

Social Analysis & Interaction

Tatiana Braescu

Seminar „Visual Analytics“ Autumn, 2007

Supervisor

Prof. Dr. Andreas Kerren

1

Purpose of the presentation

To present an overview of analysis and visualization techniques that reveal:

- “who revises whom” in Wikipedia
- “us vs. them” conflict patterns between groups of users in Wikipedia
- “who’s connected to who” on bibliographic collaboration networks

To offer insights on design considerations for asynchronous collaboration in visual analysis environments:

- works parallelization
- communication
- social organization

“Who revises Whom” in Wikipedia

Ulrik Brandes and Jürgen Lerner

Department of Computer & Information Science University of Konstanz

- controversial topics (abortion, gun control)
- delicate historic events
- important political persons

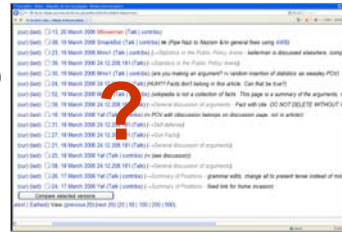


Figure 1: Small part of the revision history of the page Gun politics in Wikipedia.

A set of analysis and visualization techniques that reveal:

- The dominant authors of a page
- The roles they play and the alters they confront

Tools to understand how authors collaborate in the presence of controversy

Who revises Whom: Motivation

“Who revises Whom” in Wikipedia

First step: revision network

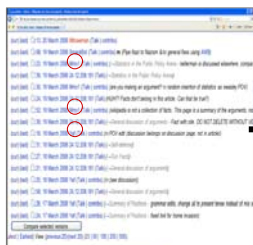


Figure 1: Small part of the revision history of the page Gun politics.

```

<page><title>Gun politics</title>
...
<revision><timestamp>2006-03-18T22:31:41Z</timestamp>
<contributor><ip>24.12.208.181</ip></contributor>
<comment>? Self-defense ?</comment>
</revision>
<revision><timestamp>2006-03-18T23:18:38Z</timestamp>
<contributor><username>Yaf</username></contributor>
<comment>rv POV edit (discussion belongs on discussion page, not in article)</comment>
</revision>
<revision><timestamp>2006-03-19T02:39:25Z</timestamp>
<contributor><ip>24.12.208.181</ip></contributor>
<comment>? General discussion of arguments ? Fact with cite. DO NOT DELETE WITHOUT VERY GOOD REASON!!!!!! Different placement on page acceptable.</comment>
</revision>
<revision><timestamp>2006-03-19T02:52:41Z</timestamp>
<contributor><username>Mmx1</username></contributor>
...
<comment>wikipedia is not a collection of facts. This page is a summary of the arguments, not a place to make ...
  
```

Figure 2: Six consecutive revisions of the page Gun politics in XML format.

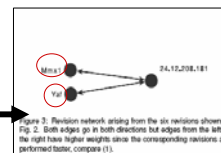


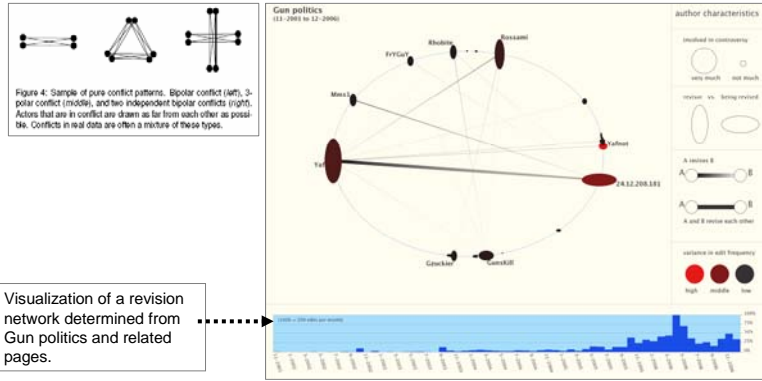
Figure 3: Revision network arising from the six revisions shown in Fig. 2. Both edges go in both directions but edges from the left to the right have higher weights since the corresponding revisions are performed faster, compare (1).

