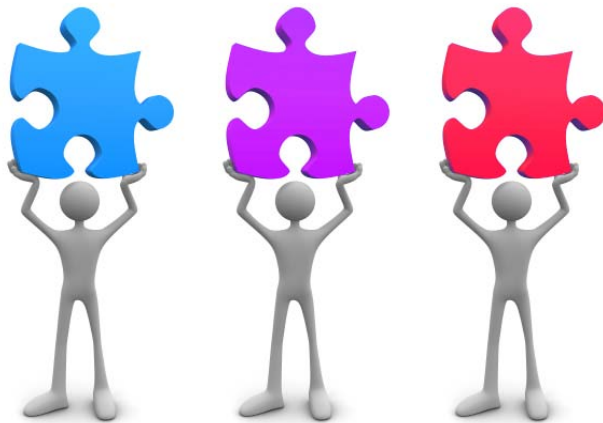
- Increasing contributors incurs process losses (Boehm, 1981; Steiner, 1972)
- Diminishing returns with added people (Hill, 1982; Sheppard, 1993)
 - Super-linear increase in communication pairs
 - Linear increase in added work
- In the extreme, costs may exceed benefits to quality (Brooks, 1975)
- The more you can support coordination, the more benefits from adding people

“Adding manpower to a late software project makes it later”

Brooks, 1975

Research question

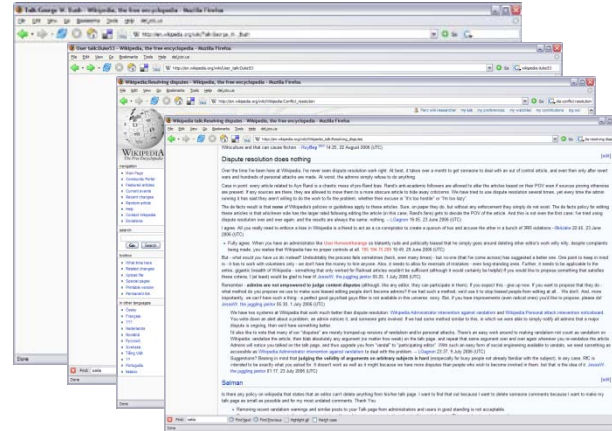
To what degree are editors in Wikipedia working independently versus coordinating?



Research infrastructure

- Analyzed entire history of Wikipedia
 - Every edit to every article
- Large dataset (as of 2008)
 - 10+ million pages
 - 200+ million revisions
 - 2.5+ Tb
- Used distributed processing
 - Hadoop distributed filesystem
 - Map/reduce to process data in parallel
 - Reduce time for analysis from weeks to hours

Types of work



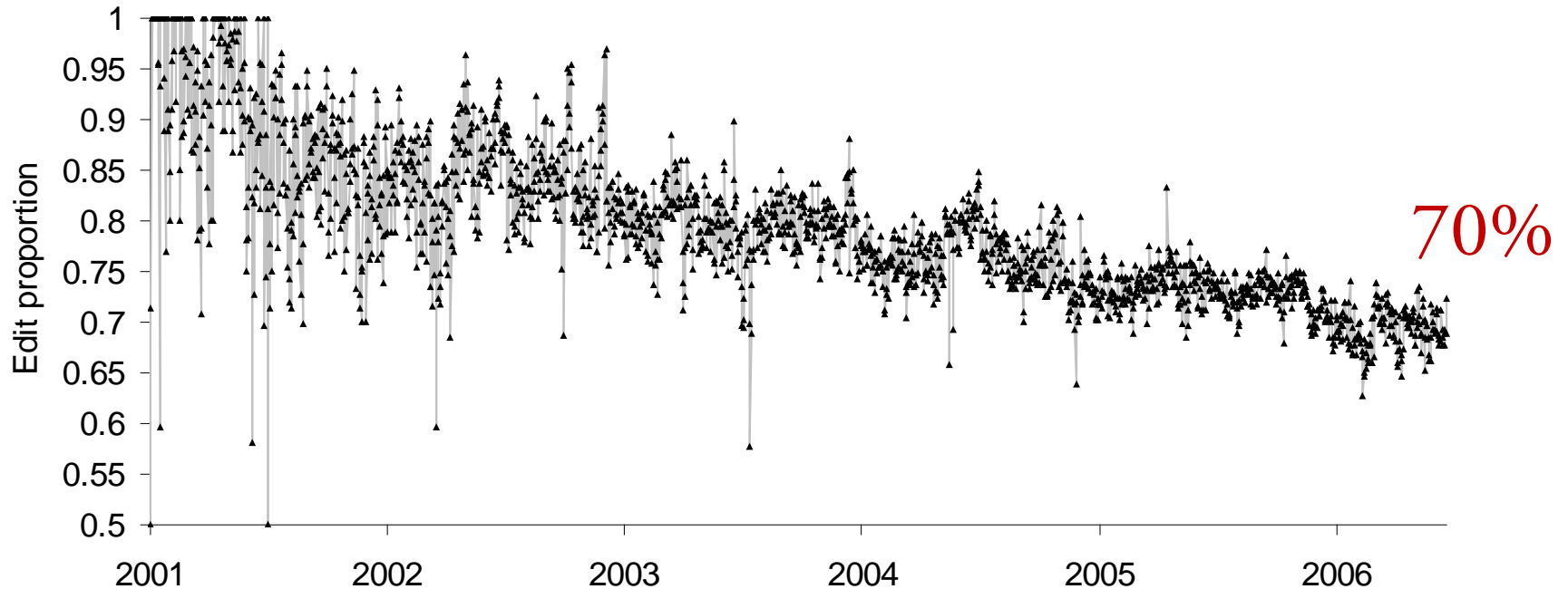
Direct work
Editing articles

Indirect work
User talk, creating policy

Maintenance work
Reverts, vandalism

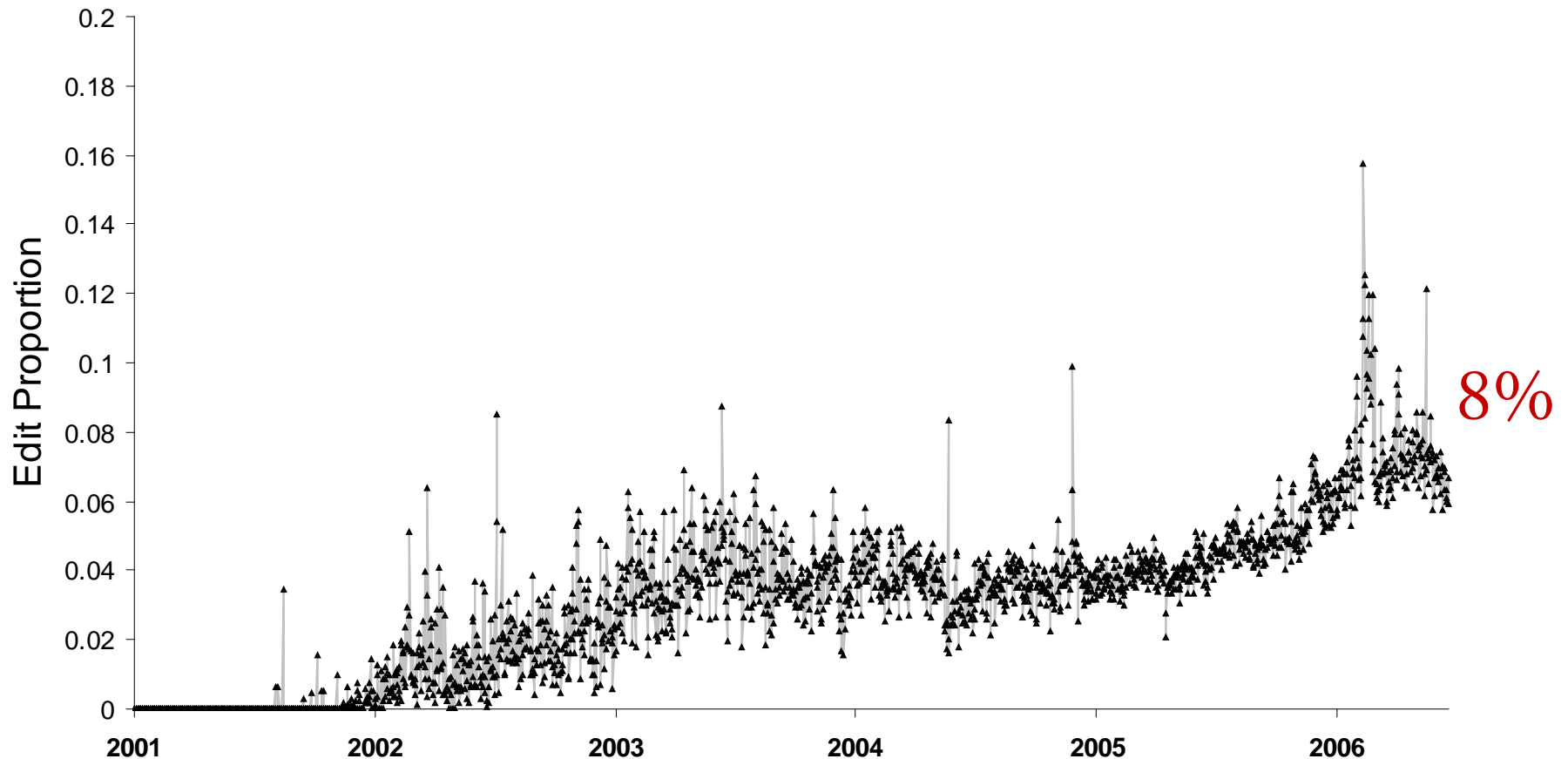
Less direct work

- Decrease in proportion of edits to article page



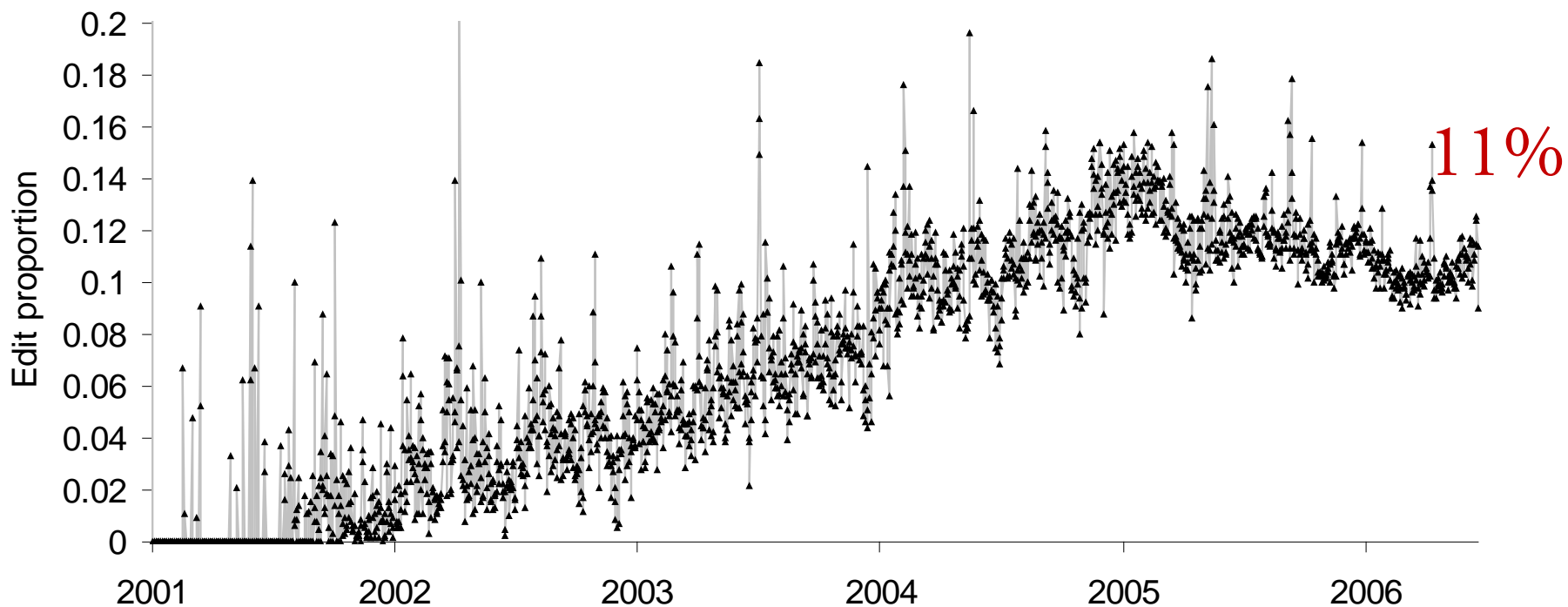
More indirect work

- Increase in proportion of edits to user talk



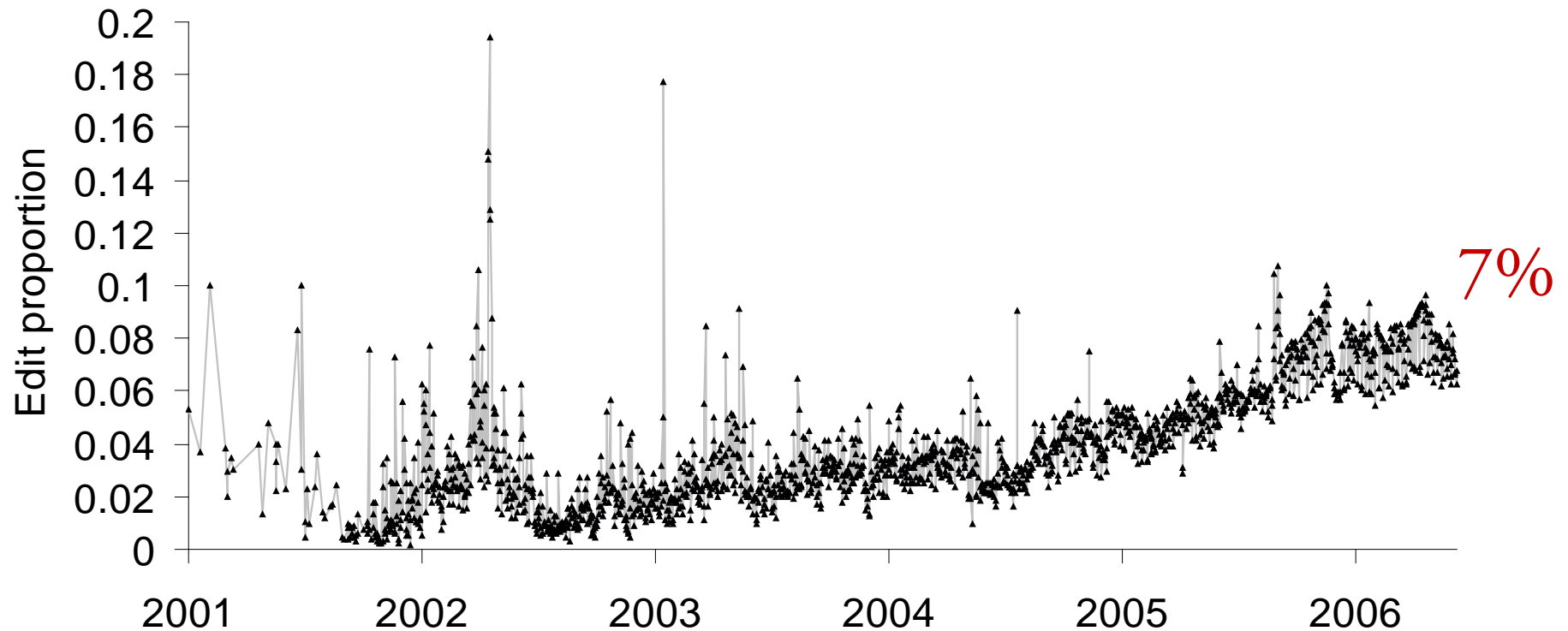
More indirect work

- Increase in proportion of edits to user talk
- Increase in proportion of edits to policy pages



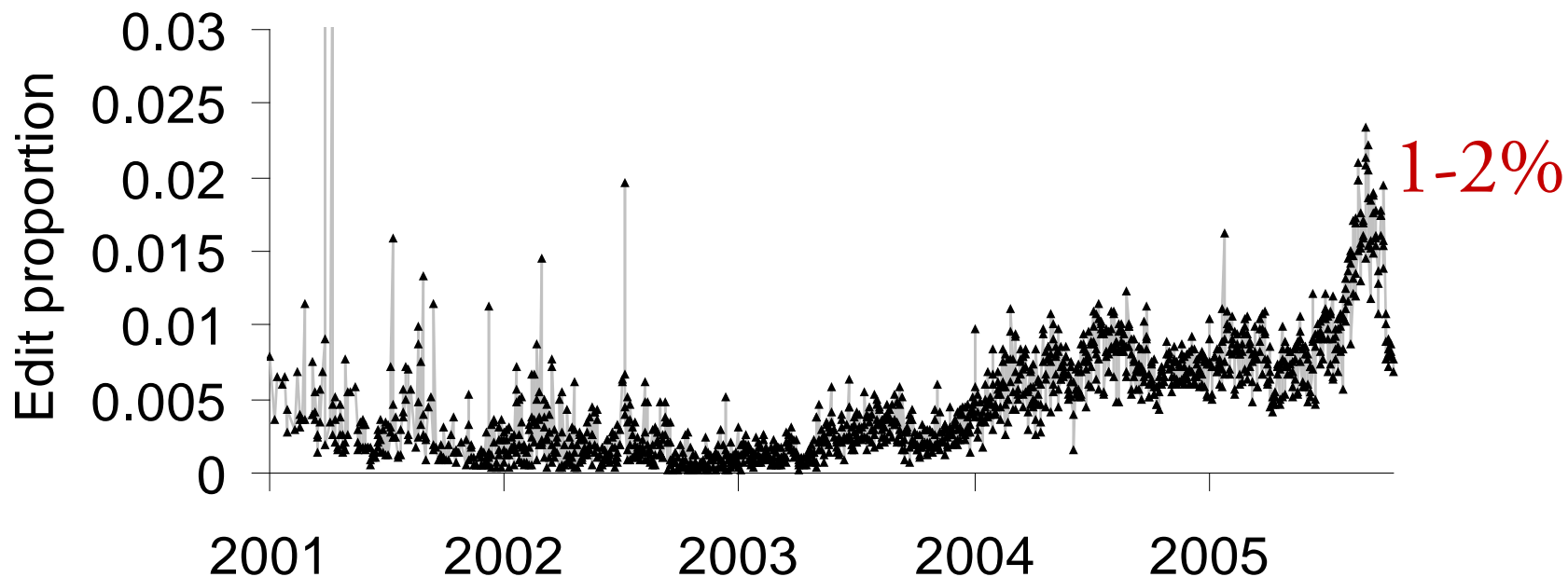
More maintenance work

- Increase in proportion of edits that are reverts



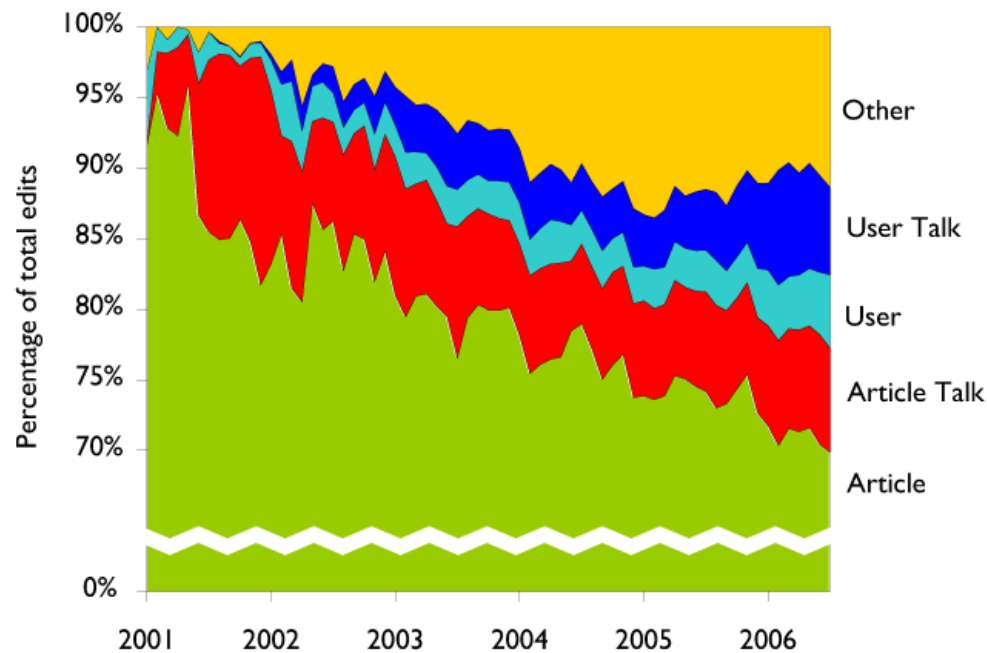
More wasted work

- Increase in proportion of edits that are reverts
- Increase in proportion of edits reverting vandalism



Global level

- Coordination costs are growing
 - Less direct work (articles)
 - + More indirect work (article talk, user, procedure)
 - + More maintenance work (reverts, vandalism)



Research question

How does coordination impact quality?



Coordination types

- Explicit coordination
 - Direct communication among editors planning and discussing article
- Implicit coordination
 - Division of labor and workgroup structure
 - Concentrating work in core group of editors

Leavitt, 1951; March & Simon, 1958; Malone, 1987; Rouse et al., 1992; Thompson, 1967

Explicit coordination: “Music of Italy”

planning

and see a list of open tasks.

+ GA This article has been **rated** as **GA-Class** on the **quality scale**.

High This article has been **rated** as **High-importance** on the **importance scale**.

This article has been rated but has no comments. If appropriate, please review the article and leave comments [here](#) to identify the strengths and weaknesses of the article and how best to improve it.

Contents [hide]

- 1 regions
- 2 sound files
- 3 Steps forward
- 4 citation needed
- 5 small change to lead
- 6 another change
- 7 imported style
- 8 Vassar Ross
- 9 GA Pass
- 10 Fair use rationale
- 11 Music history

Contents [hide]

- 1 regions
- 2 sound files
- 3 Steps forward
- 4 citation needed?
- 5 small change to lead

regions [edit]

I adjusted that section by shifting a few sentences so that we get a smoother flow from lead to northern to southern.
Jeffmatt 05:35, 24 August 2006 (UTC)

Looks good - I'm also fine with adding a paragraph on film music, though I'm not entirely sure I agree with the placement. I think I'll make some small edits tonight and hopefully make a couple sound samples (may not have time to upload them, but I'll make them).
Tuf-Kat 00:20, 25 August 2006 (UTC)

Explicit coordination: “Music of Italy”

coverage

and see a list of open tasks.

+ GA This article has been **rated** as **GA-Class** on the **quality scale**.

High This article has been **rated** as **High-importance** on the **importance scale**.

This article has been rated but has no comments. If appropriate, please review the article and leave comments [here](#) to identify the strengths and weaknesses of the article and how best to improve it.

Contents [hide]

- 1 regions
- 2 sound files
- 3 Steps forward
- 4 citation needed
- 5 small change
- 6 another change
- 7 imported style
- 8 Vasco Rossi
- 9 GA Pass
- 10 Fair use rationale
- 11 Music history

Contents [hide]

- 1 regions
- 2 sound files
- 3 Steps forward
- 4 citation needed?
- 5 small change to lead

regions [\[edit\]](#)

I adjusted that section by shifting a few sentences so that we get a smoother flow from lead to northern to southern.
[Jeffmatt](#) 05:35, 24 August 2006 (UTC)

Looks good - I'm also fine with adding a paragraph on film music, though I'm not entirely sure I agree with the placement. I think I'll make some small edits tonight and hopefully make a couple sound samples (may not have time to upload them, but I'll make them).
[Tuf-Kat](#) 00:20, 25 August 2006 (UTC)

Explicit coordination: “Music of Italy”

and see a list of open tasks.

+ GA This article has been **rated** as **GA-Class** on the **quality scale**.

High This article has been **rated** as **High-importance** on the **importance scale**.

This article has been rated but has no comments. If appropriate, please review the article and leave comments [here](#) to identify the strengths and weaknesses of the article and how best to improve it.

