INSTITUTE OF VISUAL COMPUTING

TU Graz

SCIENCE
PASSION
TECHNOLOGY

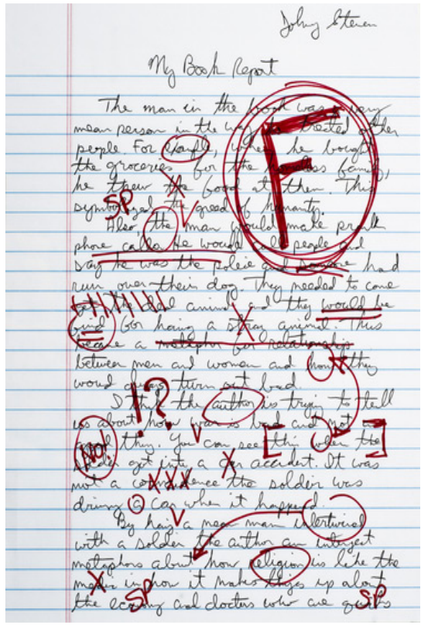
Journal Club: Time Curves

Julian Rakuschek

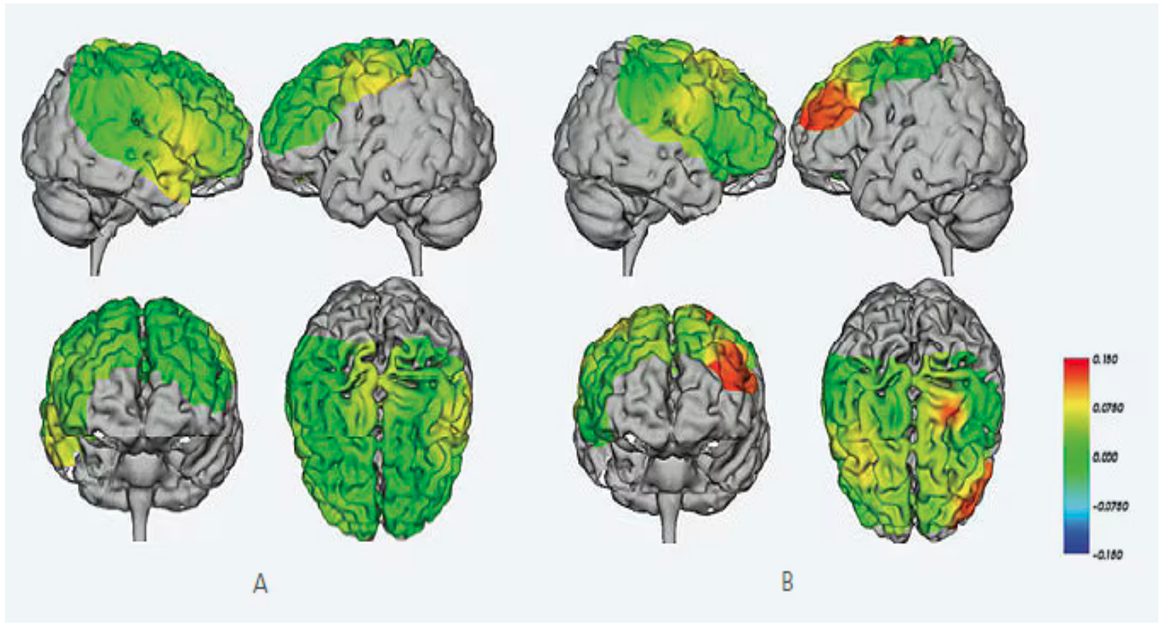
23.01.2025

Motivation

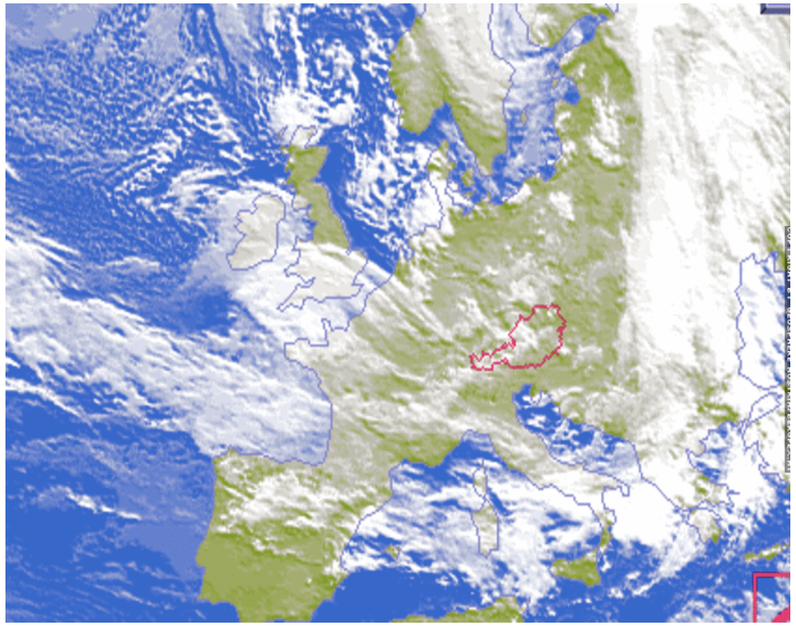
Temporal Patterns



Article Revision



Brain Activity



Climate

Domain specific visualizations exist

CHI 2004 | Paper 24-29 April | Vienna, Austria

Studying Cooperation and Conflict between Authors with *history flow* Visualizations

Fernanda B. Viégas
MIT Media Lab
Cambridge, MA 02139 USA
fviegas@media.mit.edu

Martin Wattenberg
IBM Research
Cambridge, MA 02142 USA
mwatten@us.ibm.com

Kushal Dave
IBM Research
Cambridge, MA 02142 USA
kdave@us.ibm.com

ABSTRACT
The Internet has fostered an unconventional and powerful style of collaboration: "wiki" web sites, where every visitor has the power to become an editor. In this paper we investigate the dynamics of Wikipedia, a