Who revises Whom: Input data

“Who revises Whom” in Wikipedia

Next step: **visual analysis and visual representation**

What Position Do They Take? How they are involve?



Who revises Whom: Visualization

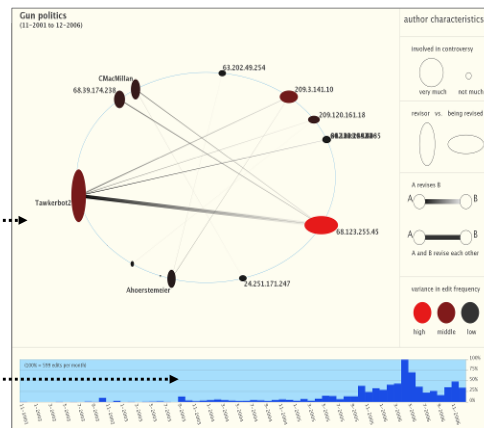
“Who revises Whom” in Wikipedia

Next step: **Filtering**

- Restriction to Time Intervals
- Restriction to Relevant Sub-networks

Network clustering reveals a relevant sub-network of the revision network of Gun politics

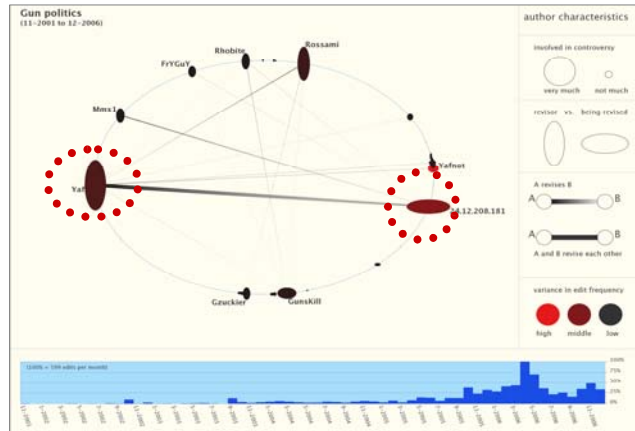
Filtering in time: a peak in the revision plot of Gun politics during 2003 has been caused by authors that vanish in the global image



Who revises Whom: Visualization

“Who revises Whom” in Wikipedia

Final step: identified recurrent patterns of confrontation



Who revises Whom: Patterns

7

“Who revises Whom” in Wikipedia

CONCLUSION

- + reveal the authors that are the most involved in controversy
- + network visualizations show “who confronts whom” and “who plays which role”
- + identified some recurrent patterns of confrontation
- + can be applied to Wikipedia articles in any language without the need for adapting NLP algorithms
- the revision network should take into account whose text has been changed during a revision
- the interpretation of the revisor vs. revised pattern can be quite different.

Who revises Whom: Conclusion

8

Purpose of the presentation

To present an overview of analysis and visualization techniques that reveal:

- “who revises whom” in Wikipedia
- “us vs. them” conflict patterns between groups of users in Wikipedia
- “who’s connected to who” on bibliographic collaboration networks.

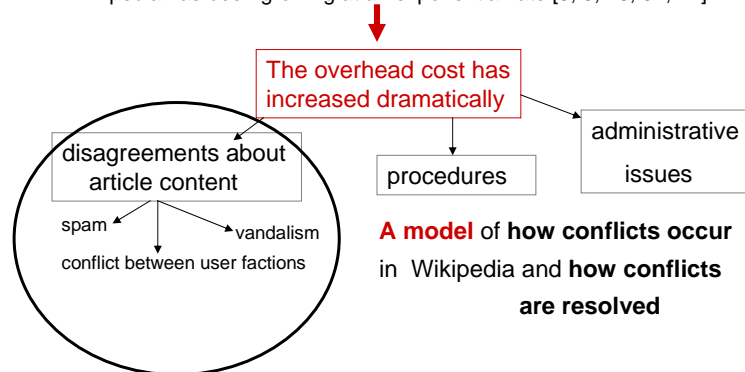
To offer insights on design considerations for asynchronous collaboration in visual analysis environments:

- works parallelization
- communication
- social organization

“Us vs. Them” conflict patterns between groups of users in Wikipedia

Bongwon Suh, Ed H. Chi, Bryan A. Pendleton and Aniket Kittur
Palo Alto Research Center

Wikipedia has been growing at an exponential rate [5, 6, 18, 32, 47]



“Us vs. Them” conflict patterns between groups of users in Wikipedia

The user conflict model: base & methods

- users' editing history
- the relationships between user edits (“reverts”)

A revert is defined as undoing the actions of another editor in whole or in part

Two different methods

Data-driven :
to generate a small fingerprint of each revision

+

User-labeled:
to capture partial reverts

provides **converging evidence on the true change in reverts over time**

Users Total	3,769,347
Users who made at least one revert	402,454
Revisions Total	58,545,791
Reverts (MD5 hash method)	3,711,638
Self-reverts	582,373
Pages with at least one revert	721,866
Pages with 50 reverts or more	9,973
Reverts (Comment method)	2,422,482
Vandalism (Comment with vandal, rvv, etc)	577,643
Reverts (Union of both methods)	3,917,008

Table 1. User, Revision, Revert and Vandalism Statistics

“Us vs. Them” conflict patterns between groups of users in Wikipedia

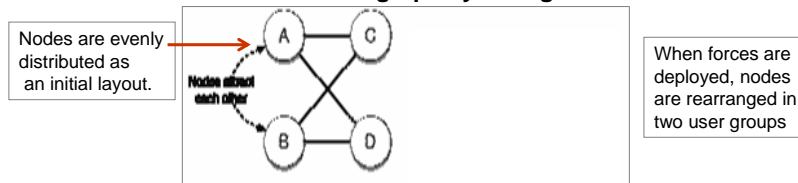
The user conflict model : design choices and layout principles

Problem using reverts to identify conflicts:

- multiple users are often involved in chains of reverts
- edit history is typically long and tedious to browse
- various types of reverts - the “revert duel”
 - the “self-reverts”
 - reverts by multiple users

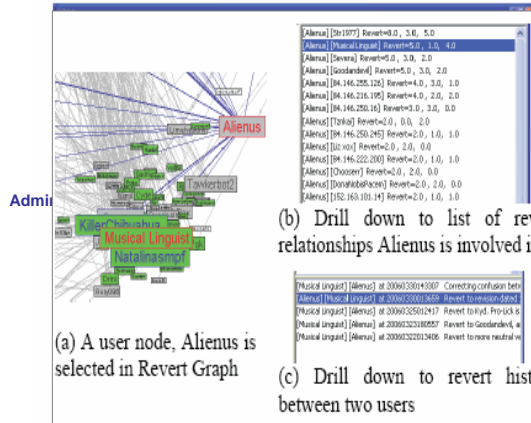
- ✓ Disregard Self Revert
- ✓ Degree of Conflict
- ✓ Conflict Group
- ✓ Identity Based Revert
- ✓ Immediate Revert Only

A force-directed graphlayout algorithm



“Us vs. Them” conflict patterns between groups of users in Wikipedia

The user conflict model → visualization tool (Revert Graph)



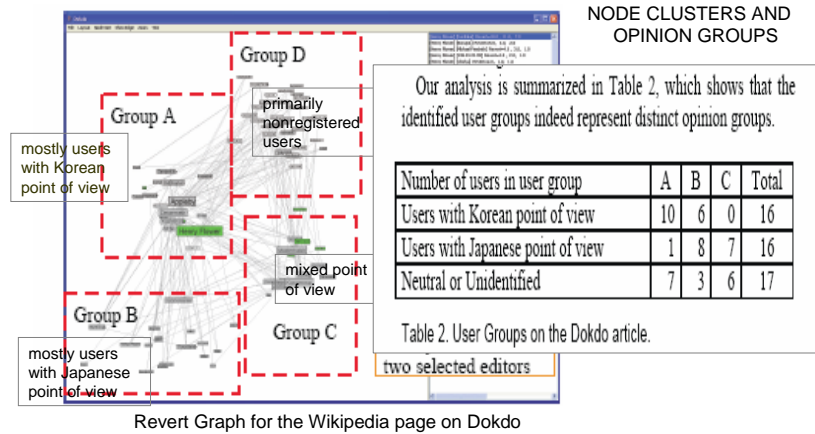
How to use the tool?