Contents [hide]

- 1 regions
- 2 sound files
- 3 Steps forward
- 4 citation needed
- 5 small change to lead
- 6 another change
- 7 imported style
- 8 Vasco Rossi
- 9 GA Pass
- 10 Fair use rationale
- 11 Music history

Contents [hide]

- 1 regions
- 2 sound files
- 3 Steps forward
- 4 citation needed?
- 5 small change to lead

regions

I adjusted that section by shifting a few sentences so that we get a smoother flow from lead to northern to southern.
Jeffmatt 05:35, 24 August 2006 (UTC)

Looks good - I'm also fine with adding a paragraph on film music, though I'm not entirely sure I agree with the placement. I think I'll make some small edits tonight and hopefully make a couple sound samples (may not have time to upload them, but I'll make them). **Tuf-Kat** 00:20, 25 August 2006 (UTC)

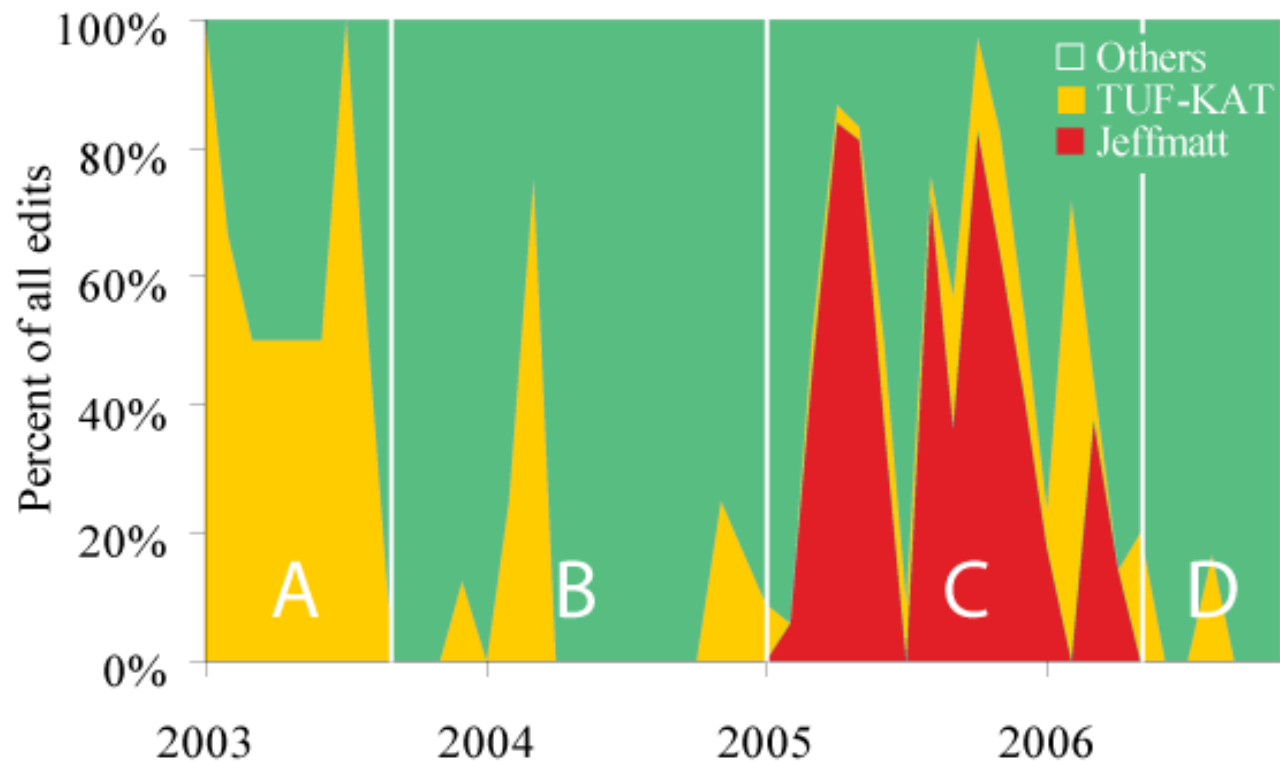
readability

Coordination types

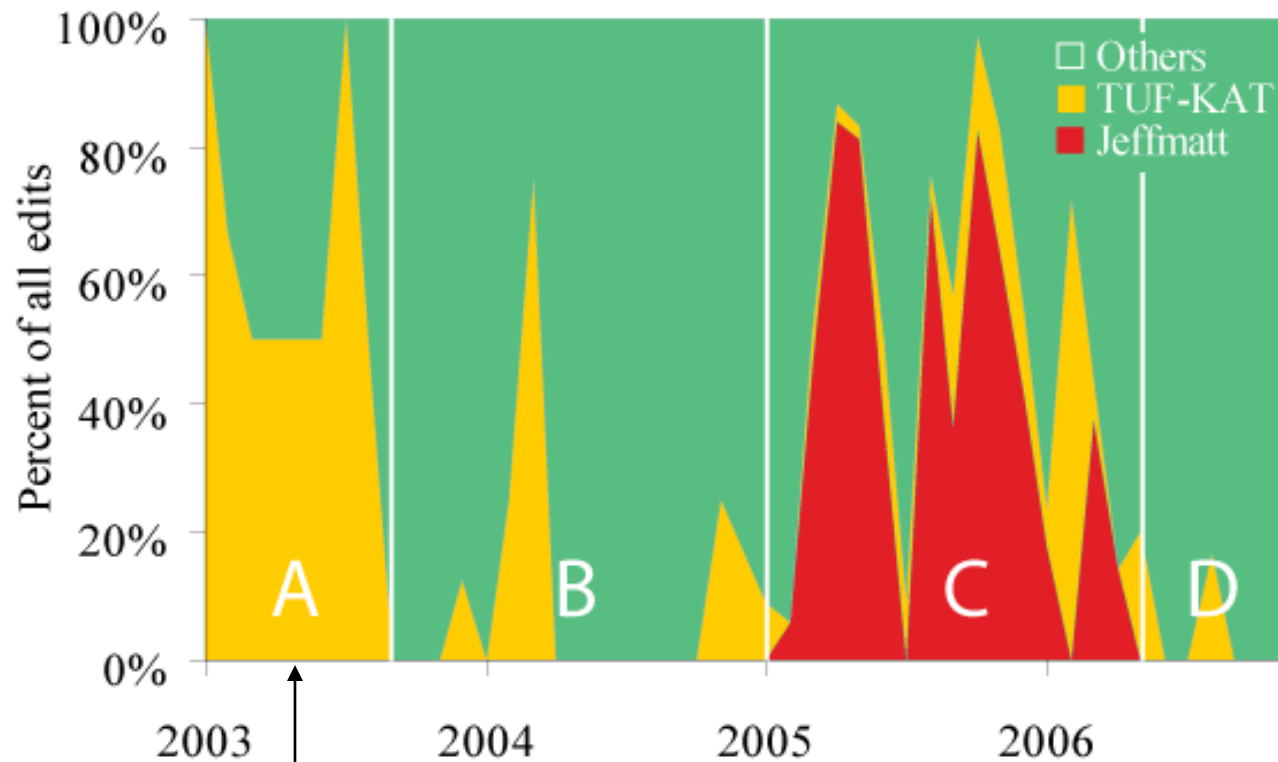
- Explicit coordination
 - Direct communication among editors planning and discussing article
- Implicit coordination
 - Division of labor and workgroup structure
 - Concentrating work in core group of editors

Leavitt, 1951; March & Simon, 1958; Malone, 1987; Rouse et al., 1992; Thompson, 1967

Implicit coordination: “Music of Italy”

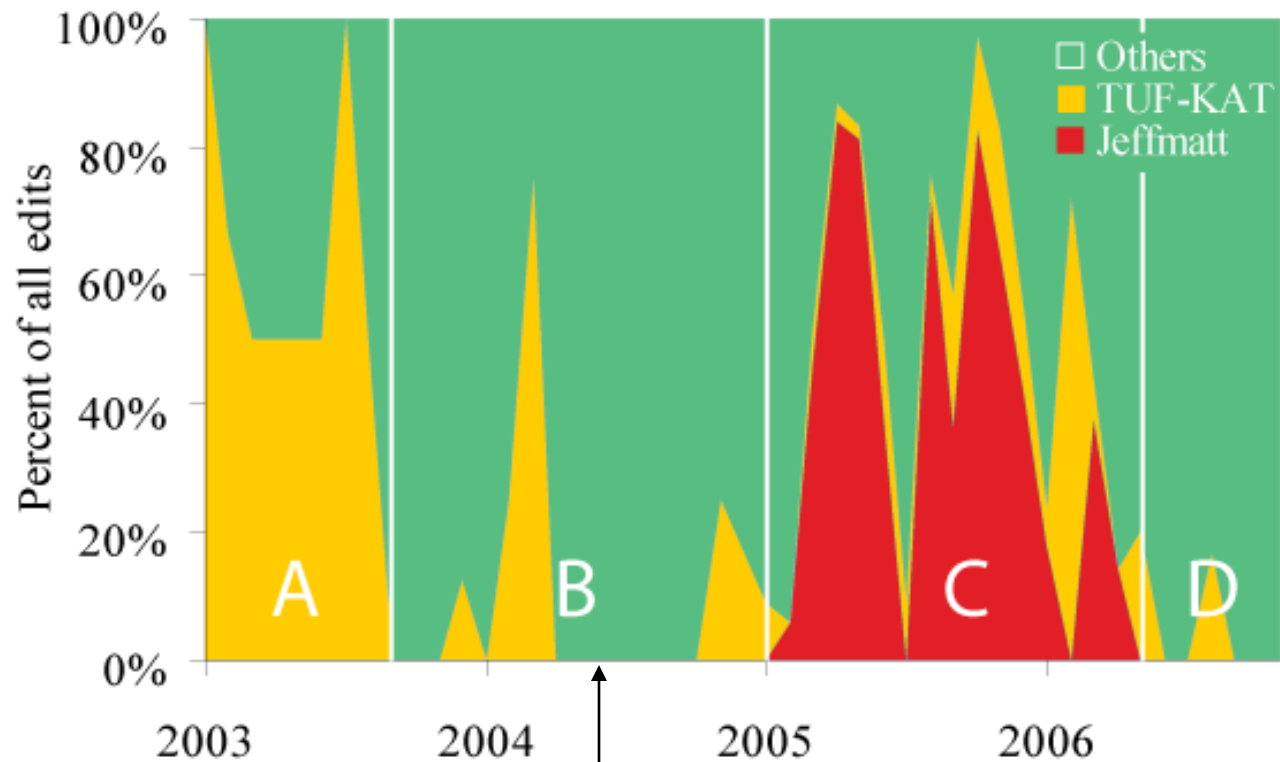


Implicit coordination: “Music of Italy”



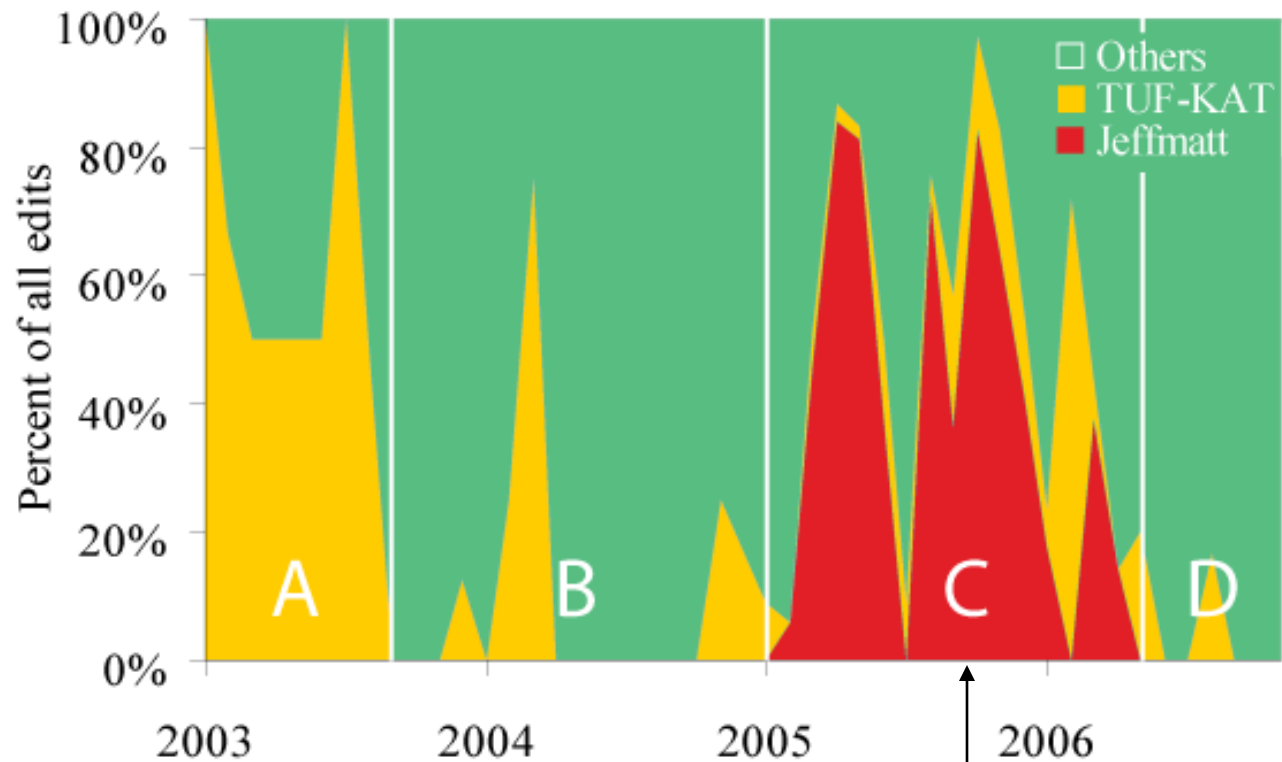
TUF-KAT: Set scope and structure

Implicit coordination: “Music of Italy”



Filling in by many contributors

Implicit coordination: “Music of Italy”



Restructuring by Jeffmatt

Research question

- What factors lead to improved quality?
 - More contributors
 - Explicit coordination
 - Number of communication edits
 - Implicit coordination
 - Concentration among editors

Measuring concentration

- If an article has 100 edits and 10 editors, it could have:
 - 10 editors making 10 edits each



Measuring concentration

- If an article has 100 edits and 10 editors, it could have:
 - 10 editors making 10 edits each
 - 1 editor making 90 edits



Measuring concentration

- If an article has 100 edits and 10 editors, it could have:
 - 10 editors making 10 edits each
 - 1 editor making 90 edits
- Measure concentration with Gini coefficient



Measuring concentration

- If an article has 100 edits and 10 editors, it could have:
 - 10 editors making 10 edits each
 - 1 editor making 90 edits
- Measure concentration with Gini coefficient

Gini = 0



Measuring concentration

- If an article has 100 edits and 10 editors, it could have:
 - 10 editors making 10 edits each
 - 1 editor making 90 edits
- Measure concentration with Gini coefficient

Gini = 0










Gini ~ 1



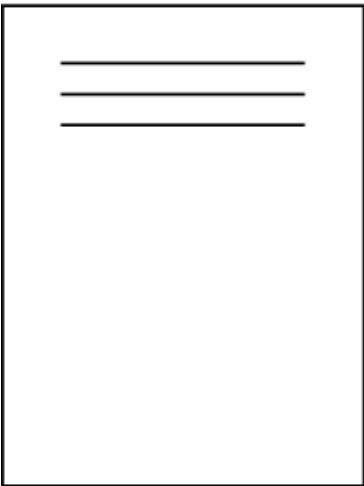
Measuring quality

- Wikipedia 1.0 quality assessment scale
 - Over 900,000 assessments
 - 6 classes of quality, from “Stub” up to “Featured”
 - Top 3 classes require increasingly rigorous peer review
- Validated community assessments with non-expert judges ($r = .54^{***}$)

Label	Criteria	Reader's experience	Editor's experience	Example
 FA (FA-Class)	Reserved exclusively for articles that have received "Featured article" status, and meet the current criteria for featured articles.	Definitive. Outstanding, thorough article; a great source for encyclopedic information.	No further additions are necessary unless new published information has come to light, but further improvements to the text are often possible.	Tourette Syndrome [ⓘ] (as of July 2007)
 FL (FL-Class)	Reserved exclusively for articles that have received "Featured lists" status, and meet the current criteria for featured lists.	Definitive. Outstanding, thorough list, a great source for encyclopedic information.	No further additions are necessary unless new published information has come to light, but further improvements to the text are often possible.	FBI Ten Most Wanted Fugitives [ⓘ] (as of January 2008)
 A (A-Class)	Provides a well-written, reasonably clear and complete description of the topic, as described in <i>How to write a great article</i> . It should be of a length suitable for the subject, with a well-written introduction and an appropriate series of headings to break up the content. It should have sufficient external literature references, preferably from reliable, third-party published sources with a reputation for fact-checking and accuracy (peer-reviewed where appropriate). Should be well illustrated, with no copyright problems. At the stage where it could at least be considered for featured article status, corresponds to the "Wikipedia 1.0" standard.	Very useful to readers. A fairly complete treatment of the subject. A non-expert in the subject matter would typically find nothing wanting. May miss a few relevant points.	Minor edits and adjustments would improve the article, particularly if brought to bear by a subject-matter expert. In particular, issues of breadth, completeness, and balance may need work. Peer-review would be helpful at this stage.	Durien [ⓘ] (as of March 2007)
 GA (GA-Class)	The article has passed through the Good article nomination process and been granted GA status, meeting the good article standards. This should be used for articles that still need some work to reach featured article standards, but that are otherwise acceptable. Good articles that may succeed in FAC should be considered A-Class articles, but having completed the Good article designation process is not a requirement for A-Class.	Useful to nearly all readers. A good treatment of the subject. No obvious problems, gaps, or excessive information. Adequate for most purposes, but other encyclopedias could do a better job.	Some editing will clearly be helpful, but not necessary for a good reader experience. If the article is not already fully wikified, now is the time.	International Space Station [ⓘ] (as of February 2007)
 B (B-Class)	Commonly the highest article grade that is assigned outside a more formal review process. Has several of the elements described in "start", usually a majority of the material needed for a comprehensive article. Nonetheless, it has some gaps or missing elements or references, needs editing for language usage or clarity, balance of content, or contains other policy problems such as copyright, Neutral Point Of View (NPOV) or No Original Research (NOR). With NPOV a well written B-class may correspond to the "Wikipedia 0.5" or "usable" standard. Articles that are close to GA status but don't meet the Good article criteria should be B- or Start-class articles.	Useful to many, but not all, readers. A casual reader flipping through articles would feel that they generally understood the topic, but a serious student or researcher trying to use the material would have trouble doing so, or would risk error in derivative work.	Considerable editing is still needed, including filling in some important gaps or correcting significant policy errors. Articles for which cleanup is needed will typically have this designation to start with.	Jammu and Kashmir [ⓘ] (as of October 2007) has a lot of helpful material but needs more prose content and references.
 Start (Start-Class)	The article has a meaningful amount of good content, but it is still weak in many areas, and may lack a key element. For example an article on Africa might cover the geography well, but be weak on history and culture. Has at least one serious element of gathered materials, including any one of the following: = a particularly useful picture or graphic = multiple links that help explain or illustrate the topic = a subheading that fully treats an element of the topic = multiple subheadings that indicate material that could be added to complete the article	Useful to some, provides a moderate amount of information, but many readers will need to find additional sources of information. The article clearly needs to be expanded.	Substantial/major editing is needed, most material for a complete article needs to be added. This article still needs to be completed, so an article cleanup tag is inappropriate at this stage.	Real analysis [ⓘ] (as of November 2006)
 Stub (Stub-Class)	The article is either a very short article or a rough collection of information that will need much work to bring it to A-Class level. It is usually very short, but can be of any length if the material is irrelevant or incomprehensible.	Possibly useful to someone who has no idea what the term meant. May be useless to a reader only passingly familiar with the term. At best a brief, informed dictionary definition.	Any editing or additional material can be helpful.	Coffee table book [ⓘ] (as of July 2005)