prominent, thriving wiki. We make three contributions. First, we introduce a new exploratory data analysis tool, the *history flow* visualization, which is effective in revealing patterns within the wiki context and which we believe will be useful in other collaborative situations as well. Second, we discuss several collaboration patterns highlighted by this visualization tool and corroborate them with statistical analysis. Third, we discuss the implications of these patterns for the design and governance of online collaborative social spaces. We focus on the relevance of authorship, the value of community surveillance in ameliorating antisocial behavior, and how authors with competing perspectives negotiate their differences.

Categories & Subject Descriptors
H.5.2: GUI; H.5.3:n Collaborative computing, Web-based interaction.

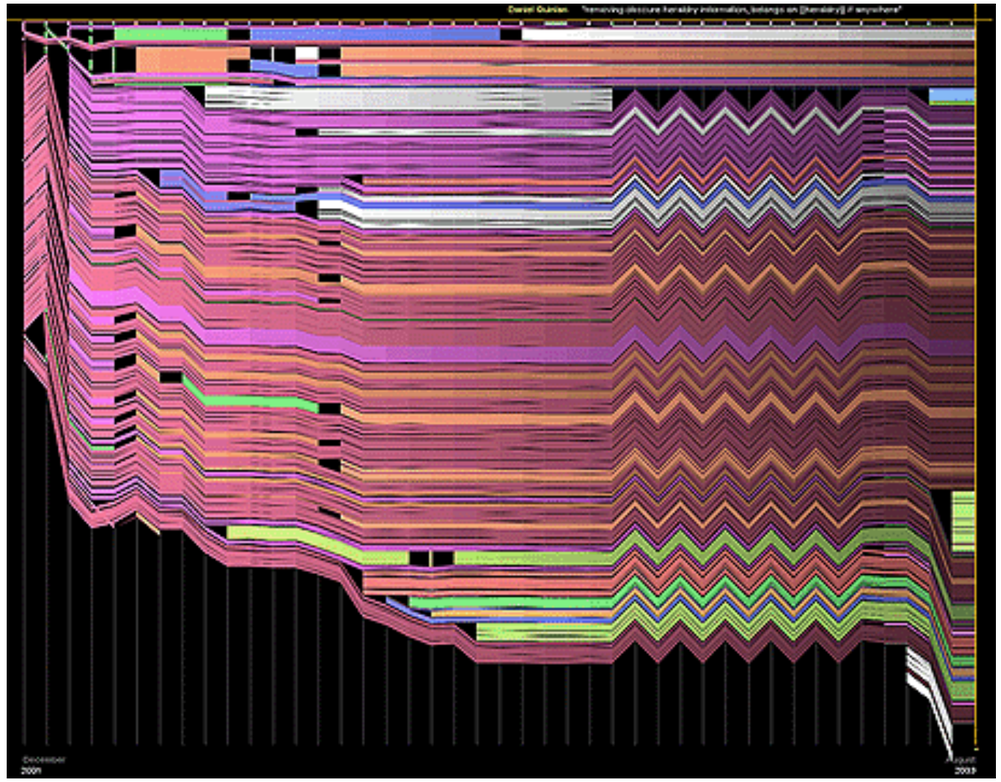
Keywords
Wiki, revision history, visualization, collaboration, document