1. The user specify an article
2. The revert history of the article is retrieved from DB
3. The tool finds a specific user node participating in editing this article
4. A node-link graph is formed and displayed
5. Click on a specific relationship
6. Simulate revert forces in the graph

Us vs. Them: Revert Graph

“Us vs. Them” conflict patterns between groups of users in Wikipedia

Visualization tool (Revert Graph): user conflict patterns



Us vs. Them: Patterns

“Us vs. Them” conflict patterns between groups of users in Wikipedia

CONCLUSION

- + helps identify important social patterns in Wikipedia
- + may be applicable to other online communities

- not every aspect of social dynamics in online collaboration systems was fully addressed
- the tool cannot detect conflicts between users who were not involved in reverts

Purpose of the presentation

To present an overview of analysis and visualization techniques that reveal:

- “who revises whom” in Wikipedia
- “us vs. them” conflict patterns between groups of users in Wikipedia
- “who’s connected to who” on bibliographic collaboration networks.

To offer insights on design considerations for asynchronous collaboration in visual analysis environments:

- works parallelization
- communication
- social organization

“Who's connected to Who” on bibliographic collaboration networks

Mustafa Bilgic ,Louis Licamele ,Lise Getoor and Ben Shneiderman
University of Maryland,College Park, MD

? **The problem** = the data may inadvertently contain several distinct references to the same underlying entity or actor.

- ✓ This visual display is misleading: incorrect number of nodes & the edges and paths are inaccurate
- ✓ Calculating of the standard social network measures, would give inaccurate results.

! **The solution**= entity-resolution to identify potential duplicates
(The process of reconciling, from the underlying data references, the actual real-world entities)

automated entity resolution

hand cleaning entity resolution

“Who's connected to Who” on bibliographic collaboration networks

D-Dupe: resolve ambiguities either by merging nodes or by marking them distinct.

Cleaning large networks by focusing on a small subnetwork containing a potential duplicate pair

Two of D-Dupe's novelties are:

1. **Stable Visual Layout Optimized for Entity Resolution**
Shows only the subnetwork relevant for the entity resolution task.
Allows visualization to scale to large networks
A stable substrate-the potential duplicates and other related entities always appear at the same location
2. **User Control for Combining Entity Resolution Algorithms**
Numerous similarity measures can be used to determine potential duplicates
Allows users to flexibly apply and interleave different measures

“Who’s connected to Who” on bibliographic collaboration networks

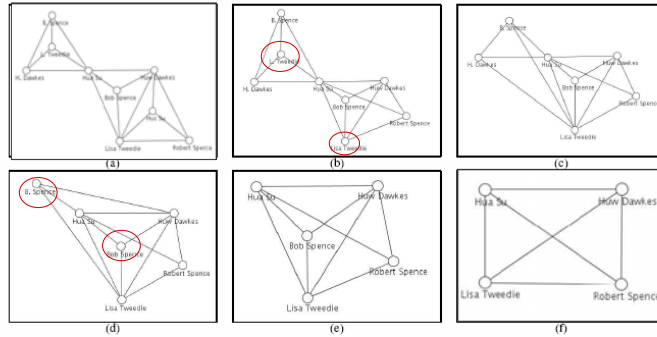
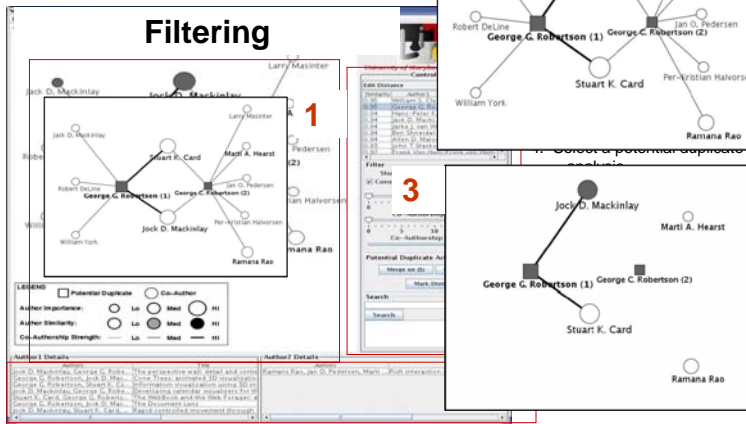