Analysis

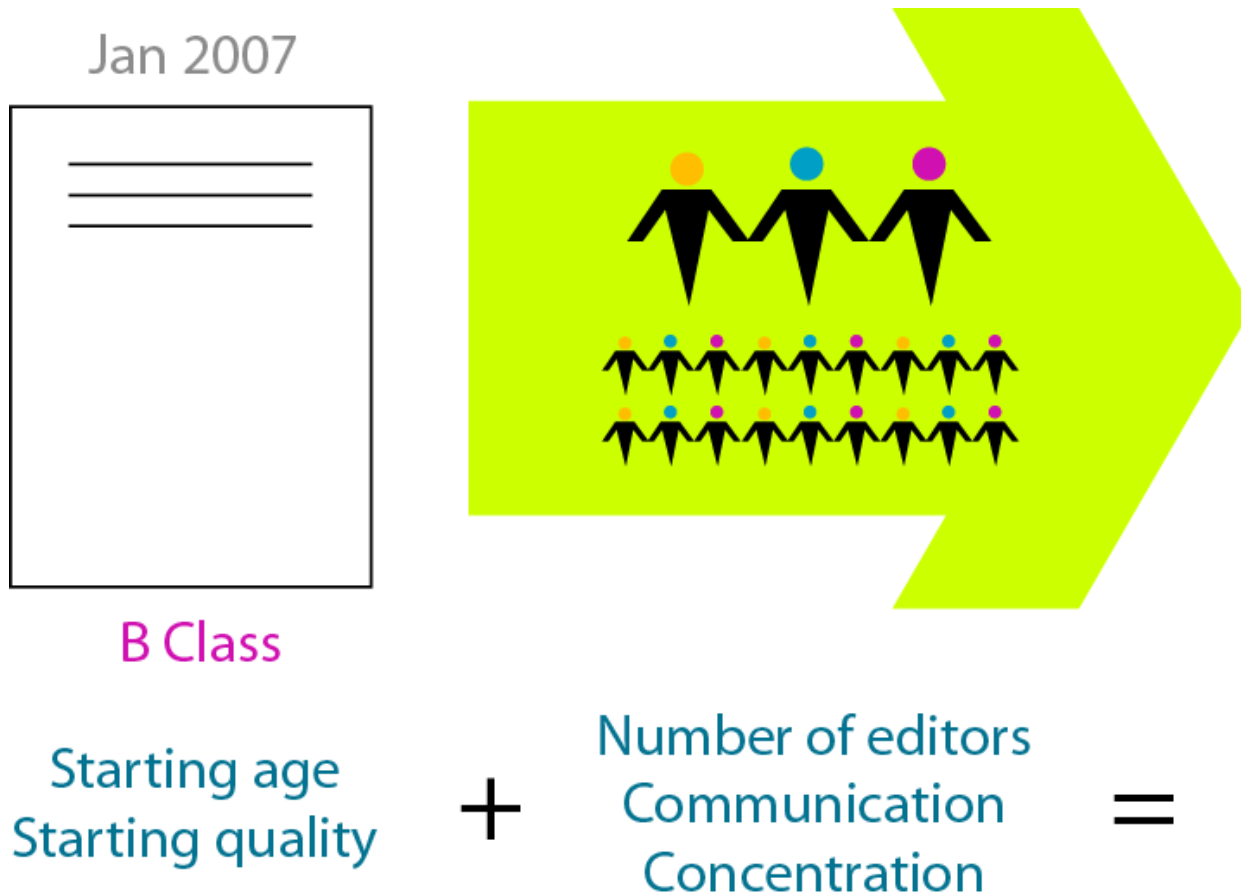
Jan 2007



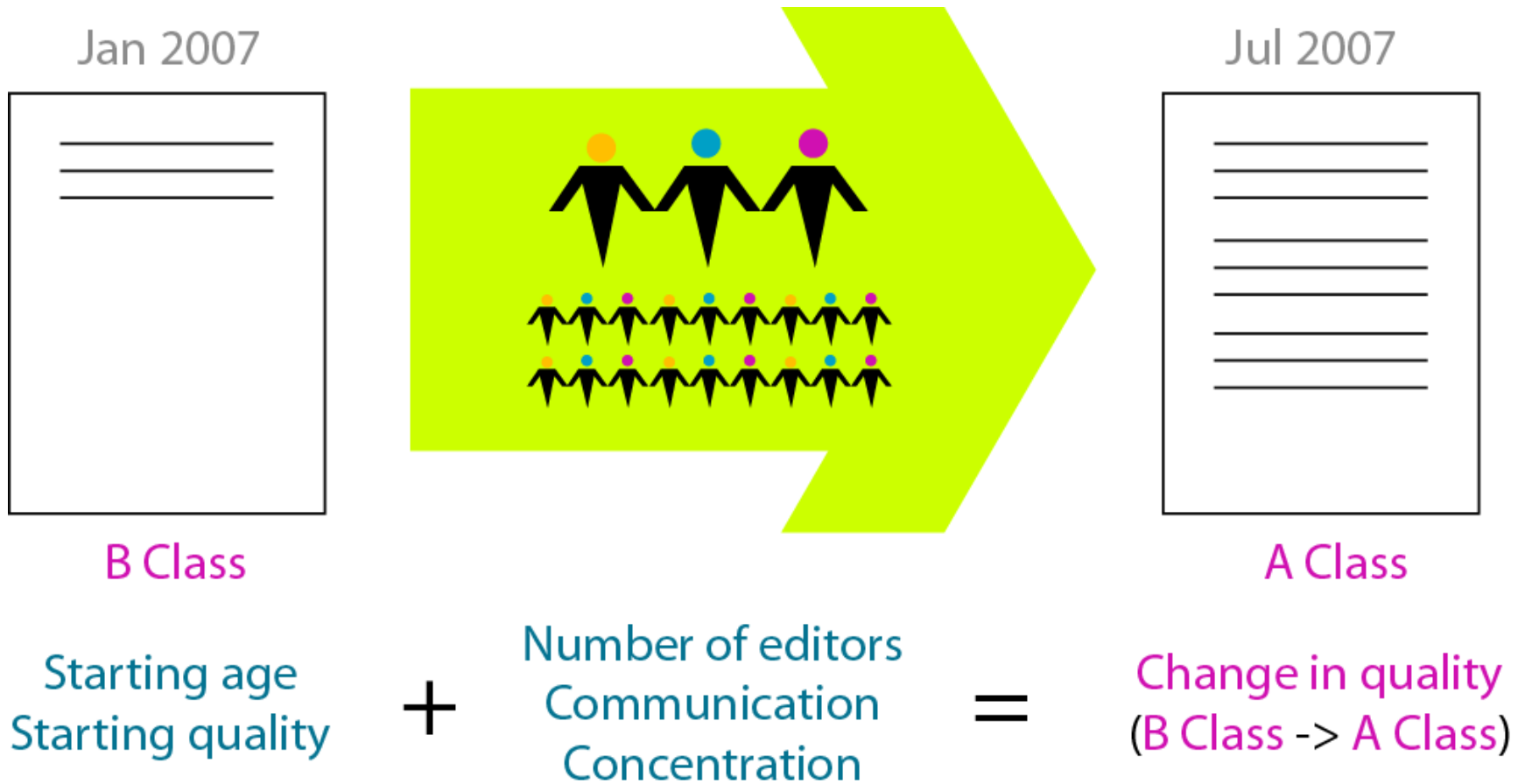
B Class

Starting age
Starting quality

Analysis

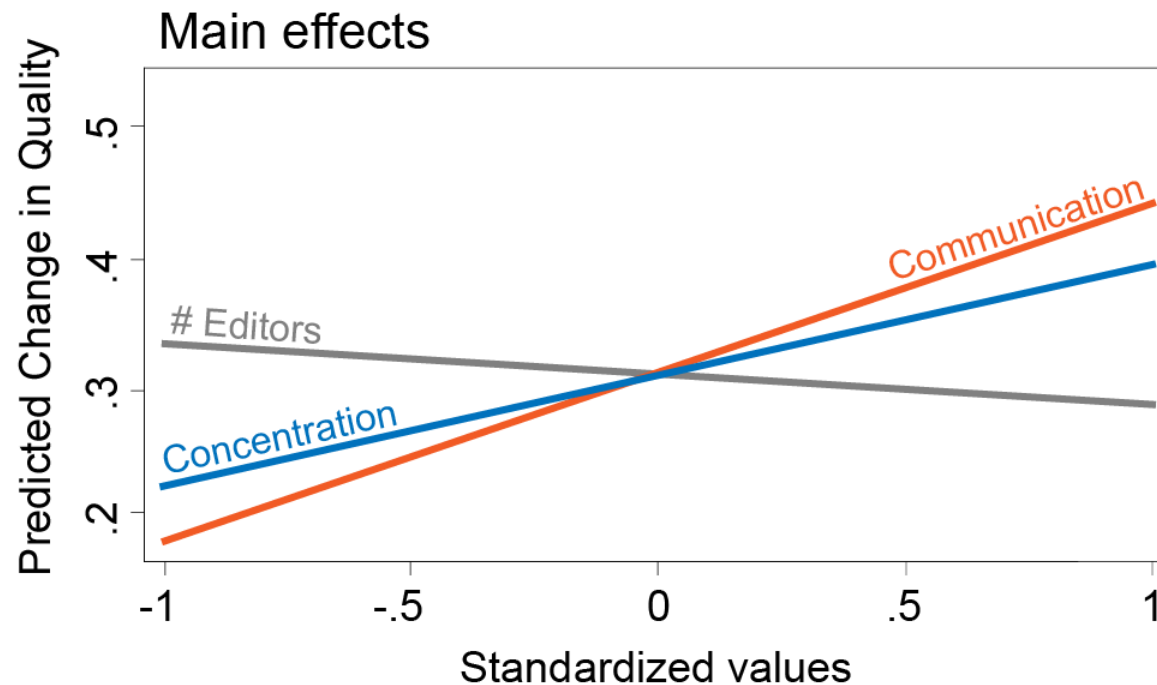


Analysis



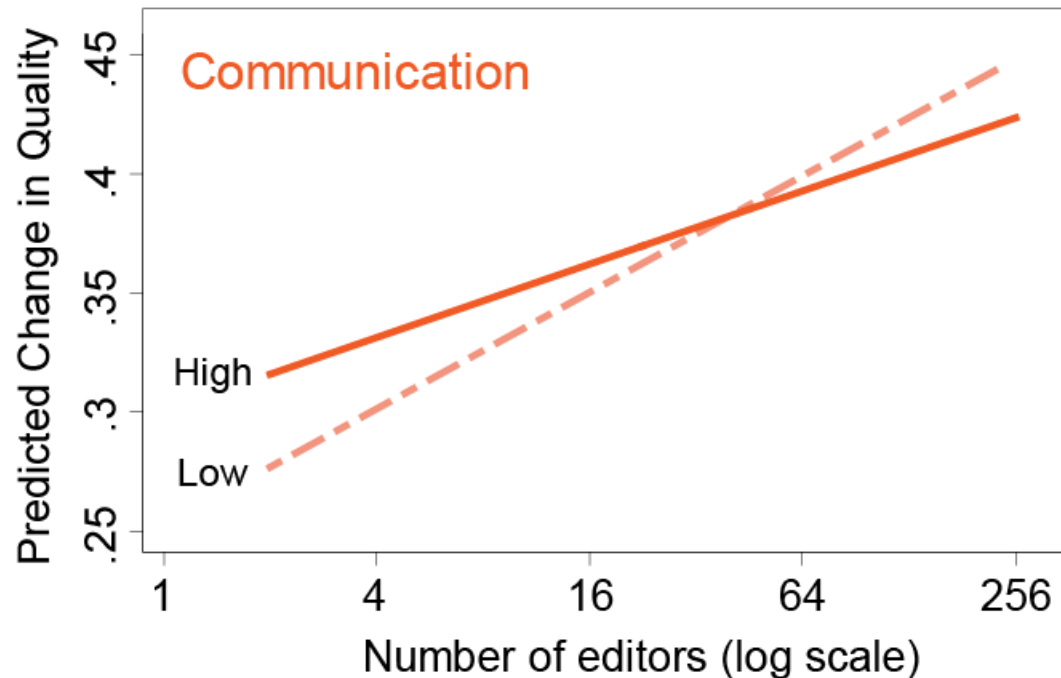
Editors + coordination

1. Editors → no effect on quality
2. Communication → increase in quality
3. Concentration → increase in quality



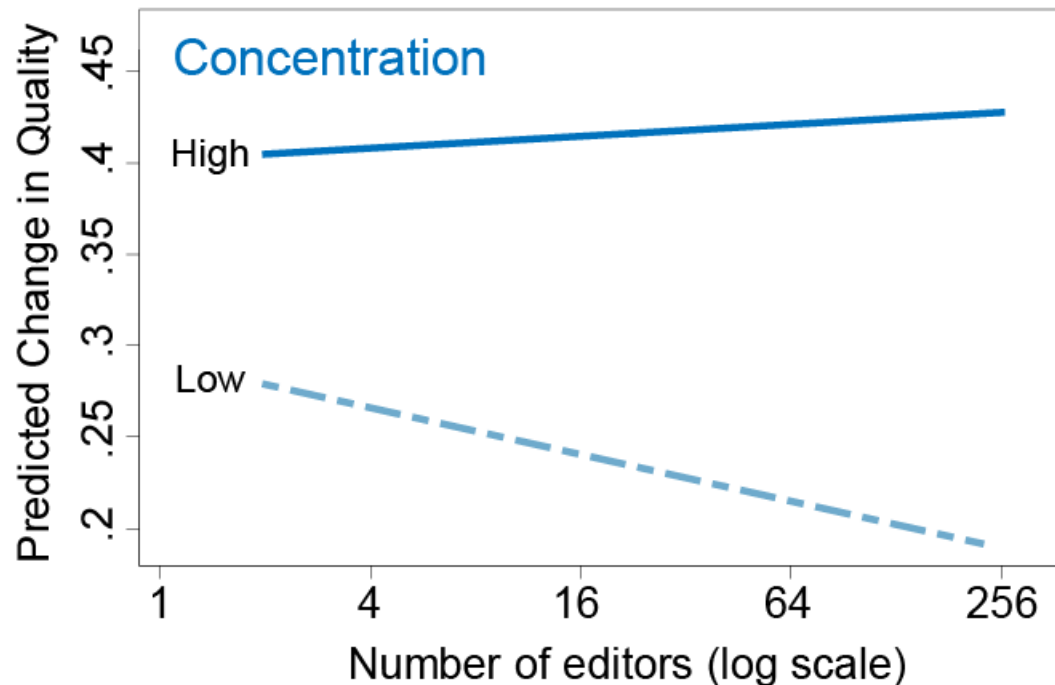
Communication x Editors

- Communication does not scale to the crowd
 - Effective with few editors
 - Ineffective with many editors



Concentration x Editors

- Concentration enables effective harnessing of the crowd
 - High concentration: more editors **increase** quality
 - Low concentration: more editors **reduce** quality















Summary

- Wikipedia includes large degree of coordination
- Adding more editors does not improve quality
 - Coordination between editors is critical
- Type of coordination is important
 - Communication does not scale to large groups
 - Concentration does scale to large groups

Measures of quality

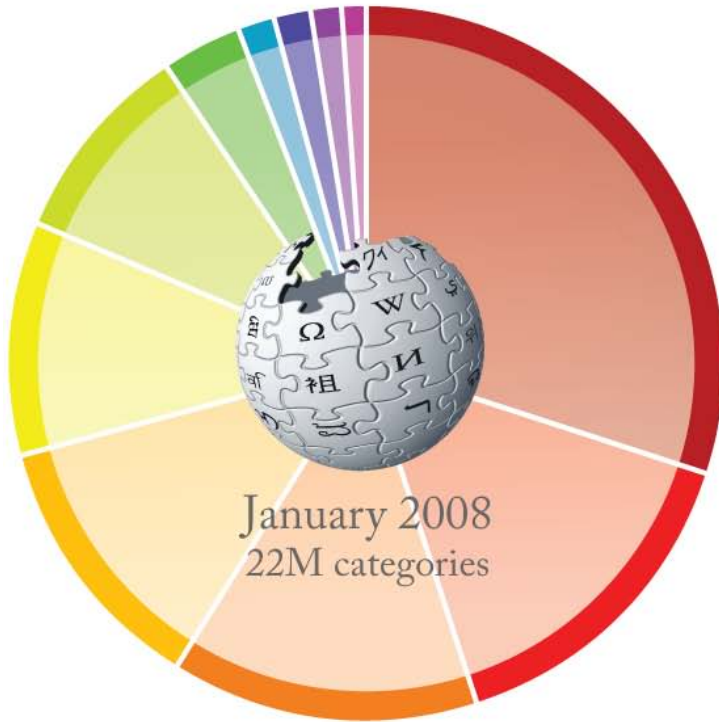
- Quality of individual topics
- Breadth and depth of coverage across topics
 - How even is the coverage?

How to quantify coverage?

Wikipedia's contents: Categorical index edit watch		
 General reference	 History and events	 Philosophy and thinking
 Culture and the arts	 Mathematics and logic	 Religion and belief systems
 Geography and places	 Natural and physical sciences	 Society and social sciences
 Health and fitness	 People and self	 Technology and applied sciences

Categories: 1879 births | 1955 deaths | People from Ulm | Academics of the Charles University | Albert Einstein | American humanists | American pacifists | American philosophers | American physicists | American socialists | American vegetarians | Cosmologists | Einstein family | ETH Zurich alumni | ETH Zurich faculty | Fellows of the Leopoldina | German agnostics | German-American Jews | German-American scientists | German humanists | German immigrants to Switzerland | German immigrants to the United States | German Jews | German Jews who emigrated to the United States to escape Nazism | German-language philosophers | German Nobel laureates | German pacifists | German physicists | German refugees | German socialists | German vegetarians | Institute for Advanced Study faculty | Jewish American scientists | Jewish American writers | Jewish inventors | Jewish skeptics | Pantheists | Jewish pacifists | Jewish philosophers | Jewish refugees | Jewish scientists | Leiden University faculty | Members of the Prussian Academy of Sciences | Naturalized citizens of the United States | Nobel laureates in Physics | Patent examiners | People associated with the University of Zurich | University of Zurich alumni | University of Zurich faculty | People from the Kingdom of Württemberg | Recipients of the Pour le Mérite (civil class) | Relativists | Stateless persons | Swiss immigrants to the United States | Swiss humanitarians | Swiss Jews | Swiss Nobel laureates | Swiss pacifists | Swiss physicists | Swiss vegetarians | Theoretical physicists | Violinists | Walhalla enshrinees | Zionists

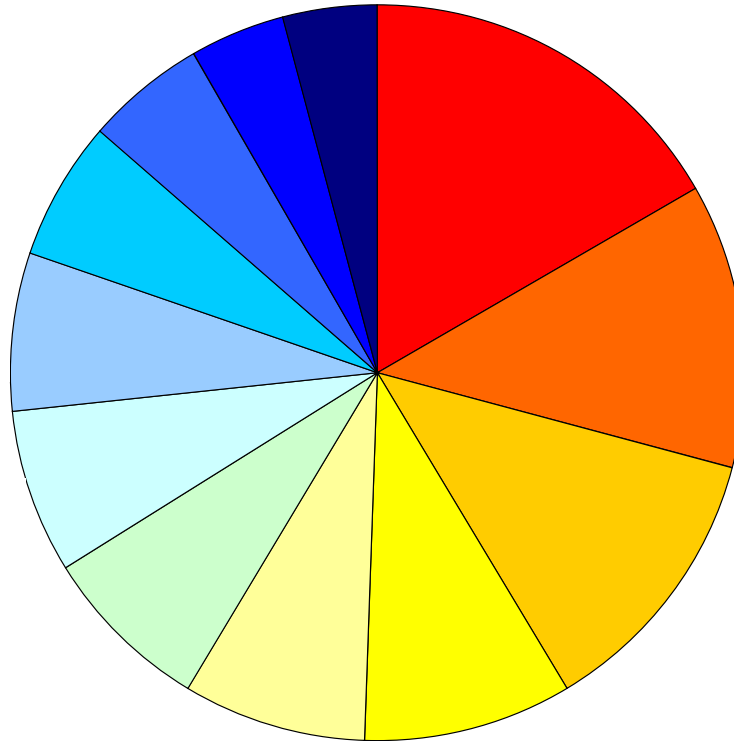
Distribution of topics in Wikipedia



Conflict in Wikipedia

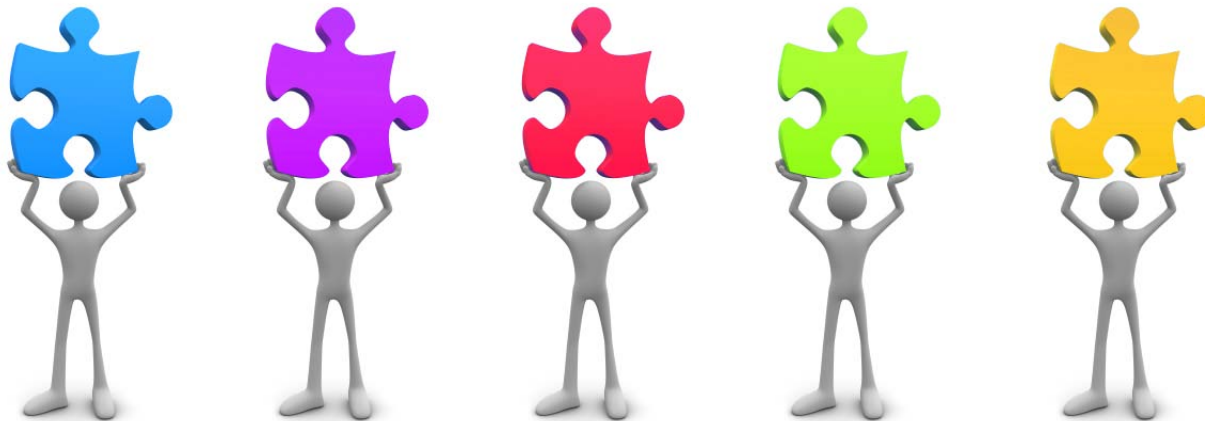
Most controversial topics

Religion



Why such uneven coverage?

- Individual self-selection of tasks
 - Scratch personal itch (Raymond, 2000)
 - Efficient user-task matching critical to success of peer production (Benkler, 2002)



Problems with self-selection

- Goals of group may not match goals of individual
- Important tasks may be neglected
 - E.g., “Britney Spears” vs. “Immune system”
- Unrewarding tasks may not get done
 - Reverting vandalism, fixing citation formatting
- These are big problems if you want a wiki to achieve the community’s goals

When people identify with a group...



...they change their behavior

Benefits of group identification

Effort
Participation
Task performance
Intrinsic motivation

Increased activity

Control by the organization
Organizationally beneficial decision making

Alignment to group goals

Information sharing
Coordinated action
Cooperation

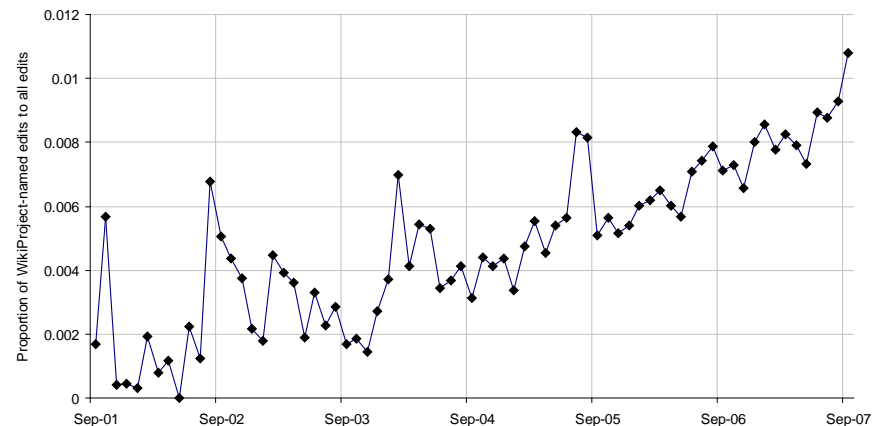
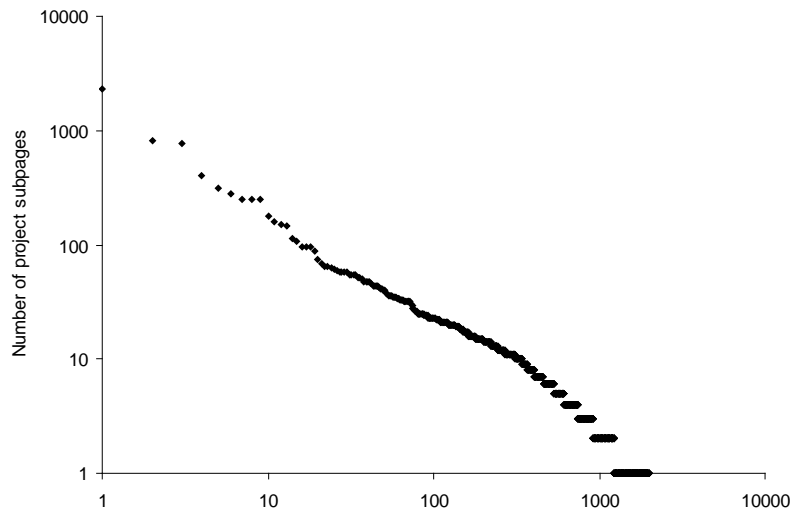
Increased coordination

Defense of the organization
Organizational citizenship behaviors

Good citizenship behavior

Wikiprojects

- ~2000 groups within Wikipedia
- Focused on specific topics
 - Military history
 - Sports
 - Tropical cyclones
- Growing in importance



WikiProject tour: Military History

- Largest WikiProject in Wikipedia
http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Military_history
- Over 700 active (and 350 inactive) members
http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Military_history/Members
- Over 80k articles in scope

The screenshot shows the Wikipedia:WikiProject Military history page. At the top, there's a navigation bar with links for 'project page', 'discussion', 'edit this page', and 'history'. Below this, the title 'Wikipedia:WikiProject Military history' is displayed, followed by the tagline 'From Wikipedia, the free encyclopedia'. A central box contains a message about the WikiProject's purpose: 'This is a WikiProject, a collaboration area and open group of editors dedicated to improving Wikipedia's coverage of a particular topic, or to organizing some internal Wikipedia process.' Below this, a blue banner announces the 'February 2008 project coordinator selection process'. The main content area is divided into sections: 'Welcome to the Military history WikiProject', 'Goals' (providing guidelines for articles), 'Scope' (defining the project's focus on historical or modern-day warfare), and 'Structure' (describing internal features like coordinators, departments, and task forces). A 'Contents' table of contents is provided, listing sections from '1 Scope' to '4.6 Portals'. On the right side, a sidebar titled 'Military history WikiProject' lists various project pages and tasks, including 'Main project page', 'Discussion archives', 'Style guide', 'Essays', 'Announcements and open tasks', 'Articles needing attention', 'Clean-up list', 'Nominations for deletion', 'New articles', 'Assessment department', 'Content department', 'Outreach department', 'Stress hotline', 'Review department', 'Image, map, and media resources', 'Showcase', 'Featured picture gallery', 'Portals', 'Project organization', and 'Templates'.

Influencing members

Production

Assessment

Clean up

Competition

Military history WikiProject announcements and open tasks
[watch](#) • [edit](#) • [discuss](#)

ARTICLES NEEDING ATTENTION • [ARTICLE & OTHER CONTENT REVIEWS](#) • [CLEAN-UP LIST](#) • [STRESS HOTLINE](#) • [WRITING CONTESTS](#)

- The February 2008 [project coordinator](#) selection process is starting. We are aiming to elect nine coordinators to serve for the next six months; if you are interested in running, please [sign up here](#) by February 14!
- The January [newsletter](#) is now available!
- Discussion about [top-level operational categories](#) continues.
- A discussion about [non-English operational codenames](#) has begun.
- A [new page for military history essays](#) has been created; project members are encouraged to post their own advice and opinions on matters within the project's purview.
- A [category restructuring workshop](#) is now open.
- [Help with converting deprecated infoboxes](#) is needed; any assistance is welcome!

Featured article candidates
[Angolan Civil War](#) • [Le Paradis massacre](#) • [Operation Camargue](#) • [Lal Masjid siege](#) • [Samuil of Bulgaria](#) • [Glorious First of June](#)

Featured article review
[Civil Air Patrol](#) • [Battle of Jutland](#) • [F-4 Phantom II](#)

Other featured content candidates
[Vietnam Veterans Memorial design](#) • [Operation Desert Storm](#)

Good article candidates
Full list...

Peer review
[Battle of Messines](#) • [Auxiliaries \(Roman military\)](#) • [List of Indian Air Force Bases](#) • [South Atlantic Medal](#) • [Heuschrecke 10](#) • [Battle of Kleidion](#) • [HMAS Melbourne \(R21\)](#) • [173rd Airborne Brigade Combat Team \(United States\)](#) • [United States Army](#) • [Operation Camargue](#) • [USS Bridgeport \(AD-10\)](#)

Picture peer review
[Auckland War Memorial Museum](#)

Articles needing attention
[...to referencing and citation](#) • [...to coverage and accuracy](#) • [...to structure](#) • [...to grammar](#) • [...to supporting materials](#) • [...to tagging](#)

[View full version \(with task force lists\)](#)

The problem with self-selection

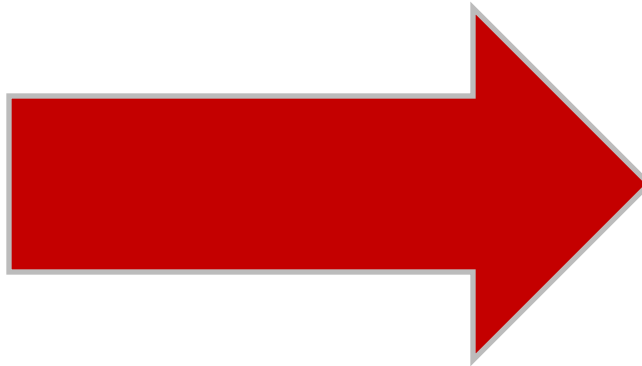
“Much of the work that members must do to sustain a successful WikiProject—almost anything beyond the actual writing of articles—is tedious, often unrewarding, and usually unappreciated.”

Solution: Group identification

“Only where... project members are willing to **share in the less exciting work** can a WikiProject muster the energy and direction to produce excellent articles **systematically rather than incidentally.**”

“To be effective, a WikiProject must foster not only interest in the topic of the project, but also an **esprit de corps** among its members.”

Joining a group in Wikipedia...



This is a [WikiProject](#), a collaboration area and open group of editors dedicated to improving Wikipedia's coverage of a particular topic, or to organizing some internal Wikipedia process.