INTRODUCTION
Online communities have long allowed people with conflicting perspectives and values to meet and talk—but usually without any need to resolve their differences. Indeed, given the endless arguments often found in traditional online forums, asking that a large group reach consensus online may seem impossible. In recent years, however, new online technologies have arisen that, by their nature, favor consensus building by community members. One example of such a technology is a special kind of web site known as a "wiki." Invented in 1995 by Ward Cunningham [14][9], a defining feature is that any reader of the site may also be an author. Each page has an "edit this page" link at the bottom, allowing users to change the content of the page. This interface supports a higher level of consensus building because a user who disagrees with a statement can very easily delete it. In this sense, the text on wiki pages is content that has survived the critical eye of the community. Since Cunningham's original implementation, wikis have become popular for many purposes both public and private, ranging from knowledge management to education [1][5].

This paper is an examination of the largest public wiki, wikipedia.org (or simply "Wikipedia"), which is a thriving site despite a seemingly unlikely model for success. The founders of Wikipedia wished to create a free online encyclopedia. Rejecting the traditional method of having each article written by an expert and subjected to review, fact-checking and editing, they took the opposite tack: on Wikipedia, content can be added or changed at any time by anyone on the Internet. To many, this approach—so vulnerable to mistakes, ignorance and malice—seems a flatly ridiculous way of producing a serious reference tool. The mystery of Wikipedia is that despite the obvious potential drawbacks of its openness, it has enjoyed significant success. It currently contains articles on more than 100,000 subjects, and from July 2002 to July 2003, it averaged 150,000 page views and 3,300 edits per day [18]. It has attracted many writers, but—more importantly—many readers, suggesting that the articles are worth reading.

In this paper, we describe our investigation into how and why such an open and vulnerable system works. Wikipedia generously makes public its database of articles, along with all past revisions of those articles, providing a rich record of interactions between authors. Mining this vast data set is a challenge: to tackle it, we created a new visualization method, dubbed *history flow*, designed to show relationships between multiple document versions. Exploratory analysis with this visualization revealed complex patterns of cooperation and conflict. We also describe some initial statistical corroboration for the patterns we find. Finally, we propose several hypotheses based on these analyses for how and why this collective authoring environment succeeds.

Our chief conclusion is that Wikipedia and its audience must be viewed as a system in which constant change is a source of strength as well as weakness. The site is subject to frequent vandalism and inaccuracy, just as skeptics might suspect—but the active Wikipedia community rapidly and effectively repairs most damage. Indeed, one type of



Revisions of the "Chocolate" Wikipedia article

The Problem: They are not easy to adapt for other domains!