Figure 1 gives an overview of the deduplication process on a small portion of bibliographic dataset used for the InfoVis 2004 Contest

“Who’s connected to Who”: Conclusion

“Who’s connected to Who” on bibliographic collaboration networks

D-Dupe: interface



dataset
possible
thms
ing to
plicates.
pair for
context
ply
des are:
ation:
[links/dd](#)

“Who’s connected to Who”: Conclusion

“Who's connected to Who” on bibliographic collaboration networks

CONCLUSION

- + D-Dupe's layout and interaction principles can be used in other social networks
 - + Use an interface which effectively combines visual and analytic information for data cleaning in an interactive tool.
- The actors should have properties that can be used by the attribute similarity functions.

Purpose of the presentation

To present an overview of analysis and visualization techniques that reveal:

- “who revises whom” in Wikipedia
- “us vs. them” conflict patterns between groups of users in Wikipedia
- “who's connected to who” on bibliographic collaboration networks.

Design considerations for asynchronous collaboration in visual analysis environments:

- works parallelization
- communication
- social organization

Design considerations for asynchronous collaboration

Jeffrey Heer, Maneesh Agrawala

University of California, Berkeley

Premise: to fully support sensemaking, interactive visualization should also support social interaction

The problem

collaboration mechanisms for supporting social interaction are not immediately clear

- How should collaboration be structured?
- What shared artifacts can be used to coordinate contributions?
- What are the most effective communication mechanisms?

Design consideration: Motivation

23

Design considerations for asynchronous collaboration

A set of design considerations

1. Division and allocation of work
2. Common ground and awareness
3. Reference and deixis
4. Incentives and engagement
5. Identity, trust, and reputation
6. Group dynamics
7. Consensus and decision making

Consensus and discussion

voting or ranking systems

prediction markets : individuals can be given a limited amount of points or currency

Information distribution and presentation

discussion

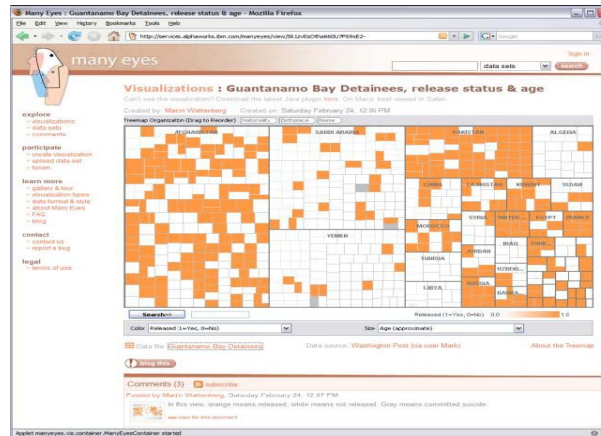
Design consideration: key issues

24

Design considerations for asynchronous collaboration

Asynchronous Collaborative Visualization Systems

ManyEyes



Design consideration: Visualization Systems

25

Design considerations for asynchronous collaboration

CONCLUSION

By partitioning work across both time and space, asynchronous collaboration offers greater scalability for group-oriented analysis

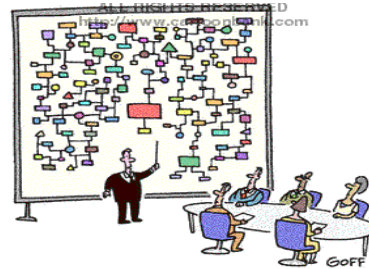
Design consideration: Conclusion

26

Social Analysis & Interaction

Final

Why we need these tools?



"And that's why we need a computer."

Human existence **depends on collaborative problem solving.**