Please see the [Guide to WikiProjects](#) and the [Directory of WikiProjects](#) for more information.

Shortcuts:
[WP:SPORT](#)
[WP:SPORTS](#)



...should lead to:

Benefits of group identification

Effort
Participation
Task performance
Intrinsic motivation

Increased activity

Control by the organization
Organizationally beneficial decision making

Alignment to group goals

Information sharing
Coordinated action
Cooperation

Increased coordination

Defense of the organization
Organizational citizenship behaviors

Good citizenship behavior

H1. Increased activity



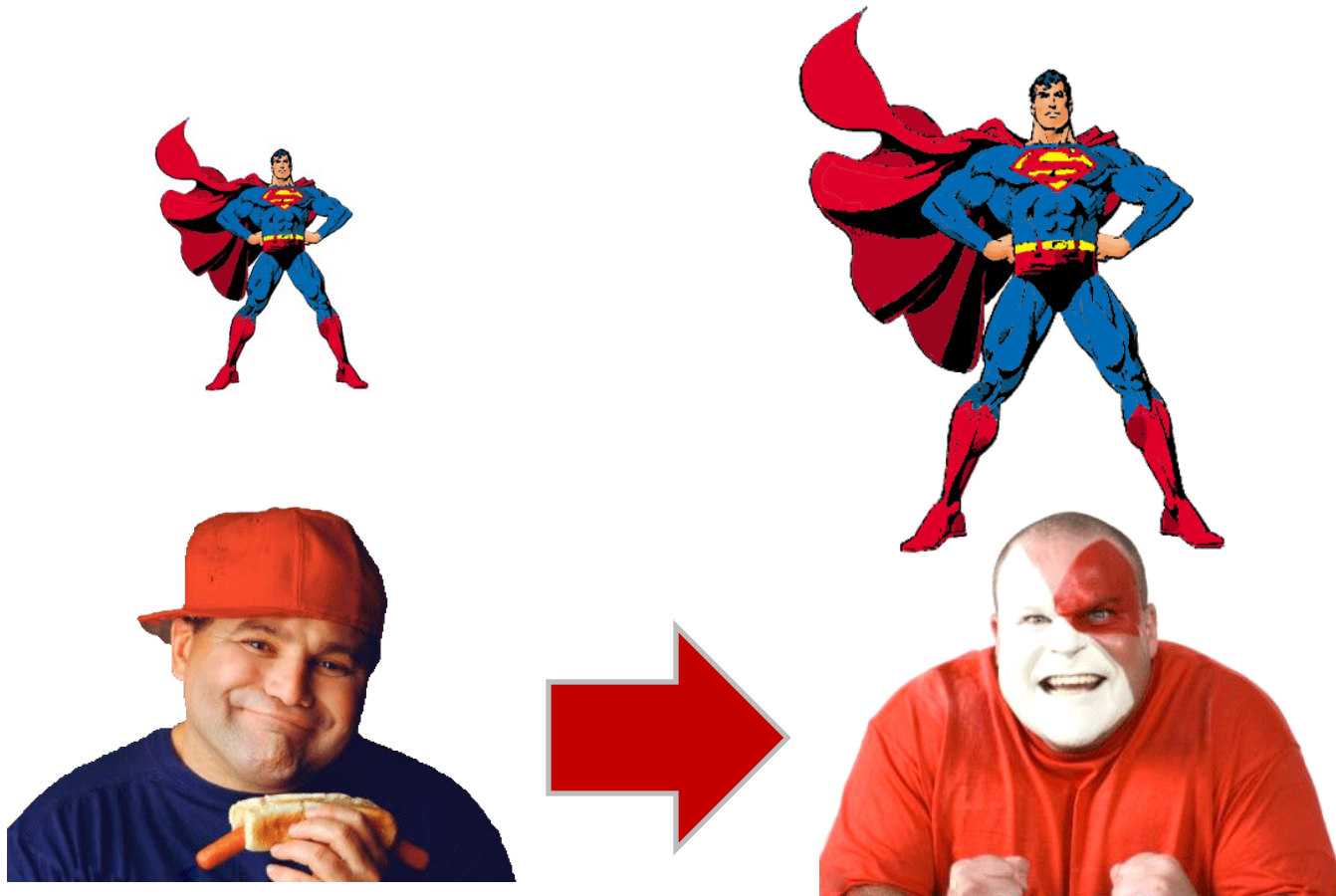
H2. Alignment to group goals



H3. Increased coordination



H4. Good citizenship behaviors



The data

- Wikipedia database dump from October 2007
 - 144 million revisions
 - ~2.5 TB text
- Chose 73 WikiProjects
 - 22 of the top 40 largest
 - 50 states + D.C.
- Sampled 125 members and matched controls
- ~8000 editors, 133,000 person-months of observations
- Used Yahoo!'s M45 cluster
 - Enabled us to process data in hours instead of months

Analysis method

- Negative binomial regression
- Hierarchical linear modeling with editor as random effect

Independent variables

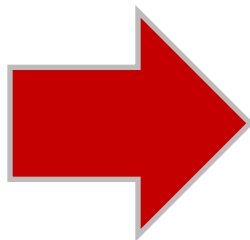
- Becomes member
- Has joined project
- In-project work (article)
- Coordination work
- Age

Dependent variables

- Number of edits
- Vandalism fighting

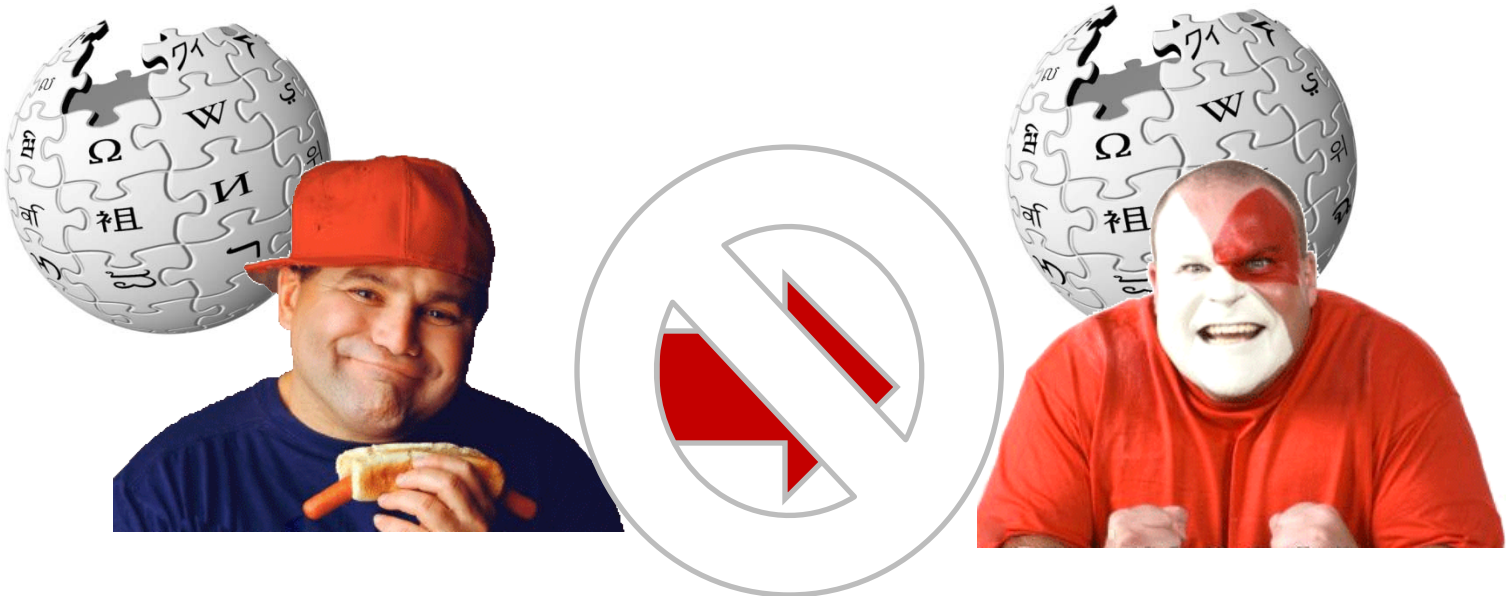
Pre-post analysis

- Comparing behavior before joining to behavior after joining
- Controls for individual differences, such as motivation, expertise, interests, etc.



H1. Increased activity

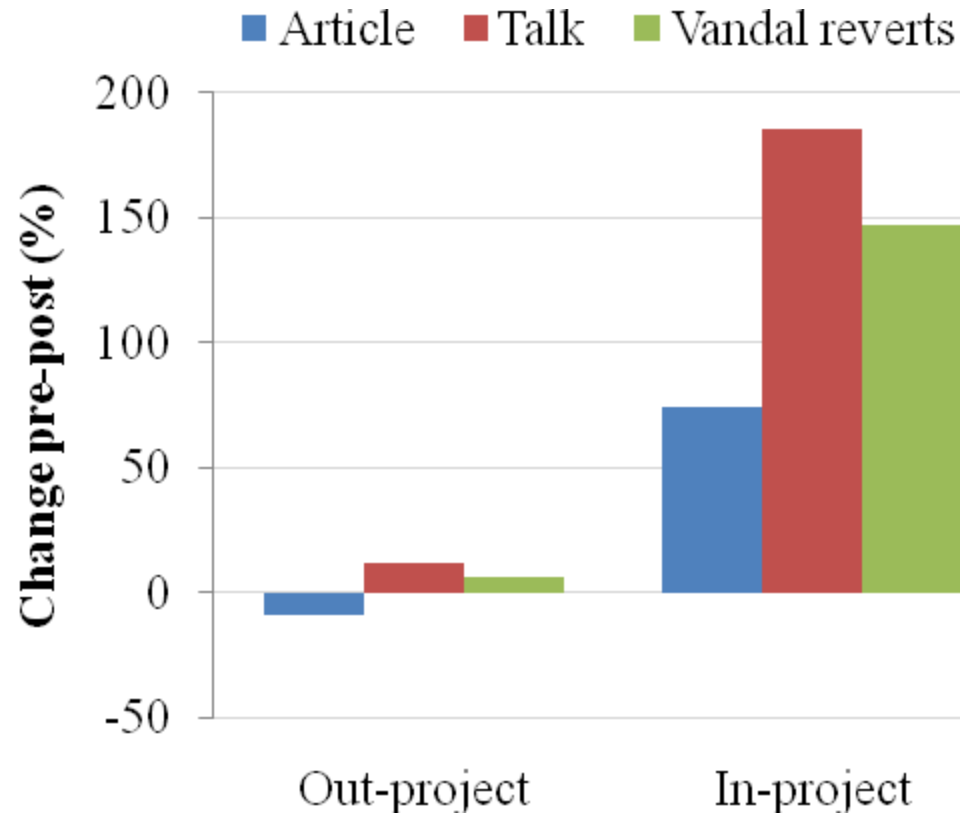
1.6% increase in total editing
(statistically significant, but not practically)



H2. Alignment to group goals



H2. Alignment to group goals



Out-project



In-project

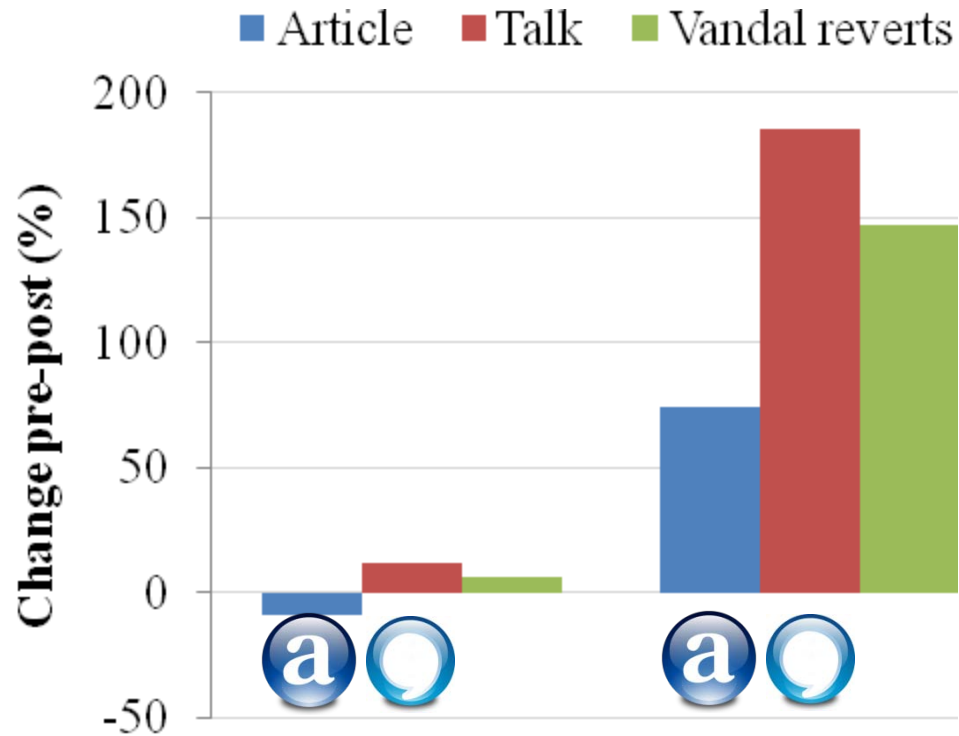


H3. Increased coordination

15% increase in overall communication



H3. Increased coordination



Out-project

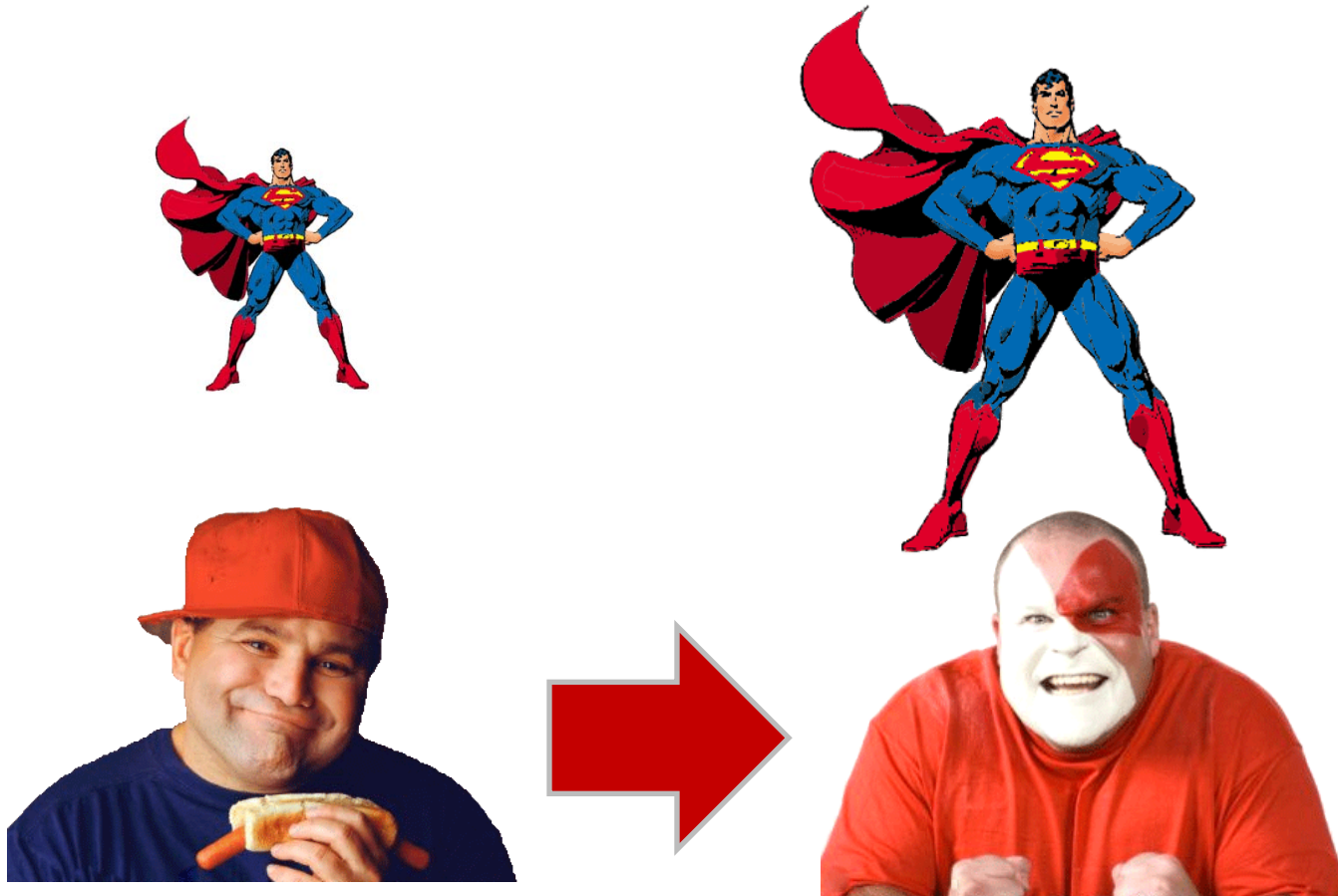


In-project

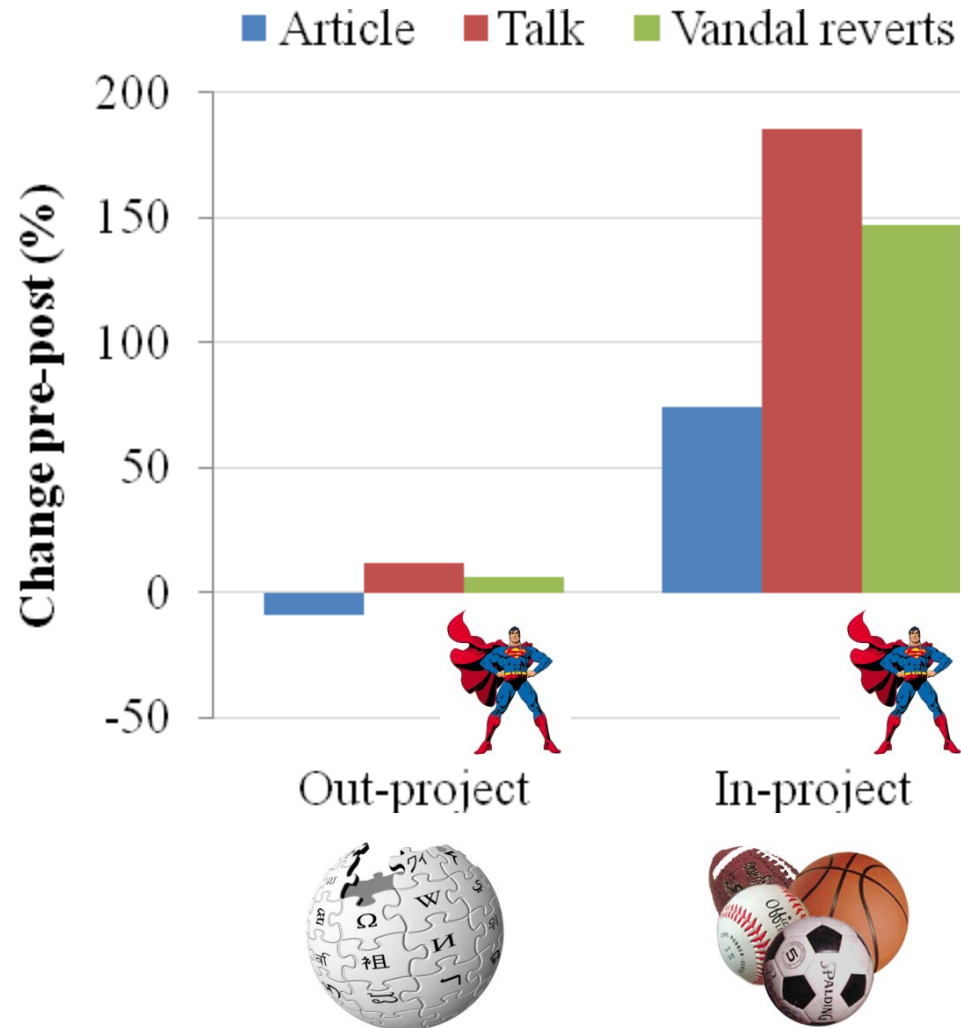


H4. Good citizenship behaviors

50% increase in vandalism fighting



H4. Good citizenship behaviors



Conclusions

- Self-selection of tasks can be a problem for many wiki communities
 - Do work important to group
 - Do tedious, unrewarding work
- Group identification can help
 - Align work with goals of group
 - Increase coordination
 - Increase good citizenship behaviors





TOOLS FOR SOCIAL COLLABORATION

Profits and perils of user-generated content

- Content in Wikipedia can be added or changed by anyone
- Because of this, has become one of the most important information resources on the web
 - Top 10 most popular websites (Alexa.com)
 - Millions of contributors
- Also causes problems
 - Conflict between contributors
 - Unknown trustworthiness

Denning et al. (2005)

- Risks with using Wikipedia
 - Accuracy of content
 - Motives of editors
 - Expertise of editors
 - Stability of article
 - Coverage of topics
 - Quality of cited information

Insufficient information to evaluate trustworthiness

History flow

"Microsoft"
Wikipedia

COLOR

group

individual

text changes

text age

links

SPACING

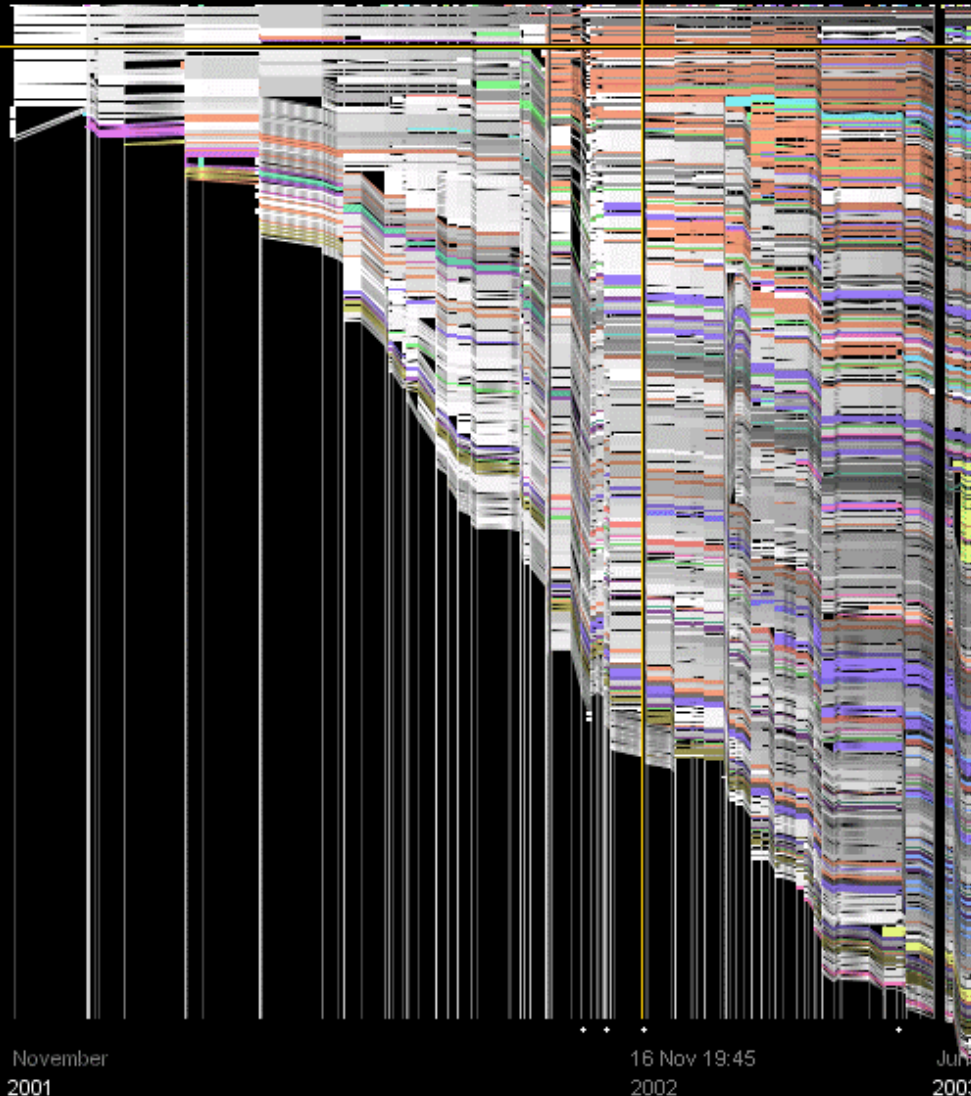
date

version

authors posts

Hannes Hirzel	4
Ed Poor	1
Taw	1
Joakim Ziegler	1
Mark Ryan	2
Conversion script	1
Ted Longstaffe	1
AxelBoldt	1
The Anome	14
PierreAbbat	1
Tarquin	1
Ellmist	1
LC	1
Wesley	1
Aldie	1
Andre Engels	4
Maveric149	1
Brion VIBBER	1
Rcingham	2
Stephen Gilbert	10
Camembert	1
DanKeshet	1
Olivier	1
Matthew Woodcraft	1
Mrwojo	1
Liftarn	1
Nixdorf	1
Hotlorp	1
Youssefsan	3
Lezek	1
Nate Silva	2
Elano	1
Snoyes	2
TakuyaMurata	1
Nknight	1
Notheruser	1
Aarrroqghhh	1

K.lee "fixing trustworthy computing link"



[compatible personal computers](#) sold. See [Microsoft Windows](#) for a general overview, and [History of Microsoft Windows](#) for a detailed history.