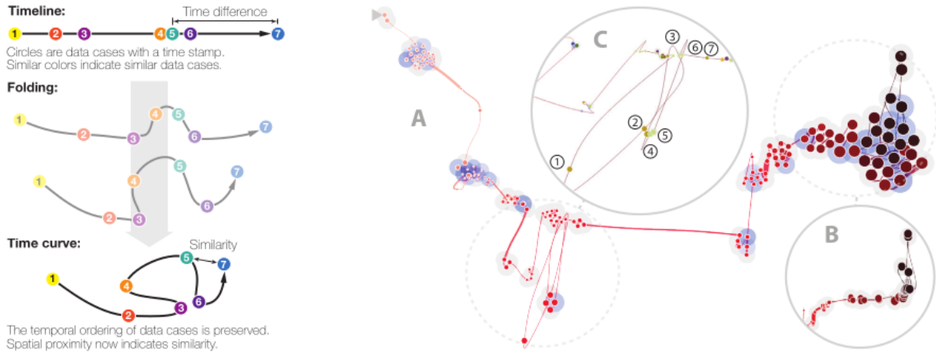
The goal: Development of dataset agnostic visualization method.

The Paper

559 IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 22, NO. 1, JANUARY 2016

Time Curves: Folding Time to Visualize Patterns of Temporal Evolution in Data

Benjamin Bach, Conglei Shi, Nicolas Heulot, Tara Madhyastha, Tom Grabowski, Pierre Dragicevic



(a) Folding time (b) History of the Wikipedia article on Palestine

Fig. 1. The time curve principle: a) a timeline is folded into itself in such a way that similar time points end up being close to each other; b) Example: a time curve showing the evolution of a Wikipedia article.

Abstract—We introduce *time curves* as a general approach for visualizing patterns of evolution in temporal data. Examples of such patterns include slow and regular progressions, large sudden changes, and reversals to previous states. These patterns can be of interest in a range of domains, such as collaborative document editing, dynamic network analysis, and video analysis. Time curves employ the metaphor of folding a timeline visualization into itself so as to bring similar time points close to each other. This metaphor can be applied to any dataset where a similarity metric between temporal snapshots can be defined, thus it is largely datatype-agnostic. We illustrate how time curves can visually reveal informative patterns in a range of different datasets.

Index Terms—Temporal data visualization, information visualization, multidimensional scaling

2016 TVCG Paper with 247 citations (Google Scholar)

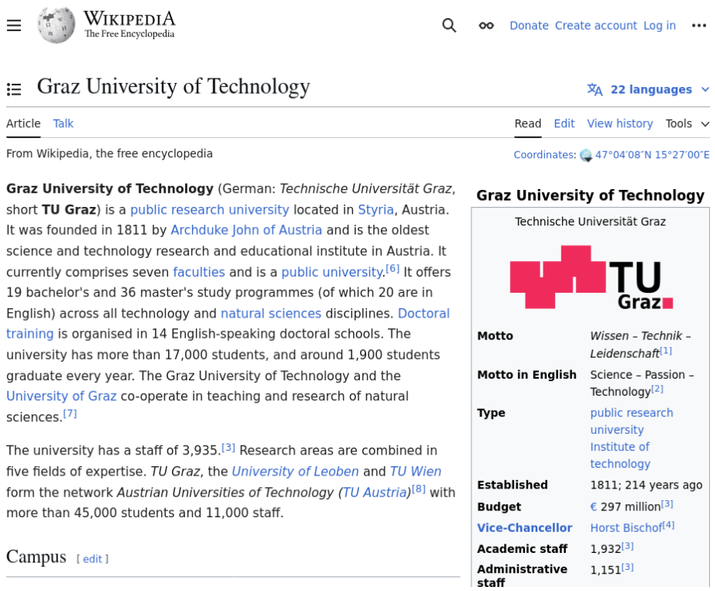
Time Curves

What do we need?

First: Snapshots of your dataset at different time points

Second: Similarity Measure

Example



Wikipedia
The Free Encyclopedia

Donate Create account Log in

Graz University of Technology

22 languages

Article Talk | Read Edit View history Tools

From Wikipedia, the free encyclopedia

Coordinates: 47°04′08″N 15°27′00″E

Graz University of Technology (German: *Technische Universität Graz*, short **TU Graz**) is a public research university located in Styria, Austria. It was founded in 1811 by Archduke John of Austria and is the oldest science and technology research and educational institute in Austria. It currently comprises seven faculties and is a public university.[6] It offers 19 bachelor's and 36 master's study programmes (of which 20 are in English) across all technology and natural sciences disciplines. Doctoral training is organised in 14 English-speaking doctoral schools. The university has more than 17,000 students, and around 1,900 students graduate every year. The Graz University of Technology and the University of Graz co-operate in teaching and research of natural sciences.[7]

The university has a staff of 3,935.[3] Research areas are combined in five fields of expertise. *TU Graz*, the *University of Leoben* and *TU Wien* form the network *Austrian Universities of Technology* (*TU Austria*)[8] with more than 45,000 students and 11,000 staff.