The company's older flagship operating system was MS-DOS, which used a [command line interface](#). Early versions of Windows required the use of MS-DOS code. This requirement was eliminated in [Windows NT](#) and its descendants, which include [Windows 2000](#) and [Windows XP](#).

[Microsoft Office](#) is the company's line of [office software](#). It includes [Word](#) (a word processor), [Excel](#) (a [spreadsheet](#)), and [PowerPoint](#) (presentation software). Microsoft also produces Microsoft Office for [Apple Macintosh](#) computers.

Hardware

Although Microsoft is primarily a software company, it also produces several [computer hardware](#) products, often to support specific software business strategies.

An early example is the Microsoft Mouse, which encouraged the use of the Windows operating system's [graphical user interface](#) (GUI). Using a GUI without a [mouse](#) was awkward, and hence the proliferation of mice would speed the widespread adoption of Windows. Later models sport scrolling wheels, extra buttons, [LED](#) motion detectors and other features.

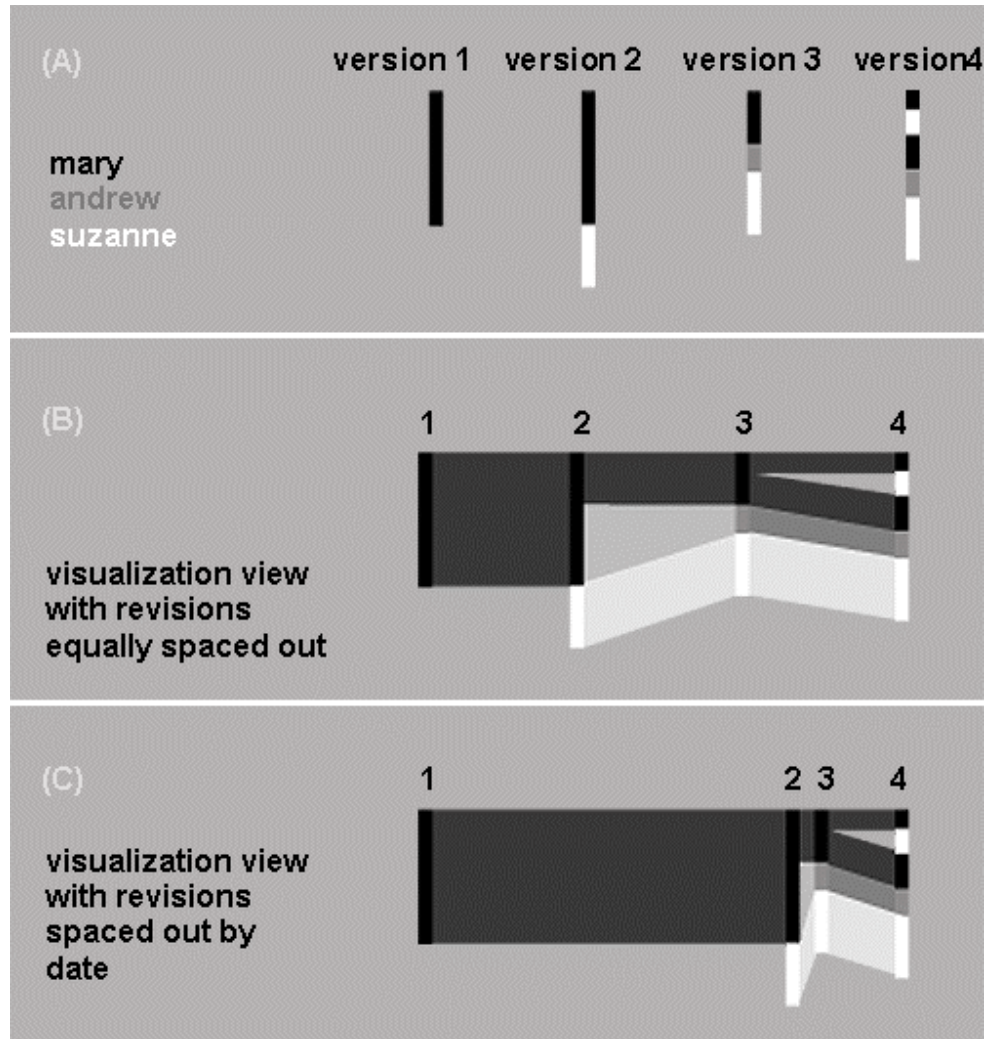
Microsoft also sells the Sidewinder line of [joysticks](#) and gamepads.

The company bought [WebTV](#), a television [Internet appliance](#) in an effort to bolster their [MSN](#) Internet service.

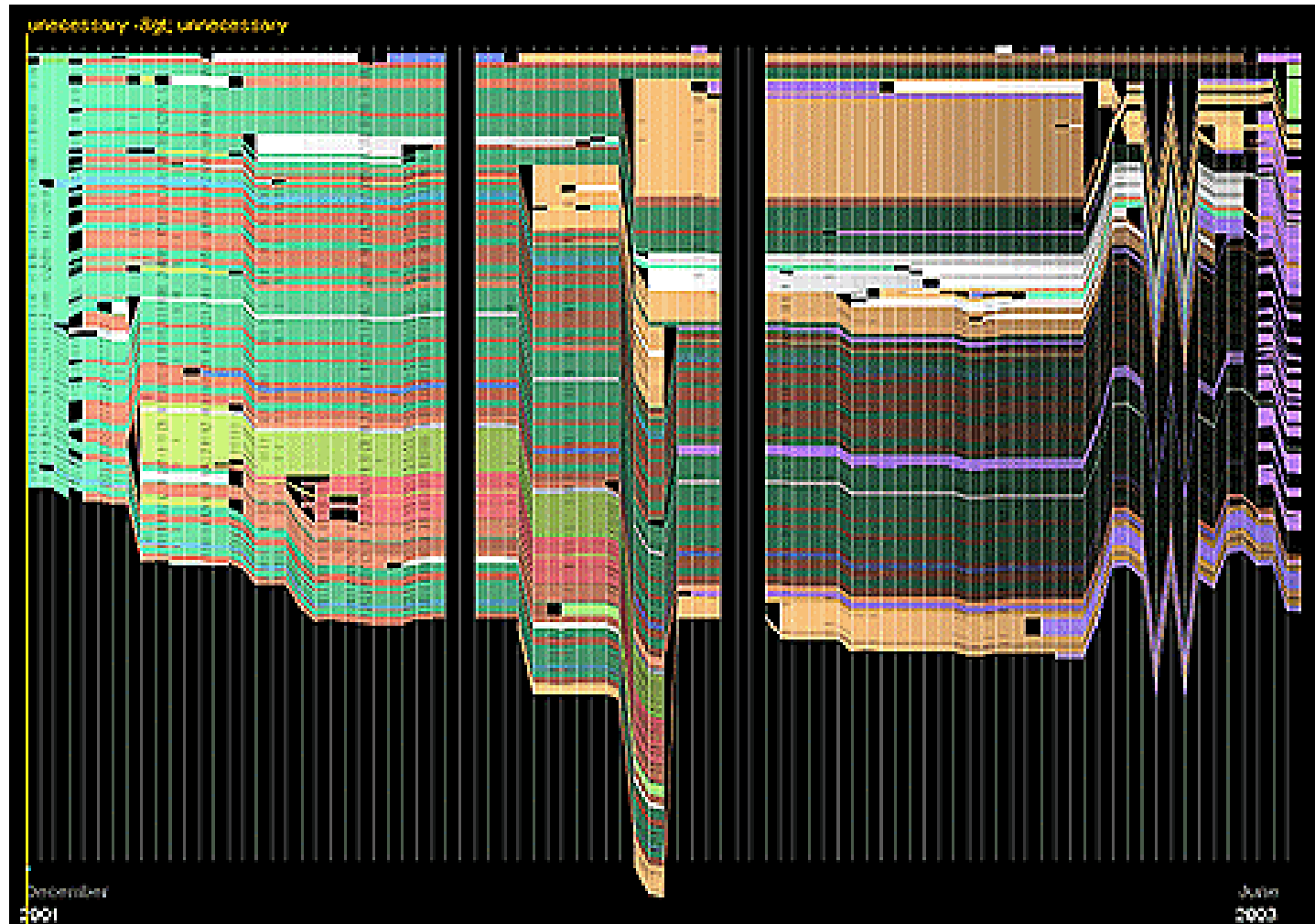
When Microsoft released the [XBox](#) in late 2001, the company entered the multi-billion dollar [game console](#) market dominated by [Sony](#) and [Nintendo](#).

History of Microsoft

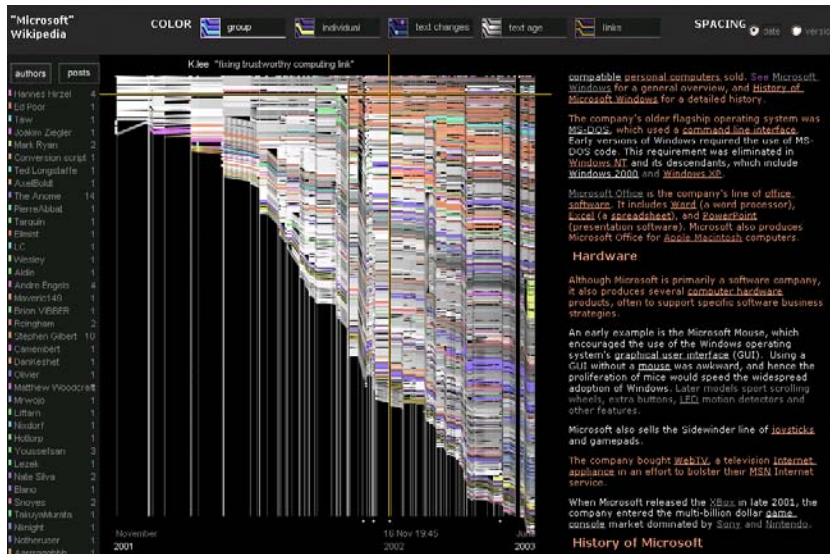
Details



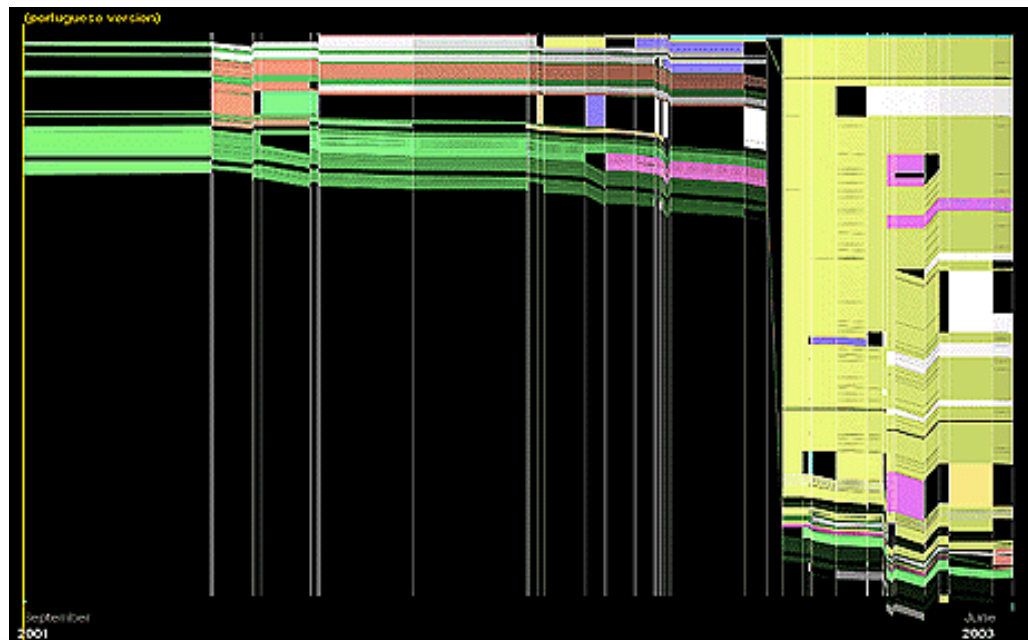
Vandalism



Anonymous contribution

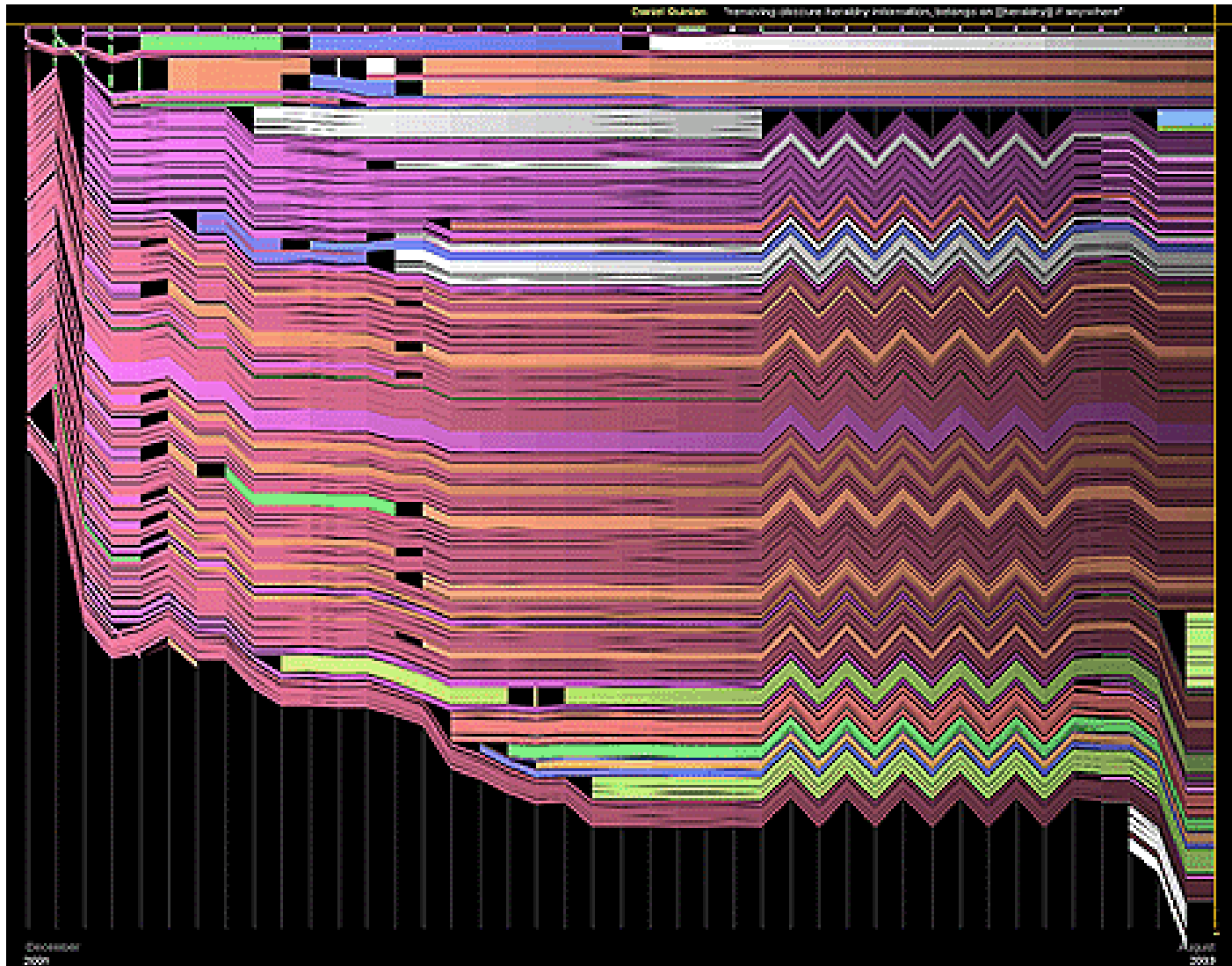


M\$: many anonymous contributors



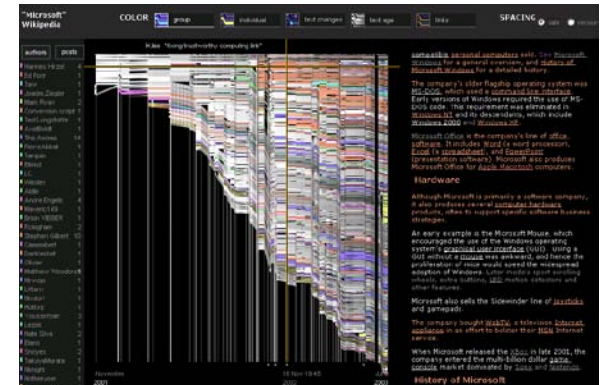
Brazil: few anonymous contributors

Edit war



Scaling visualization

- “Making sense of the history for even a single entry, however, is not straightforward. The sheer number of versions can be daunting: as of August 2003, the entry for Microsoft had 198 versions comprising 6.2 MB of text”
- As of March 2008, Microsoft had 7951 revisions, 40x as many as in their paper
 - George W. Bush has more than 37,000 revisions, almost 200x the size
- How can a visualization scale to these sizes?



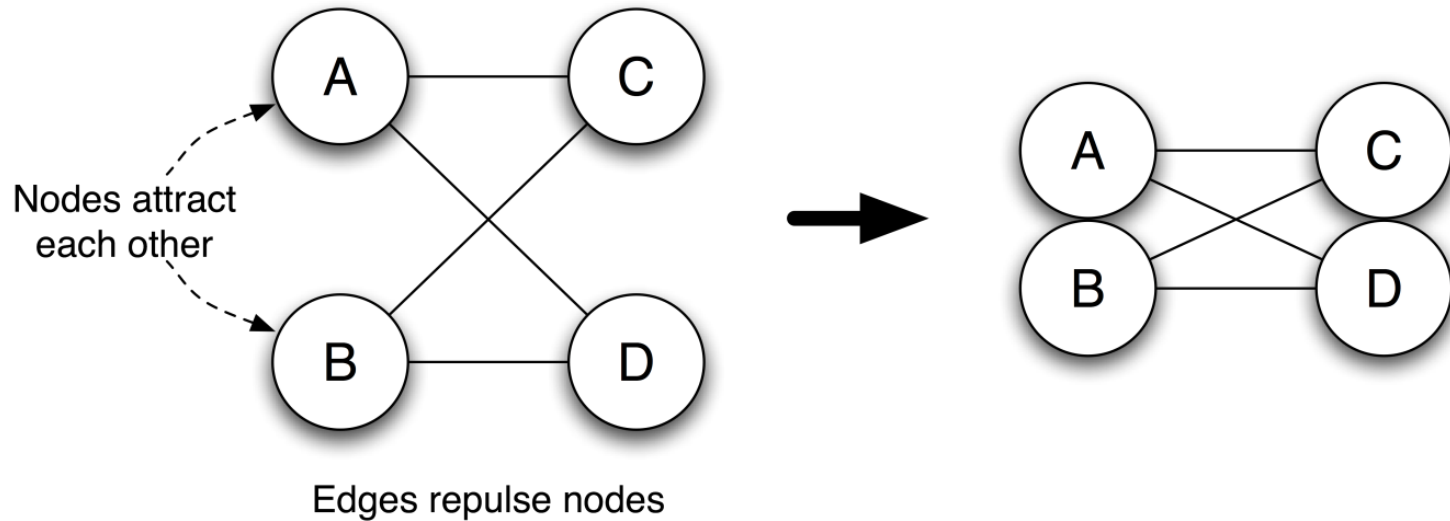
Assigning trust (Adler, Alfaro, et al.)



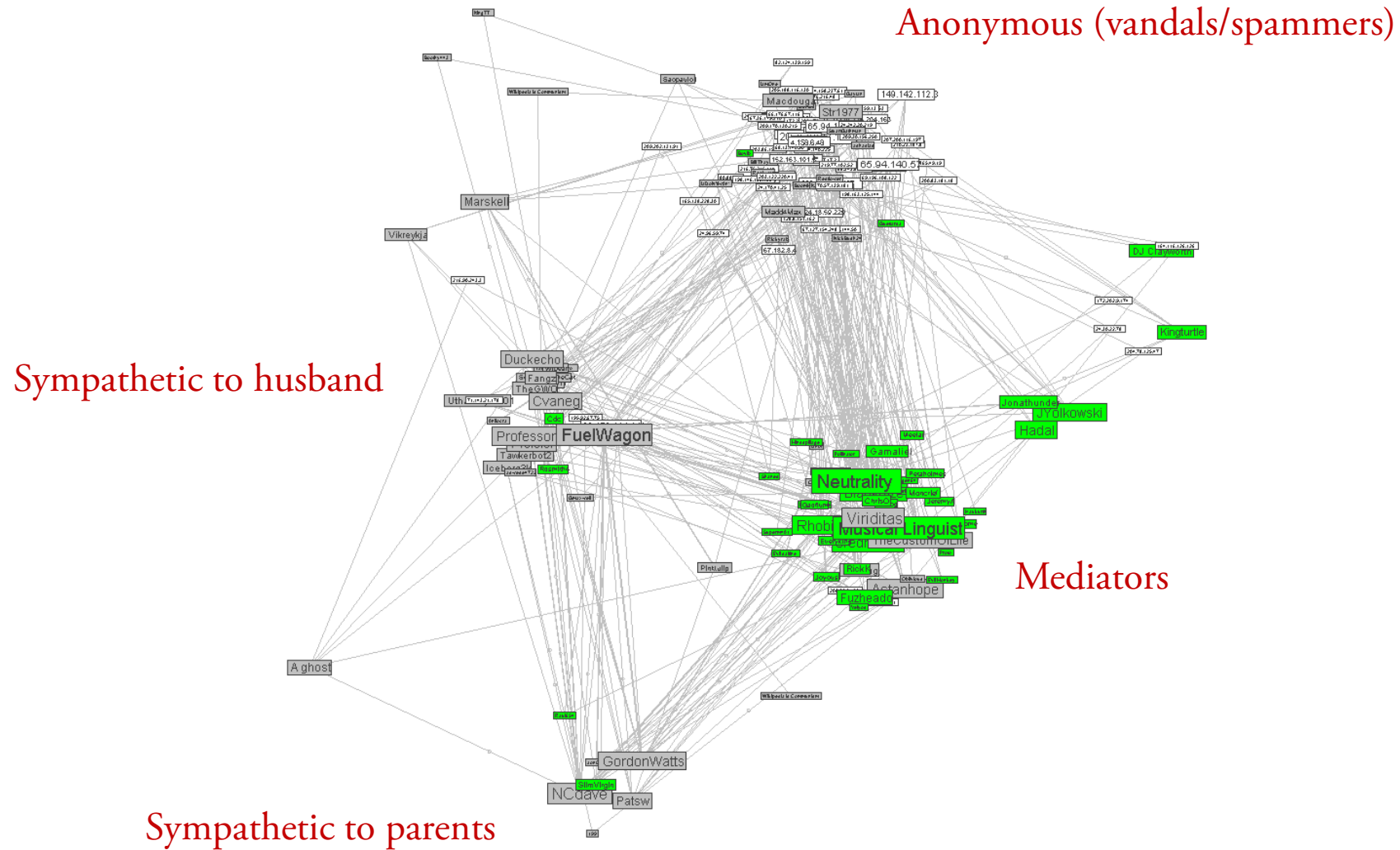
Figure 1: Trust coloring resulting from an attempt to modify the spelling of the Danish Prime Minister's last name, from Fogh, to Fjogh (in Danish, a fjog is a fool). The text background is a shade of orange that is the darker, the lower the trust of the text. The sequence consists of two consecutive revisions. Notice how the trust coloring highlights the information that has not yet been sufficiently reviewed. Subtle changes such as the above can be hard for Wikipedia visitors to spot without the help of a trust coloring.

Conflict at the user level

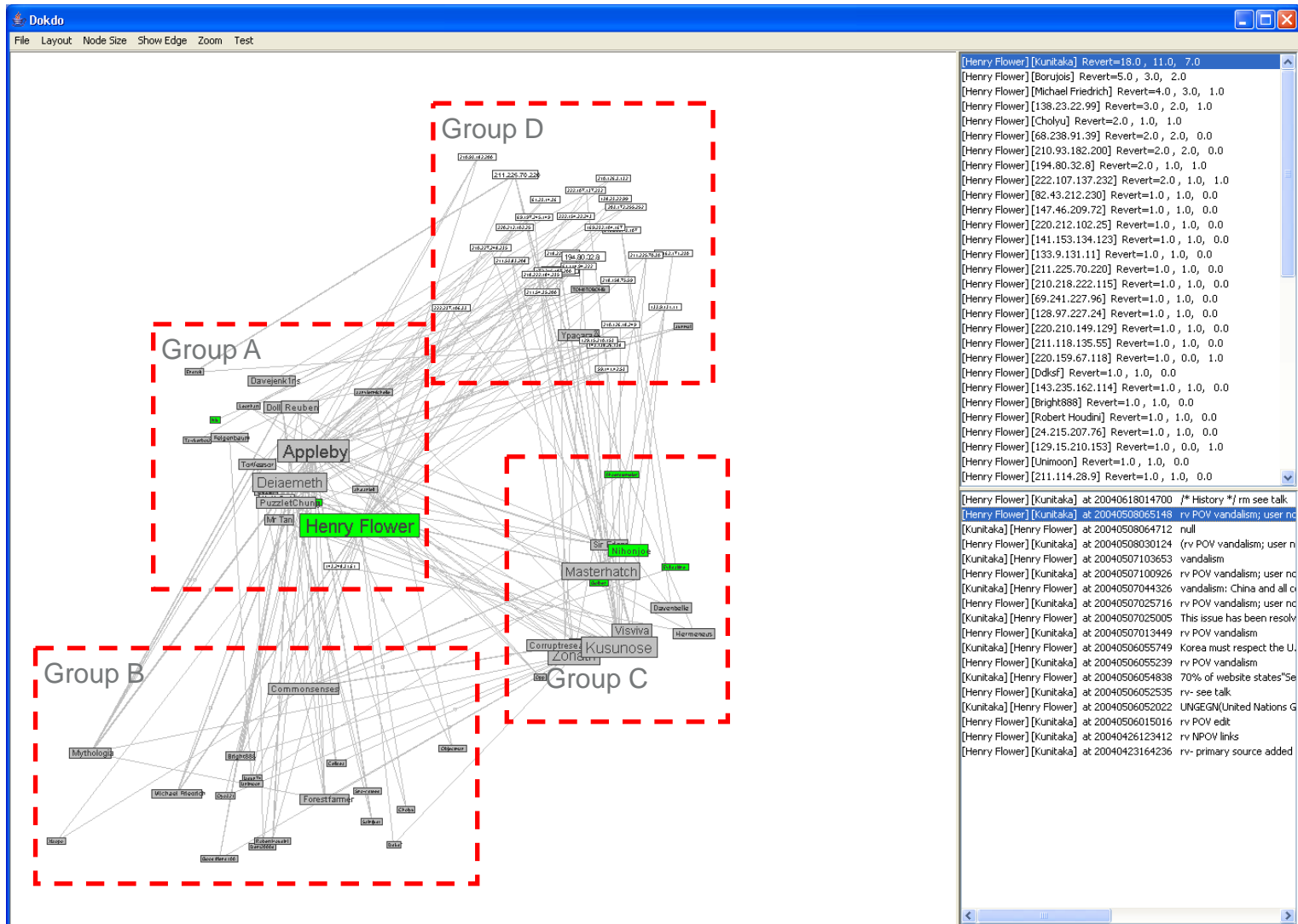
- How can we identify conflict between users?



Terry Schiavo



Dokdo/Takeshima opinion groups



Predicting conflict

- Metrics for identifying conflict in articles
 - Easily computable, scalable

Metric type	Page Type
Revisions (#)	Article, talk, article/talk
Page length	Article, talk, article/talk
Unique editors	Article, talk, article/talk
Unique editors / revisions	Article, talk
Links from other articles	Article, talk
Links to other articles	Article, talk
Anonymous edits (#, %)	Article, talk
Administrator edits (#, %)	Article, talk
Minor edits (#, %)	Article, talk
Reverts (#, by unique editors)	Article

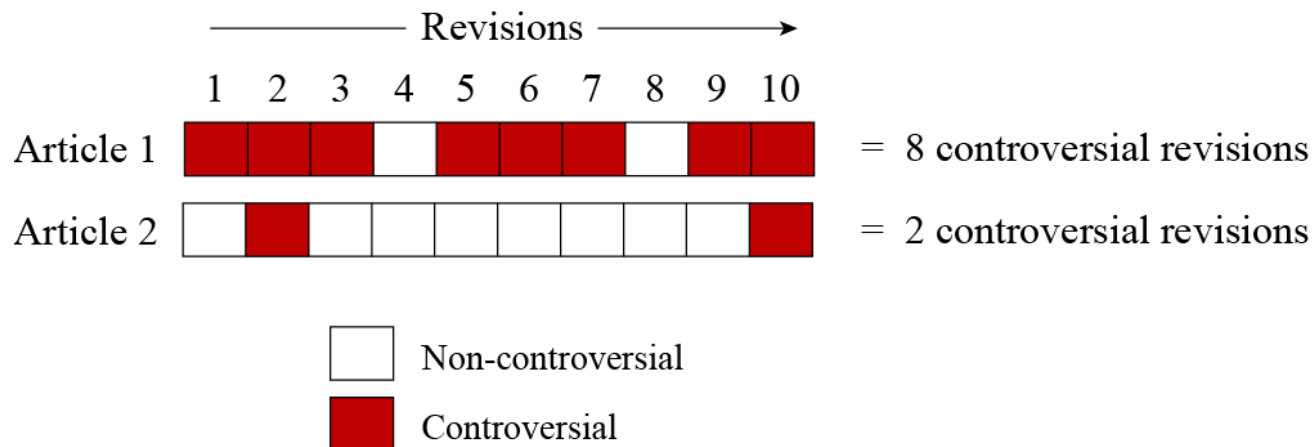
Defining conflict

- Operational definition for conflict
- Revisions tagged controversial



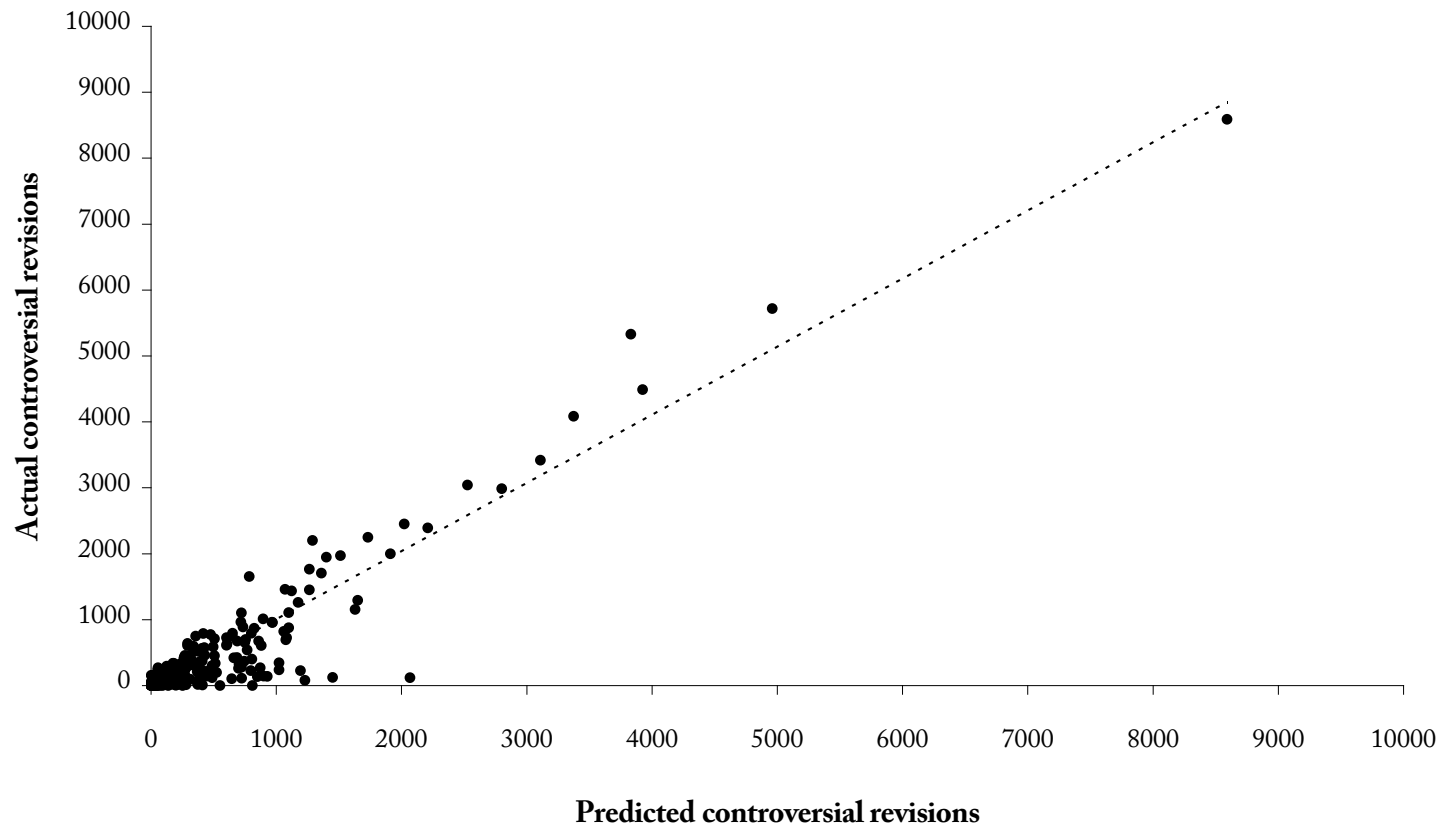
This is a **controversial** topic, which may be under **dispute**.
Please read this talk page and discuss substantial changes here before making them.
Make sure you supply full **citations** when adding information to highly controversial articles.

- Conflict revision count



Performance: Cross-validation

- 5x cross-validation, $R^2 = 0.897$



Performance: Generalization

- Detect conflict for unlabeled articles
 - Applied model to untagged articles
 - 28 articles rated by 13 expert Wikipedians
 - Significant positive correlation with predicted scores ($p < 0.013$)



Determinants of conflict

Highly weighted metrics of conflict model:

- ↑ Revisions (talk)
- ↑ Minor edits (talk)
- ↓ Unique editors (talk)
- ↑ Revisions (article)
- ↓ Unique editors (article)
- ↑ Anonymous edits (talk)
- ↓ Anonymous edits (article)

Trust

- Numerous studies surface trust-relevant information
 - Editors [Adler & Alfaro, 2007; Dondio et al., 2006; Zeng et al., 2006]
 - Stability [Suh et al., 2008]
 - Conflict [Kittur et al., 2007; Viegas et al., 2004]
- But how much impact can this have on user perceptions in a system which is inherently mutable?

What would make you trust Wikipedia more?

What would make you trust Wikipedia more?

Hypotheses

1. Visualization will impact perceptions of trust
2. Compared to baseline, visualization will impact trust both positively and negatively
3. Visualization should have most impact when high uncertainty about article
 - Low quality
 - High controversy

Design

- 3 x 2 x 2 design

Visualization


- High trust
- Low trust
- Baseline (none)

Controversial	Uncontroversial	
Abortion George Bush	Volcano Shark	High quality
Pro-life feminism Scientology and celebrities	Disk defragmenter Beeswax	Low quality

Method

- Users recruited via Amazon's Mechanical Turk
 - 253 participants
 - 673 ratings
 - 7 cents per rating
 - Kittur, Chi, & Suh, CHI 2008: Crowdsourcing user studies

Example: High trust visualization



WIKIPEDIA
The Free Encyclopedia

navigation

- [Main page](#)
- [Contents](#)
- [Featured content](#)
- [Current events](#)
- [Random article](#)

interaction

- [About Wikipedia](#)
- [Community portal](#)
- [Recent changes](#)
- [Contact Wikipedia](#)
- [Make a donation](#)
- [Help](#)

search

toolbox

- [What links here](#)
- [Related changes](#)
- [Upload file](#)
- [Special pages](#)
- [Printable version](#)

[Sign in](#) / [create account](#)

[article](#) [discussion](#) [edit this page](#) [history](#)

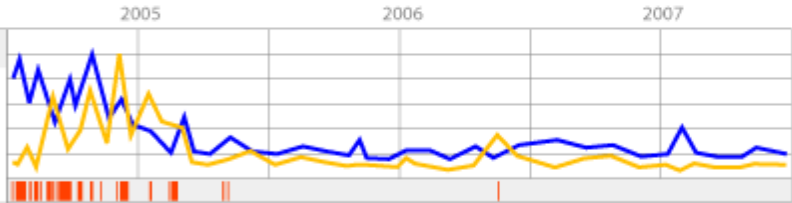
Your continued donations keep Wikipedia running!

Scientology and celebrities

From Wikipedia, the free encyclopedia


SUMMARY INFO

837 words (7% from anonymous)
Last change by [Alvvurus](#) (2832 edits)
Stability (words) in last:
day: 98% month: 96% year: 91%




Recruitment and endorsements by **Scientologist celebrities** have always been very important to the [Church of Scientology](#). [Scientology](#) actively recruits celebrities to promote Scientology to the public at large. Written Scientology policies describe recruiting celebrities in great detail. Scientology has had a written program governing celebrity recruitment since at least 1955, when [L. Ron Hubbard](#) created "Project Celebrity", offering rewards to Scientologists who recruited targeted celebrities.^[1] A Scientology policy letter of 1976 states that "rehabilitation of celebrities who are just beyond or just approaching their prime" enables the "rapid dissemination" of Scientology.^[2]

The Church of Scientology operates special [Celebrity Centres](#). Scientology policy governs the Celebrity Centres (the main one in Los Angeles and others in Paris, Nashville, and elsewhere), stating that "one of the major purposes of the Celebrity Centre and its staff is to expand the number of celebrities in Scientology." (Scientology Flag Order 2310) Another order describes Celebrity Centre's Public Clearing Division and its goal, "broad public into Scientology from celebrity dissemination"; this division has departments for planning celebrity events and routing the general public onto



A Scientology Centre on Hollywood Boulevard in Hollywood, Los Angeles, California.

Example: Low trust visualization



WIKIPEDIA
The Free Encyclopedia

navigation

- [Main page](#)
- [Contents](#)
- [Featured content](#)
- [Current events](#)
- [Random article](#)

interaction

- [About Wikipedia](#)
- [Community portal](#)
- [Recent changes](#)
- [Contact Wikipedia](#)
- [Make a donation](#)
- [Help](#)

search

toolbox

- [What links here](#)
- [Related changes](#)
- [Upload file](#)
- [Special pages](#)
- [Printable version](#)

[Sign in / create account](#)

[article](#) [discussion](#) [edit this page](#) [history](#)

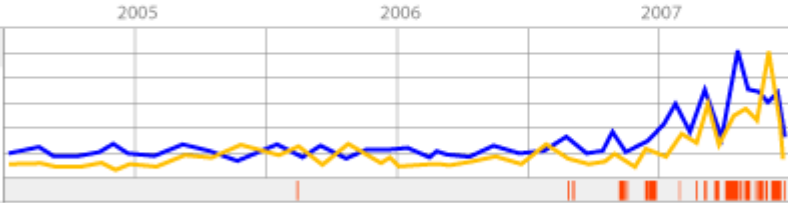
Your continued donations keep Wikipedia running!

Scientology and celebrities

From Wikipedia, the free encyclopedia

SUMMARY INFO

735 words (58% from anonymous)
Last change by [121.2.171.4](#) (1 edit)
Stability (words) in last:
day: 39% month: 32% year: 27%




Edits

- Article
- Discussion
- Reverts

Recruitment and endorsements by **Scientologist celebrities** have always been very important to the [Church of Scientology](#). [Scientology](#) actively recruits celebrities to promote Scientology to the public at large. Written Scientology policies describe recruiting celebrities in great detail. Scientology has had a written program governing celebrity recruitment since at least 1955, when [L. Ron Hubbard](#) created "Project Celebrity", offering rewards to Scientologists who recruited targeted celebrities.^[1] A Scientology policy letter of 1976 states that "rehabilitation of celebrities who are just beyond or just approaching their prime" enables the "rapid dissemination" of Scientology.^[2]

The Church of Scientology operates special [Celebrity Centres](#). Scientology policy governs the Celebrity Centres (the main one in Los Angeles and others in Paris, Nashville, and elsewhere), stating that "one of the major purposes of the Celebrity Centre and its staff is to expand the number of celebrities in Scientology." (Scientology Flag Order 2310) Another order describes Celebrity Centre's Public Clearing Division and its goal, "broad public into Scientology from celebrity dissemination"; this division has departments for planning celebrity events and routing the general public onto



A Scientology Centre on Hollywood Boulevard in Hollywood, Los Angeles, California.

Summary info: Editor

SUMMARY INFO

6731 words (46% from anonymous)

Last change by 121.24.8.218 (1 edit)

Stability (words) in last:

day: 38% month: 34% year: 26%

- % from anonymous users

SUMMARY INFO

6731 words (3% from anonymous)

Last change by valer39 (3422 edits)

Stability (words) in last:

day: 94% month: 91% year: 90%

Summary info: Editor

SUMMARY INFO

6731 words (46% from anonymous)

Last change by 121.24.8.218 (1 edit)

Stability (words) in last:

day: 38% month: 34% year: 26%

SUMMARY INFO

6731 words (3% from anonymous)

Last change by valer39 (3422 edits)

Stability (words) in last:

day: 94% month: 91% year: 90%

- % from anonymous users
- Last change by anonymous or established user

Summary info: Stability

SUMMARY INFO

6731 words (46% from anonymous)

Last change by 121.24.8.218 (1 edit)

Stability (words) in last:

day: 38% month: 34% year: 26%

- Stability of words

SUMMARY INFO

6731 words (3% from anonymous)

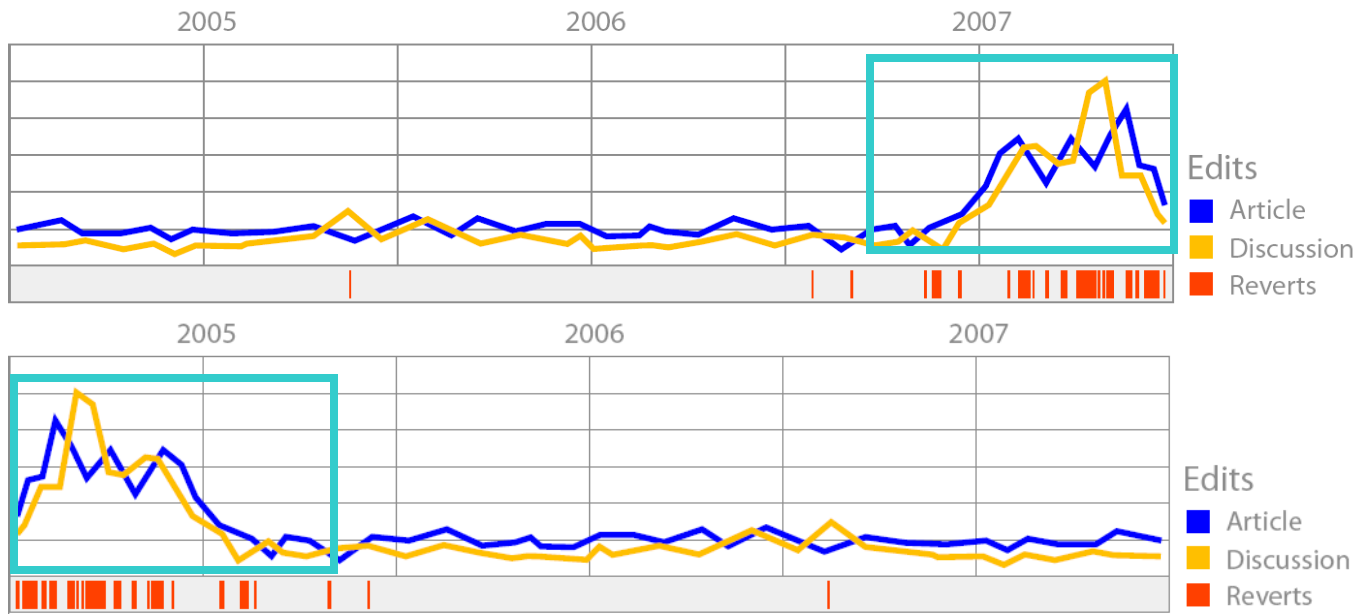
Last change by valer39 (3422 edits)

Stability (words) in last:

day: 94% month: 91% year: 90%

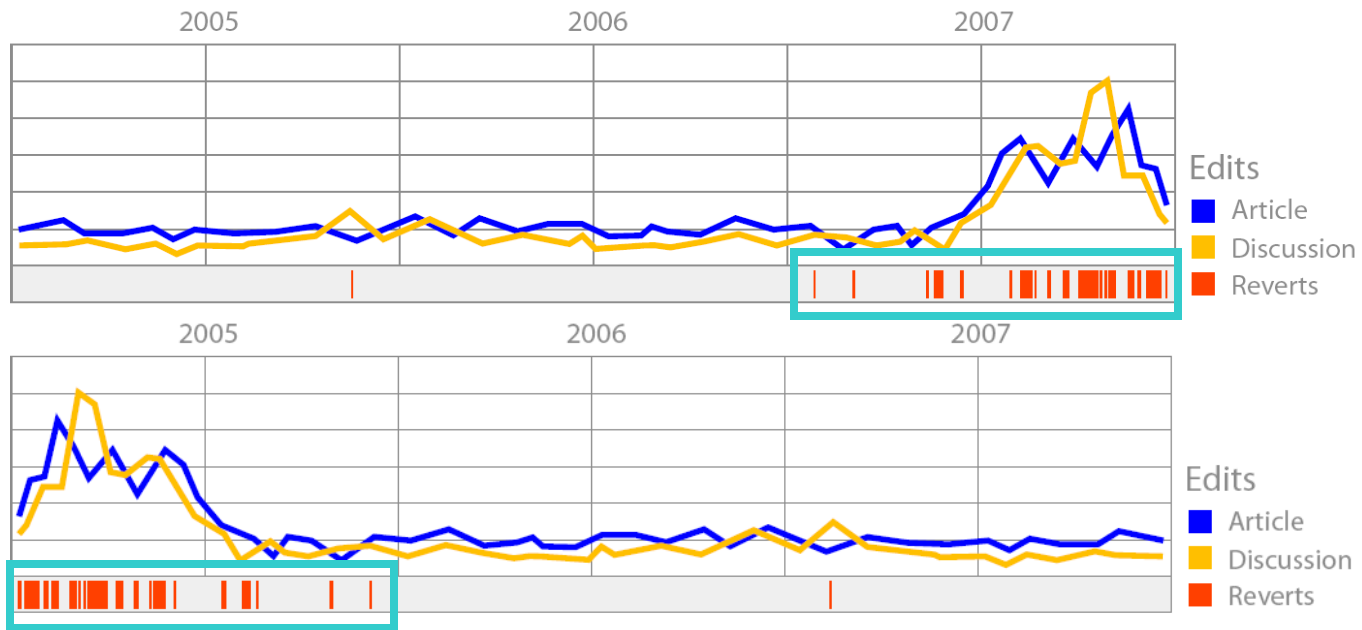
Summary info: Stability

- Instability



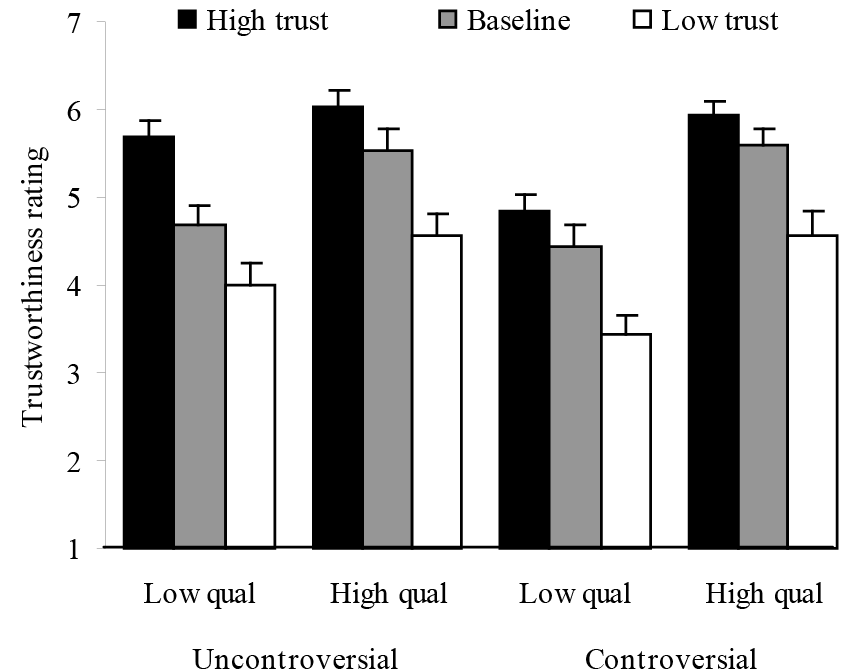
Summary info: Conflict

- Instability
- Conflict



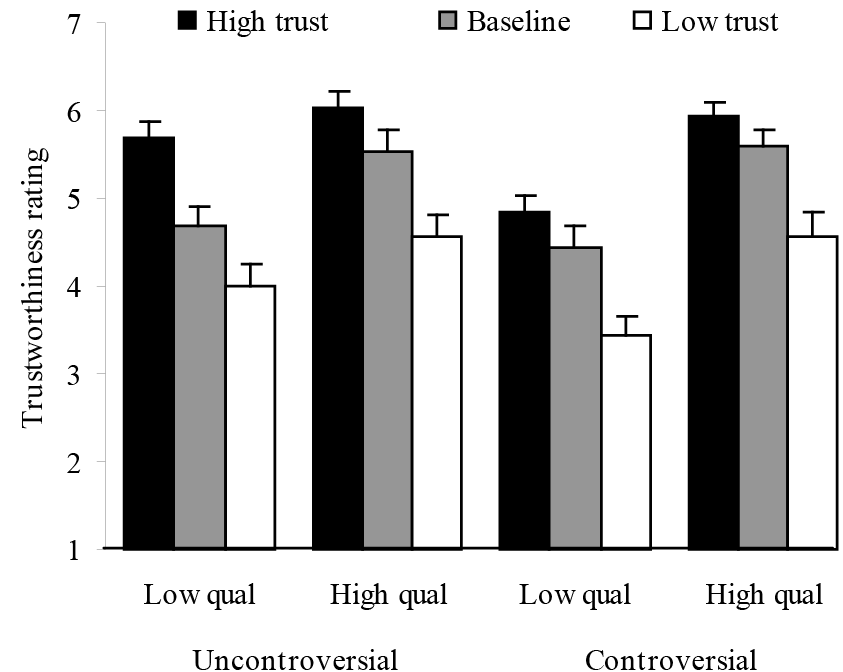
Results

1. Significant effect of visualization
 - High > low, $p < .001$
2. Both positive and negative effects
 - High > baseline, $p < .001$
 - Low < baseline, $p < .01$
3. No effect of article uncertainty
 - No interaction of visualization with either quality or controversy
 - Robust across conditions