Campus [edit]

Graz University of Technology

Technische Universität Graz

TU Graz

Motto	*Wissen – Technik – Leidenschaft*[1]
Motto in English	Science – Passion – Technology[2]
Type	public research university Institute of technology
Established	1811; 214 years ago
Budget	€ 297 million[3]
Vice-Chancellor	Horst Bischof[4]
Academic staff	1,932[3]
Administrative staff	1,151[3]

Snapshots of Wikipedia articles

Similarity measure: edit distance

Arranging Points by Similarity

Multidimensional Scaling

$$\text{Stress}_D(x_1, x_2, \ldots, x_N) = \sqrt{\sum_{i \neq j = 1,\ldots,N} (d_{ij} - \|x_i - x_j\|)^2}$$

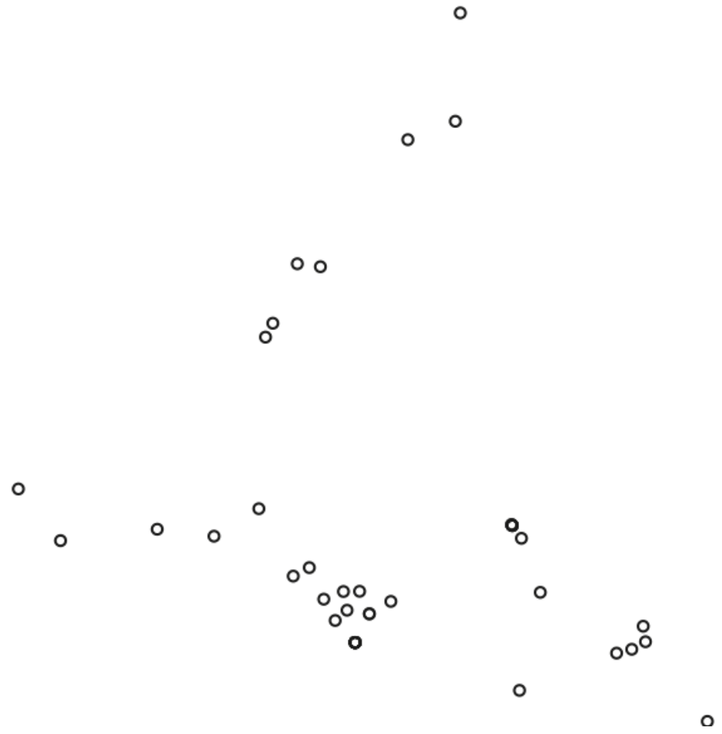
x_i . . . data points in projected space