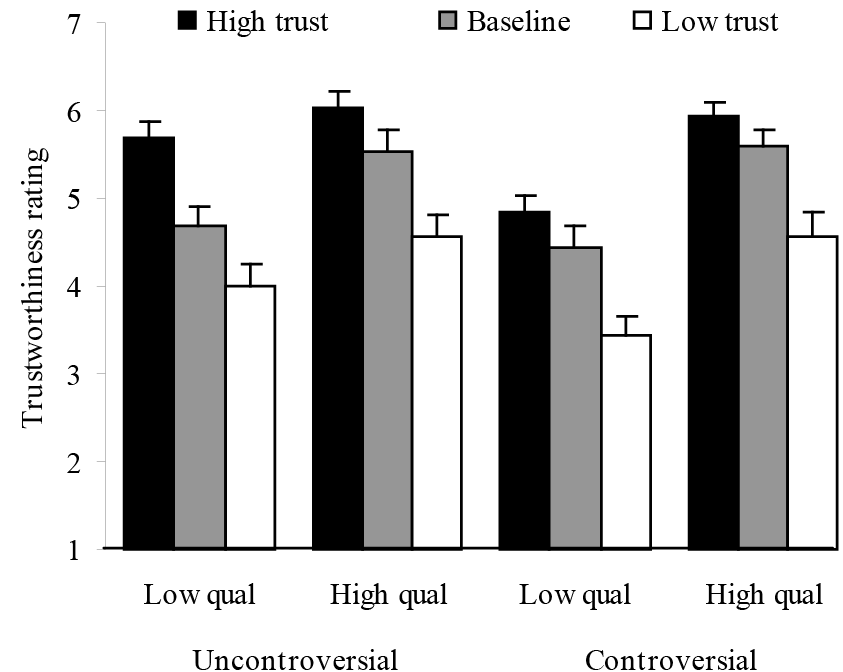
Results

1. Significant effect of visualization
 - High > low, $p < .001$
2. Both positive and negative effects
 - High > baseline, $p < .001$
 - Low < baseline, $p < .01$
3. No effect of article uncertainty
 - No interaction of visualization with either quality or controversy
 - Robust across conditions



Results

1. Significant effect of visualization
 - High > low, $p < .001$
2. Both positive and negative effects
 - High > baseline, $p < .001$
 - Low < baseline, $p < .01$
3. No effect of article uncertainty
 - No interaction of visualization with either quality or controversy
 - Robust across conditions



Future directions

- Generalizing across communities
- Supporting insight and discovery

Very many different communities

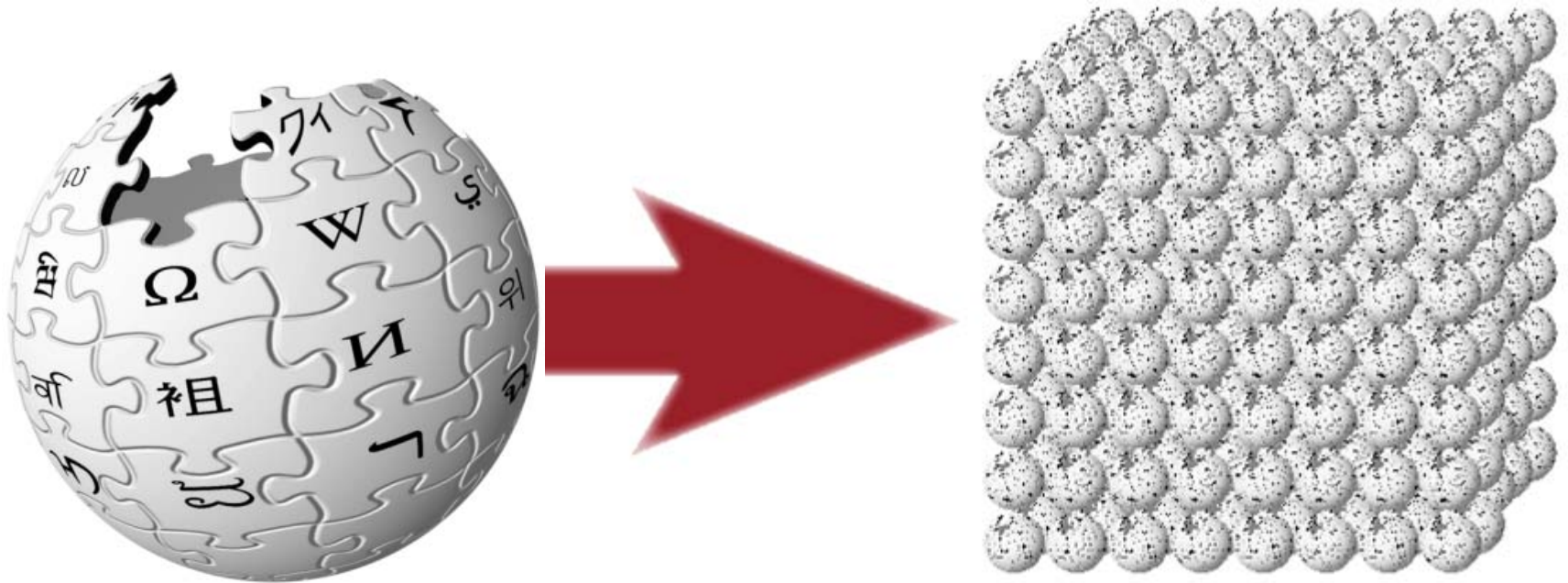


Why is it hard to compare?

- Different technologies, norms, populations
- Easy to ignore important cultural differences
- Multiple changing factors
 - Often hundreds of difficult to compare differences

Comparison: hard but important

- If we want to move towards science of online communities we must *generalize* beyond individual communities



The case for generalization

- Common coordination problems as groups grow
 - Informal norms codified and enforced
 - Informal control centralized and formalized
 - Subgroup formation
 - Increased coordination costs
 - Increased conflict

Beyond Wikipedia



Log in/ Create an account

Find a Wiki

Entertainment ▾ Gaming ▾ Lifestyle ▾ Q&A ▾ Community ▾



Welcome to Wikia
Collaborate with people who love what you love.

Get started today!

What's Hot On 

**Haiti**

**Recipes**

**American Girl Dolls**

**Zoids**

**Lego Atlantis**

what'syours?



created in the last hour 168,220 edit

**470** editors

38 editors

27 editors

15 editors

11 editors

LA X, Parts 1 & 2/Theories from Lostpedia

Mass Effect 2 Guide from Mass Effect Wiki

Victory Pose from Final Fantasy Wiki

Victory of the Daleks from Tardis

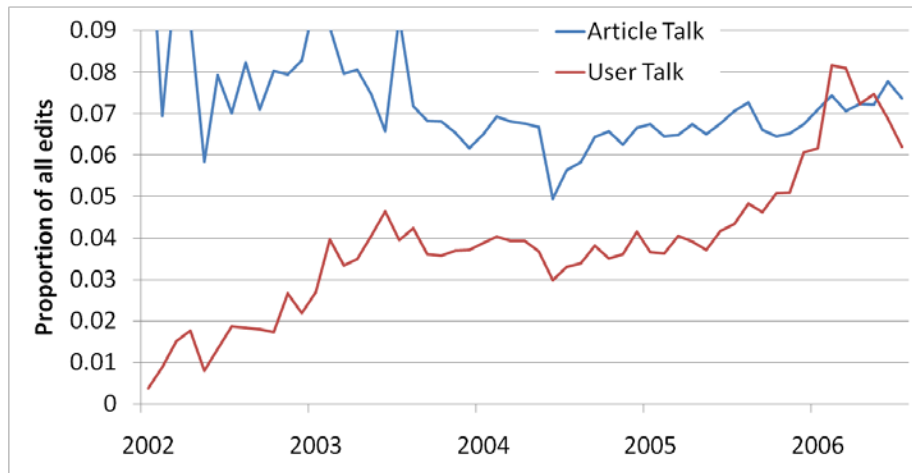
2010 from Brickipedia

Data

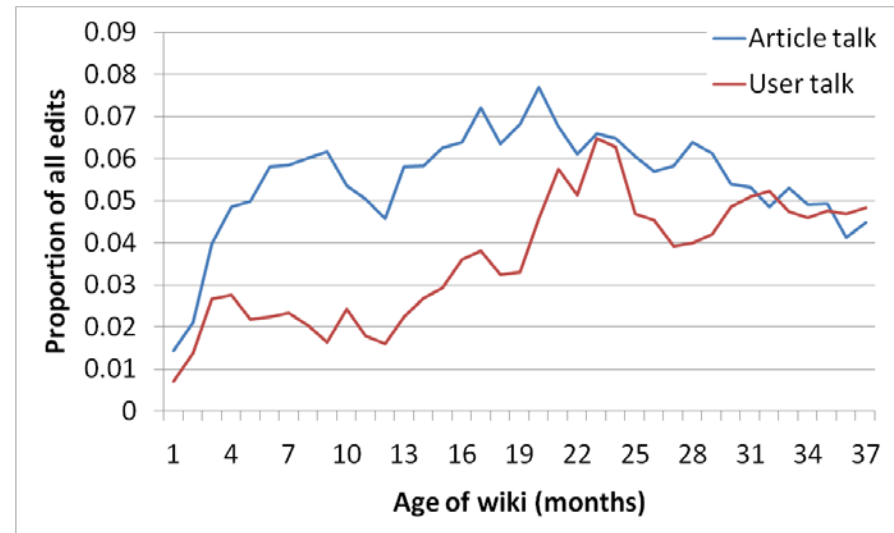
- 6811 wikis
 - Full revision histories to November 2007
 - 5M total pages
 - 23M revisions
 - 1M editors
- All based on same MediaWiki platform as Wikipedia
 - Same namespace use (mostly)
 - Many norms, policies, and conventions carried over from Wikipedia

Communication

- Similar patterns of communication growth
- User talks grows later and keeps growing



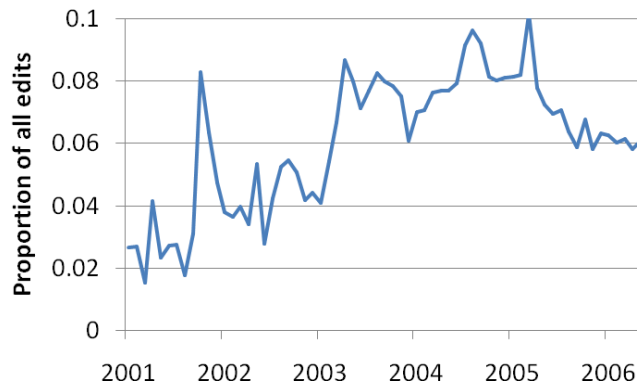
Wikipedia



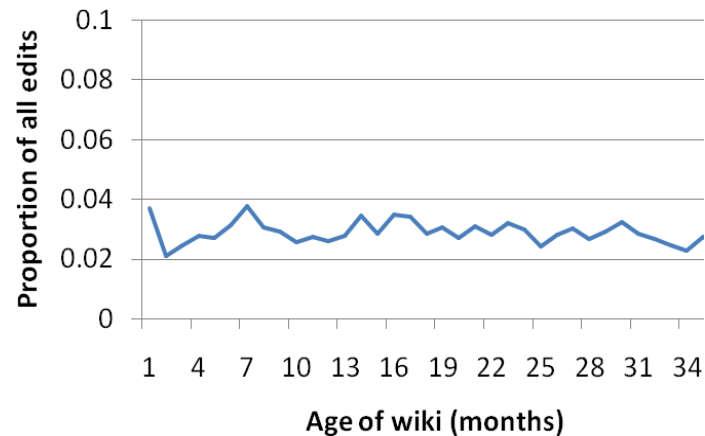
All wikis

Policies, procedures, and groups

- Much higher growth for Wikipedia
 - Even comparing to just wikis over 5k edits



Wikipedia



All wikis

From observation to intervention

- Observation is not enough
- Need experiments/interventions to establish causality
- But difficult to intervene on existing communities
 - Ethical reasons
 - Practical reasons

“Artificial” online communities

- Create ad-hoc online communities to test theories of coordination
 - Recruited from Amazon’s Mechanical Turk and other platforms
- Bring people together to collaborate on specific tasks
- Manipulate coordination mechanisms and group characteristics
 - Common policies + procedures
 - Direct communication
 - Degree of group turnover
 - Group size

Example: Michael Phelps

The screenshot shows a web browser window titled "phelps" displaying an EtherPad interface. The browser's address bar shows the URL: <http://etherpad.com/ep/pad/view/TTopqv51uQ/pdYmP3mGgr>. The EtherPad interface includes a top navigation bar with "File", "Edit", "View", "Window", and "Help" menus. Below this is a toolbar with various editing tools. The main content area is titled "Public Pad" and displays the following text:

CMU Collaborative Document Editing HT.

Special Rules:
Length: Aim for a length of 800 words. Copy and paste the content (not the citations) into a word count tool, such as <http://www.wordcounttool.com>.

EDIT CONTENT BELOW THIS LINE
=====

Michael Phelps is considered to be the greatest athlete of all times in the sport of swimming. He is most well known for winning eight Olympic gold medals in swimming during the 2008 Summer Olympic games in Beijing.

Michael Phelps was born on June 30, 1985 [1]. He is the son of Debbie and Fred Phelps, and has two older sisters, Hilary and Whitney. His sisters were both swimmers and Michael followed in their footsteps, starting his swimming career at age seven. At age 11, when it became clear that he needed better training facilities, and moved his training to North Baltimore Aquatic Club, where Bob Bowman became his new coach [4]. Phelps graduated from Towson High school in 2003 but had to delay college plans to focus on the 2004 Olympics. Bob Bowman, who he had been Michael's coach since he had first starting swimming, he was hired as the new men's swim coach for the University of Michigan. Shortly after competing in Athens, Phelps was arrested for driving under the influence of alcohol in Salisbury, Maryland [2].

[1] Michael Phelps. (2008). In Encyclopedia Britannica. Retrieved August 10, 2008, from Encyclopedia Britannica Online <http://www.britannica.com/EBchecked/topic/115571/Michael-Phelps>
[2] http://www.koonst.com/articles/Michael_Phelps_341192/part1
[3] http://en.ck12.org/wiki/Michael_Phelps
[4] http://www.ck12.com/Book/Phelps/Phelps_bk.htm

On the right side of the interface, there are sections for "View latest content", "Link to this version", "Link to read-only page", "Edit this pad", "Download as" (with options for HTML, Plain text, Microsoft Word, and PDF), and "Authors" (listing Jordan (CMU Researcher) and Doreen).

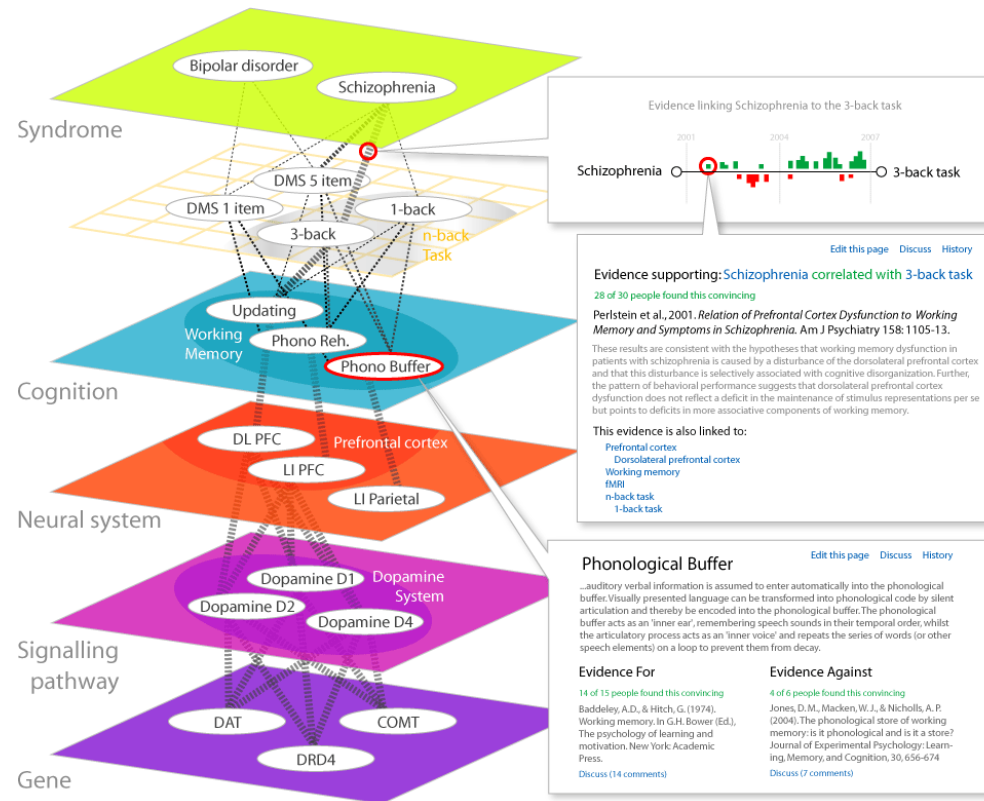
At the bottom of the browser window, there is a video player interface showing a progress bar at 00:00:03 and several control buttons (play, pause, stop, previous, next).

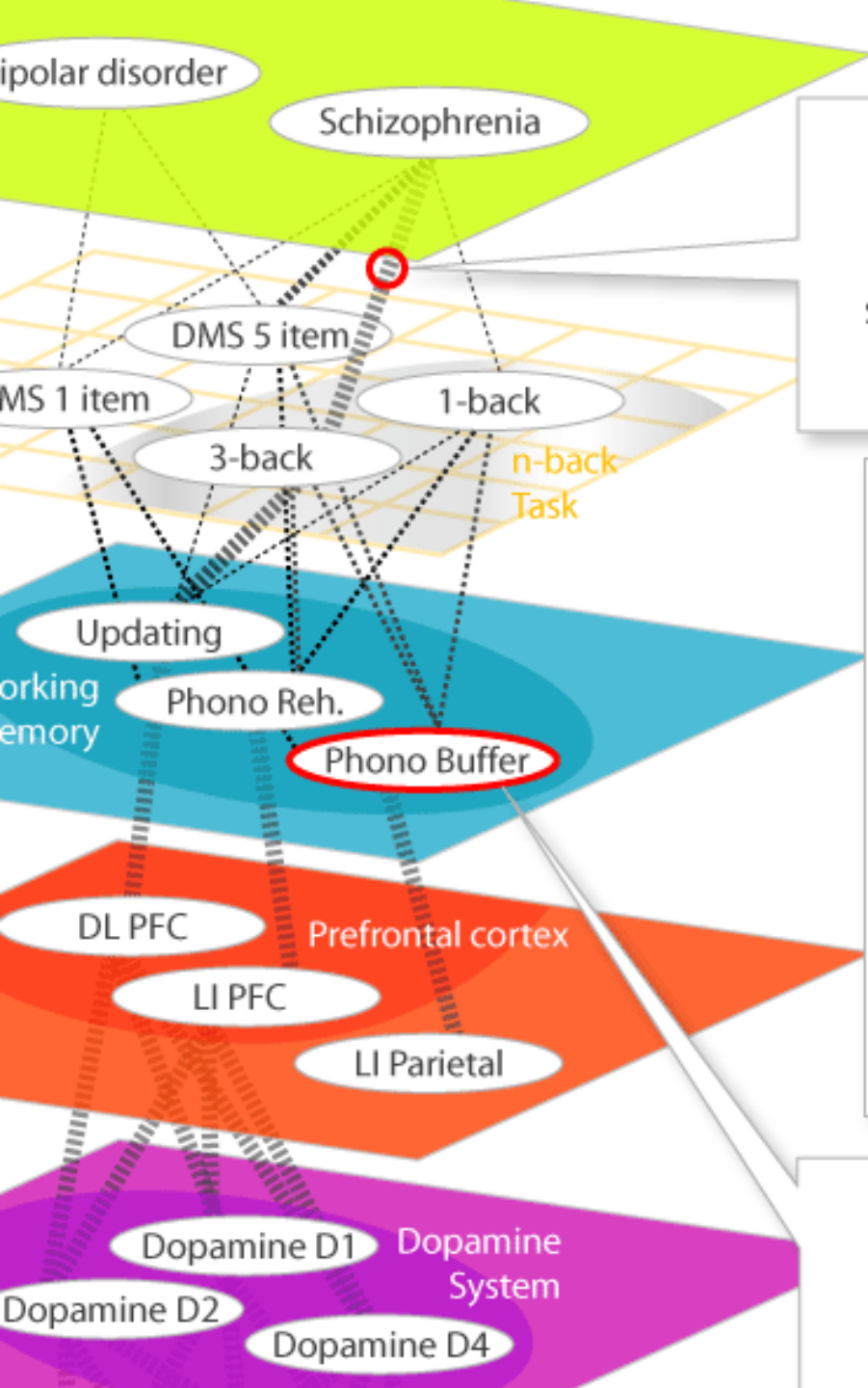
Making sense of science

- Growing fragmentation of science makes interdisciplinary insight difficult
 - Incompatible terms
 - Conflicting theories and results
 - Multiple levels of analysis
- Instead of just aggregating results, need to support discussion, consensus building, integration
- Can we use social collaboration to make sense of science?

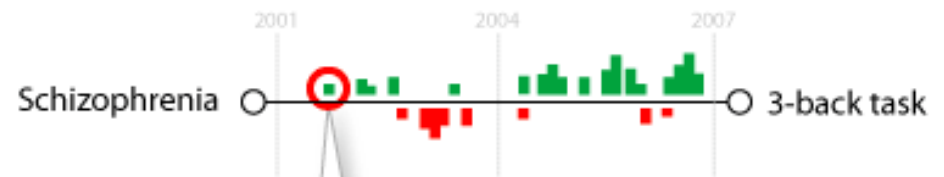
Cognitive Atlas

- Combine flexibility of wikis with structural elements of ontologies
- Enable scientific discussion and consensus building
- Visualize evidence and hypotheses
- 5 years NIH funding





Evidence linking Schizophrenia to the 3-back task



[Edit this page](#) [Discuss](#) [History](#)

Evidence supporting: Schizophrenia correlated with 3-back task

28 of 30 people found this convincing

Perlstein et al., 2001. *Relation of Prefrontal Cortex Dysfunction to Working Memory and Symptoms in Schizophrenia*. Am J Psychiatry 158: 1105-13.

These results are consistent with the hypotheses that working memory dysfunction in patients with schizophrenia is caused by a disturbance of the dorsolateral prefrontal cortex and that this disturbance is selectively associated with cognitive disorganization. Further, the pattern of behavioral performance suggests that dorsolateral prefrontal cortex dysfunction does not reflect a deficit in the maintenance of stimulus representations per se but points to deficits in more associative components of working memory.