d_{ij} . . . similarity between data points in high dimensional space

$\|x_i - x_j\|$ . . . distance between projected points

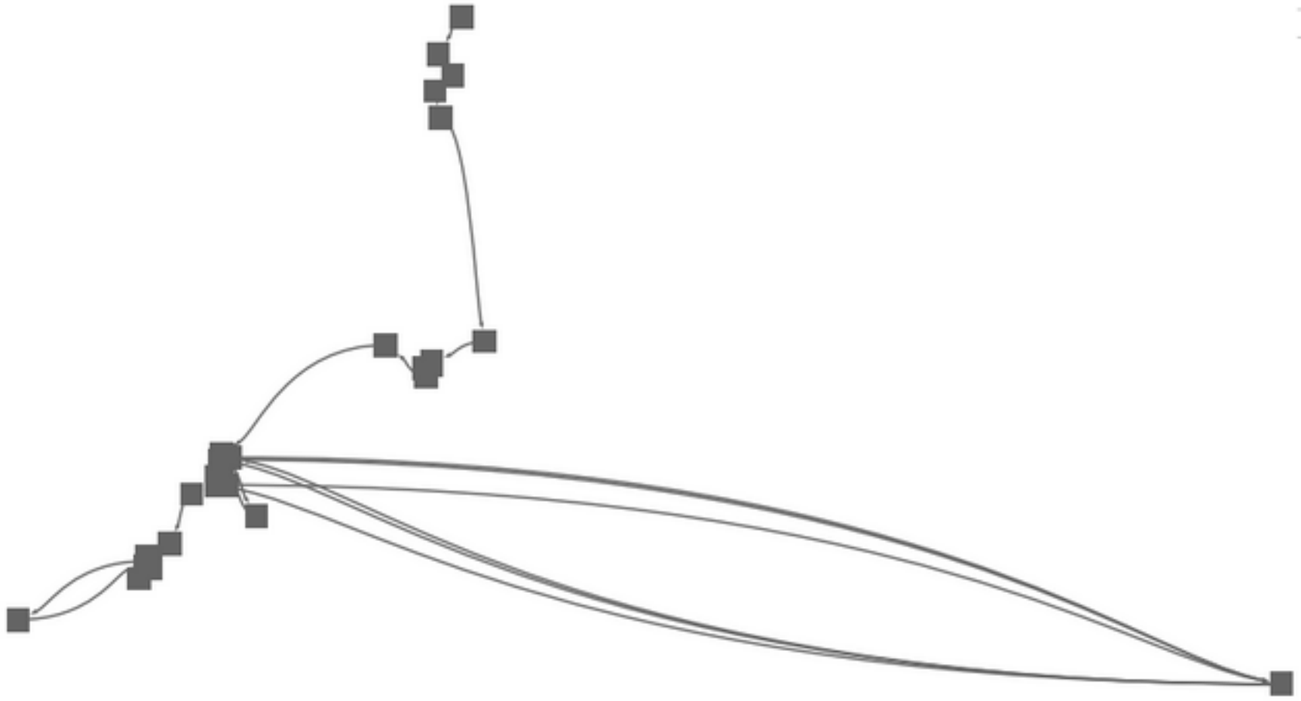
Goal: Minimise Stress

First Step



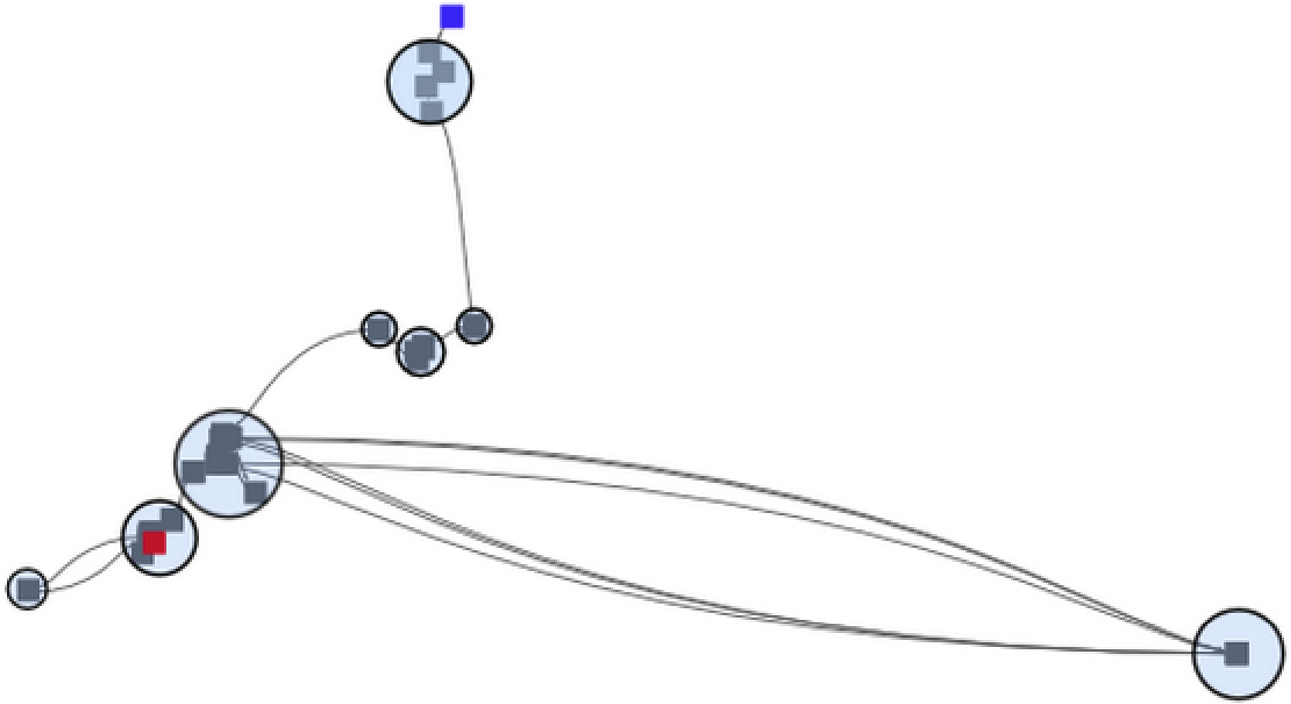