This evidence is also linked to:

- [Prefrontal cortex](#)
- [Dorsolateral prefrontal cortex](#)
- [Working memory](#)
- [fMRI](#)
- [n-back task](#)
- [1-back task](#)

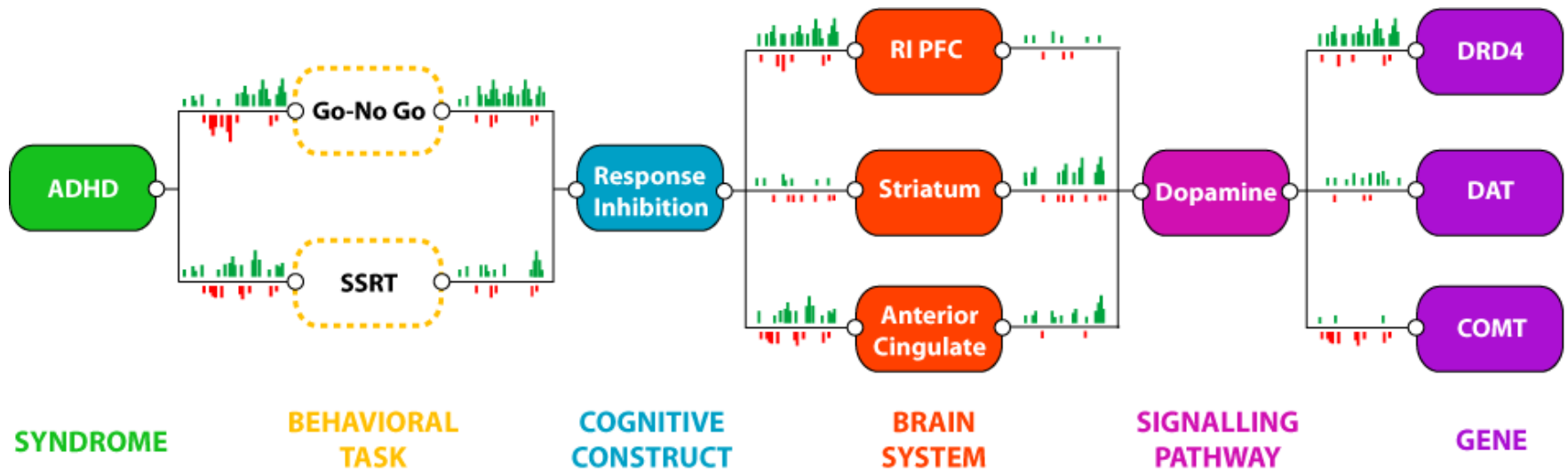
[Edit this page](#) [Discuss](#) [History](#)

Phonological Buffer

...auditory verbal information is assumed to enter automatically into the phonological buffer. Visually presented language can be transformed into phonological code by silent articulation and thereby be encoded into the phonological buffer. The phonological buffer acts as an 'inner ear', remembering speech sounds in their temporal order, whilst the articulatory process acts as an 'inner voice' and repeats the series of words (or other

Cognitive Atlas

- Visualize evidence for hypotheses spanning multiple levels of analysis



Conclusions

- Social collaboration has potential for making sense of overwhelming information
- But only if we can harness it through the appropriate kinds of coordination
- Machine learning and visualization to augment sensemaking and trust
- Towards generalized theories of collaborative production and discovery

Thanks to

Carnegie Mellon

Robert Kraut
Sara Kiesler
Jason Hong
Christos Faloutsos
Polo Chau
Bryan Pendleton
Bryant Lee
Moirra Burke
Ben Collier
Haiyi Zhu



John Levine

John Riedl
Aaron Halfaker



Keith Holyoak
John Hummel
Russ Poldrack
Bob Bjork



Ed Chi
Bongwon Suh
Bryan Pendleton
Peter Pirolli
Stuart Card



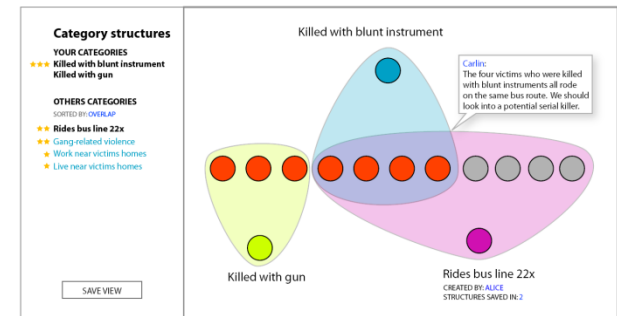
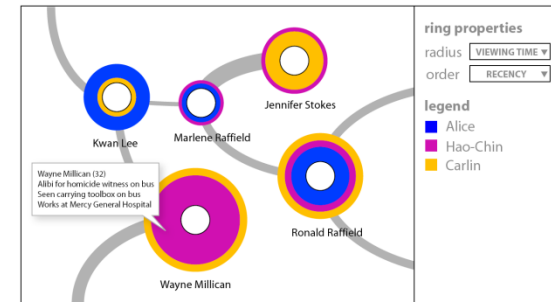
Susan Fussell

for more information:
www.kittur.org

Collaborative visualization for analysts

Scaling collaborative analysis to very many people and very much data

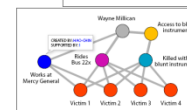
- Visualizing activity traces
- Sharing category structures
- Collaboratively building hypotheses/schemas



a. Carlin uses the tool to look for the most highly connected nodes to the four victims. Two appear: Wayne Millican and Ronald Raffield.



b. Carlin adds Wayne Millican's access to a blunt instrument to the hypothesis graph, then saves this view highlighting his multiple relations to the victims.



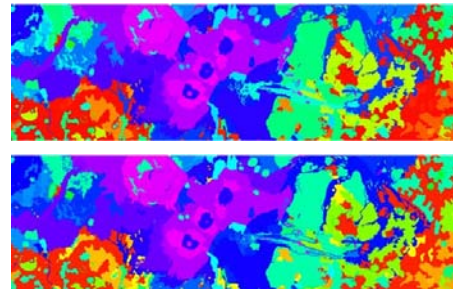
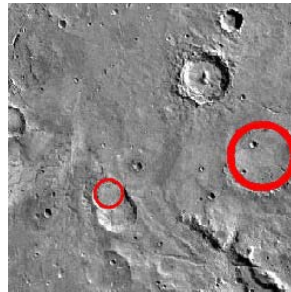
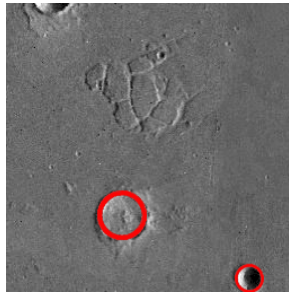
Carlin: Wayne Millican is a maintenance worker, and Hao-Chin noted that he was seen entering the 22x bus carrying a toolbox. That connects him to the victims by riding the same bus, as well as having access to a blunt instrument.

Hao-Chin: Interesting connection. It turns out Wayne also works at the same hospital as two of the victims, and leaves work around the same time as the victims got on the bus.

c. Building evidence for Carlin's hypothesis, Hao-Chin adds another connection between Wayne Millican and two of the victims.

Wisdom of crowds

- Example: NASA Clickworkers
 - 100k+ volunteers identified Mars craters from space photographs
 - Aggregate results “virtually indistinguishable” from expert geologists



experts

crowds

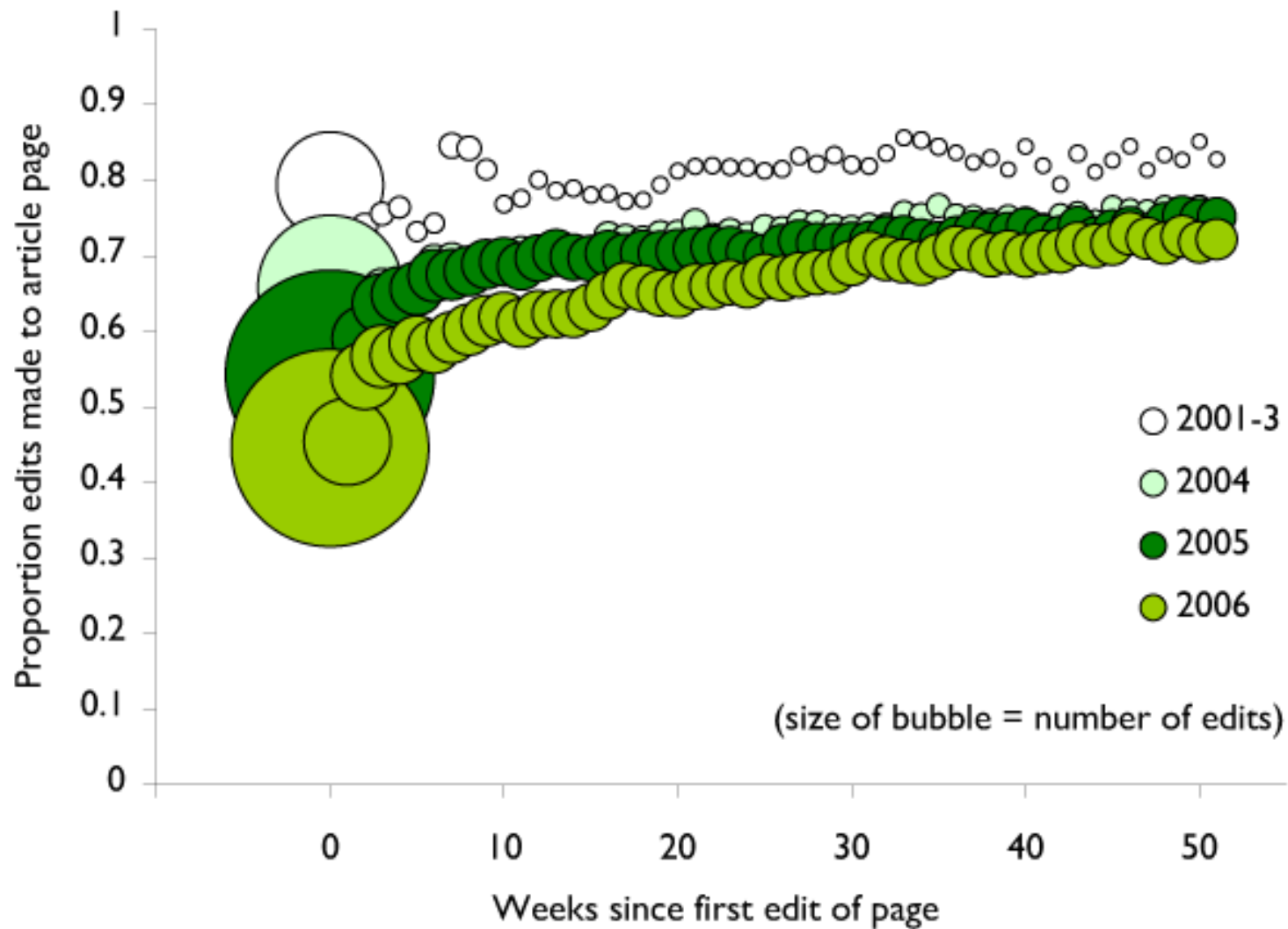
Coordination and quality

	Model 1 # Editors			Model 2 Editor Concentration			Model 3 Talk Edits			Model 4 All		
	Coef.	SE	P	Coef.	SE	P	Coef.	SE	P	Coef.	SE	P
Intercept	.442	.025	***	.329	.031	***	.362	.028	***	.304	.033	***
Initial Quality	-.149	.005	***	-.138	.005	***	-.151	.005	***	-.146	.005	***
Article Age	-.024	.004	***	-.005	.004		-.017	.004	***	-.006	.004	
# Editors	.053	.002	***	.003	.003		-.005	.003		-.020	.003	***
Editor Concentration				.791	.041	***				.600	.038	***
Editors X Concentration				.248	.017	***				.216	.020	***
Quality X Concentration				-.236	.032	***				-.222	.035	***
Age X Concentration				-.066	.027	*				-.041	.028	
# Talk Edits							.113	.005	***	.087	.004	***
Editors X Talk							-.003	.001	*	-.010	.002	***
Quality X Talk							-.012	.003	***	-.001	.003	
Age X Talk							-.009	.003	**	-.003	.003	

Table 3. Nested lagged regression analysis of the number of editors, coordination metrics (editor concentration and talk edits), and article life cycle on change in article quality.

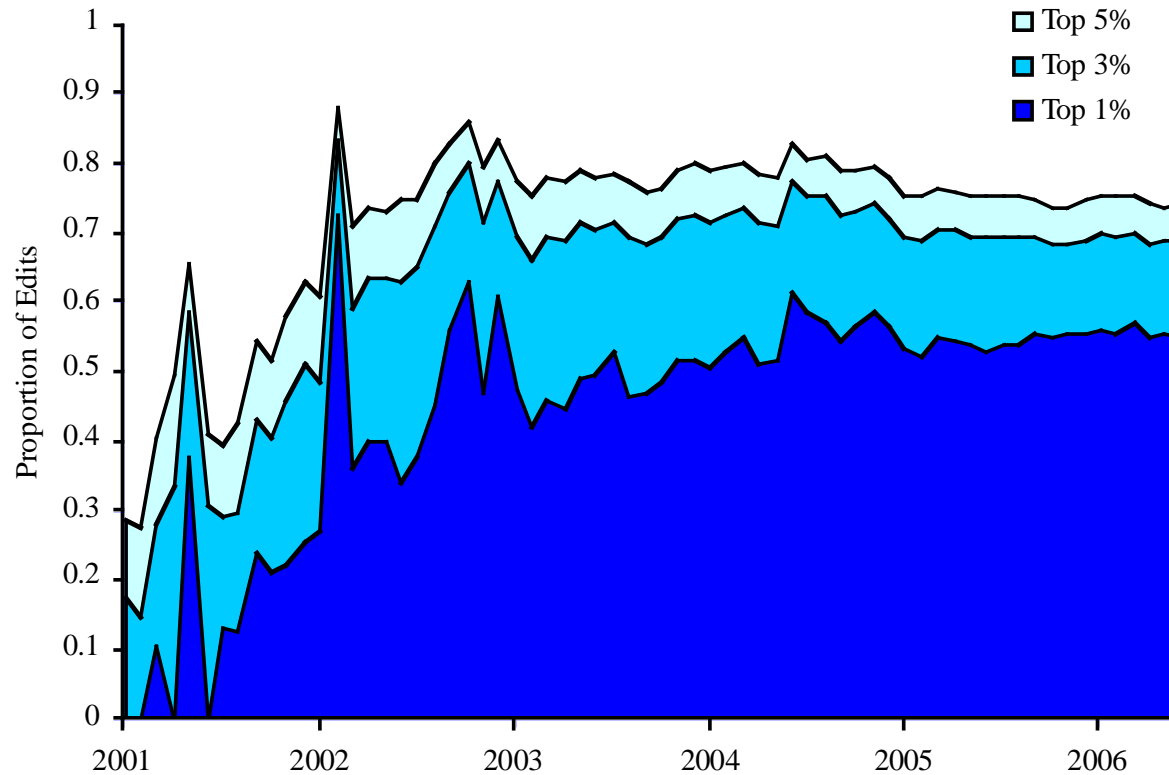
Note: *** $p < .001$, ** $p < .01$, * $p < .05$

Article lifespan

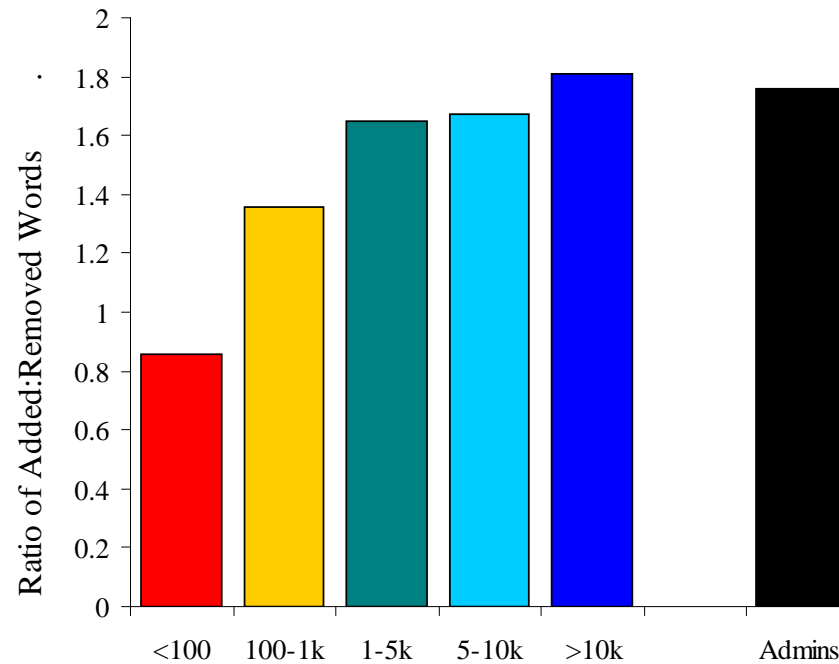


Power of the few

- Most work done by a small proportion of editors

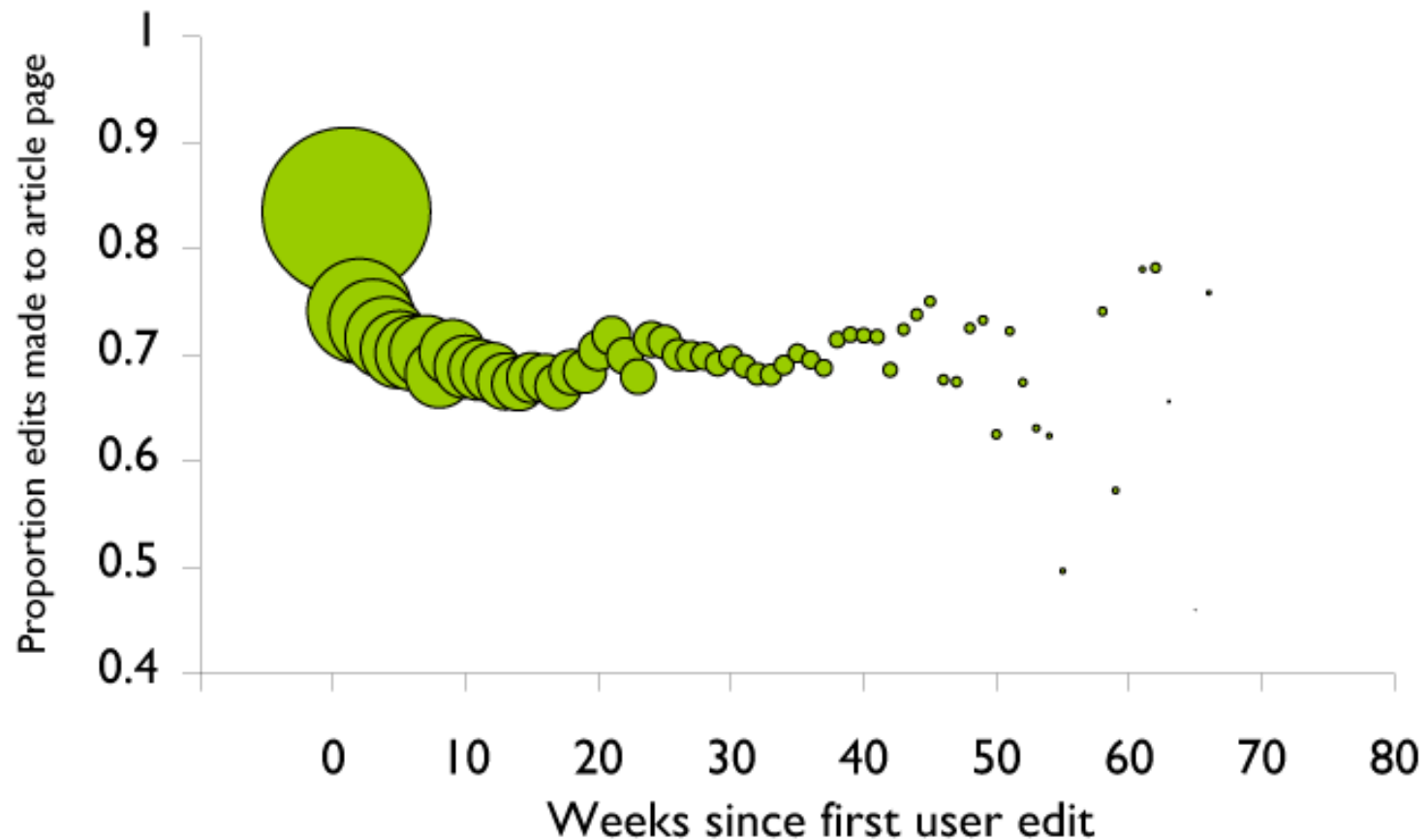


- These users do more substantial work



Coordination activity

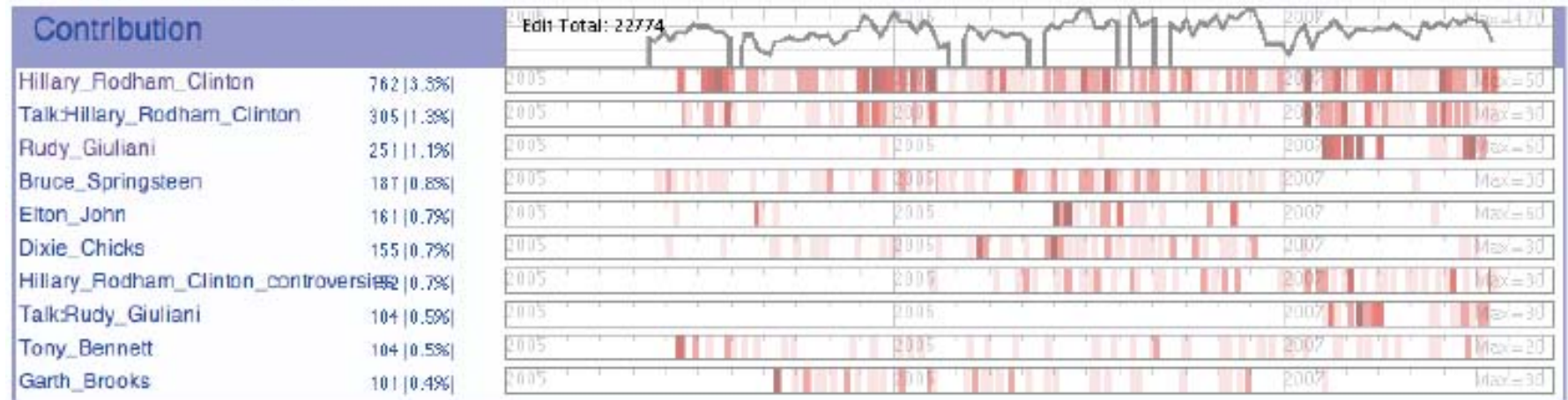
- These users do more coordination activity



WikiDashboard

User:Wasted Time R

From Wikipedia, the free encyclopedia



Defining trust

“a positive belief about the perceived reliability of, dependability of, and confidence in a person, object, or process.”

- Fogg & Tseng, 1999

Evidence-based links

[evidence](#)[discussion](#)[edit this page](#)[history](#)

[RI Prefontal Cortex](#) [\[correlated with\]](#) [Stop Signal Task](#)

Evidence For (12)

Standardized effect size = 0.4 (95% CI = [0.1 0.7], N=16)
63 out of 68 people found this convincing

Aron & Poldrack (2006). Cortical and subcortical contributions to Stop signal response inhibition: role of the subthalamic nucleus. *J Neurosci* 26:2424-33.

We propose that the role of the STN is to suppress thalamo-cortical output, thereby blocking Go response execution. These results provide convergent data for a role for the STN in Stop-signal response inhibition. They also suggest that the speed of Go and Stop processes could relate to the relative activation of different neural pathways. Future research is required to establish whether Stop-signal inhibition could be implemented via a direct functional neuroanatomic projection between IFC and STN (a "hyperdirect" pathway).

Did you find this convincing? ☒ Yes ☐ No

[Discuss \(7 comments\)](#)

Standardized effect size = 0.5 (95% CI = [0.3 0.7], N=24)
41 out of 51 people found this convincing

Chevrier et al. (2007). Dissociation of response inhibition and performance monitoring in the stop signal task using event-related fMRI. *Hum Brain Mapping* in press.

Cognitive go-phase processes included response withholding and monitoring and activated right prefrontal and midline networks. Response withdrawal additionally activated right inferior frontal gyrus and basal ganglia (caudate). Error detection invoked by failed inhibition activated dorsal anterior cingulate cortex (dACC) and right middle frontal

Evidence Against (1)

Standardized effect size = 0.3 (95% CI = [0 0.6], N=8)
18 out of 23 people found this convincing

Pliszka et al. (2006). Neuroimaging of inhibitory control areas in children with attention deficit hyperactivity disorder who were treatment naive or in long-term treatment. *Am J Psychiatry*, 163:1052-60.

In relation to comparison subjects, ADHD subjects failed to activate the anterior cingulate cortex and the left ventrolateral prefrontal cortex after unsuccessful inhibition. These findings appear in treatment-naive ADHD individuals and are unlikely to be an artifact of long-term treatment with stimulants or of abrupt termination of stimulants before imaging.

Did you find this convincing? ☒ Yes ☐ No

[Discuss \(12 comments\)](#)