Apply MDS to the dataset

Second Step



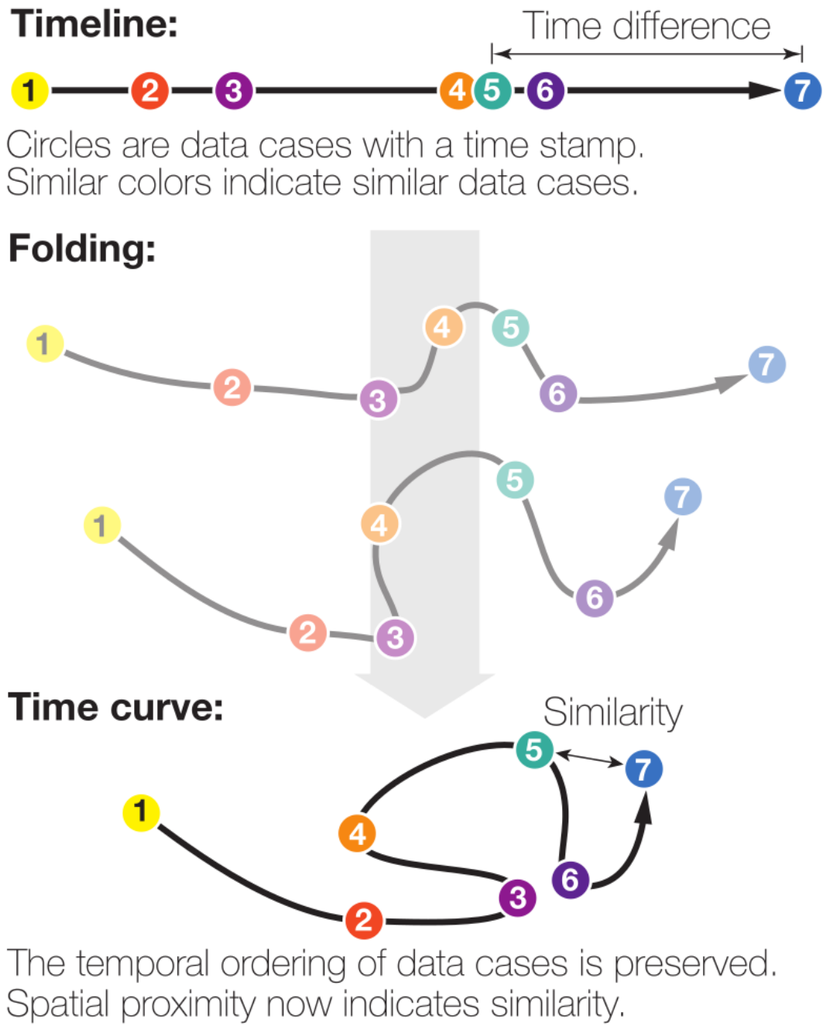
Connect the points in their temporal ordering

Third Step



Remove overlaps and color the points

Like Folding a Timeline



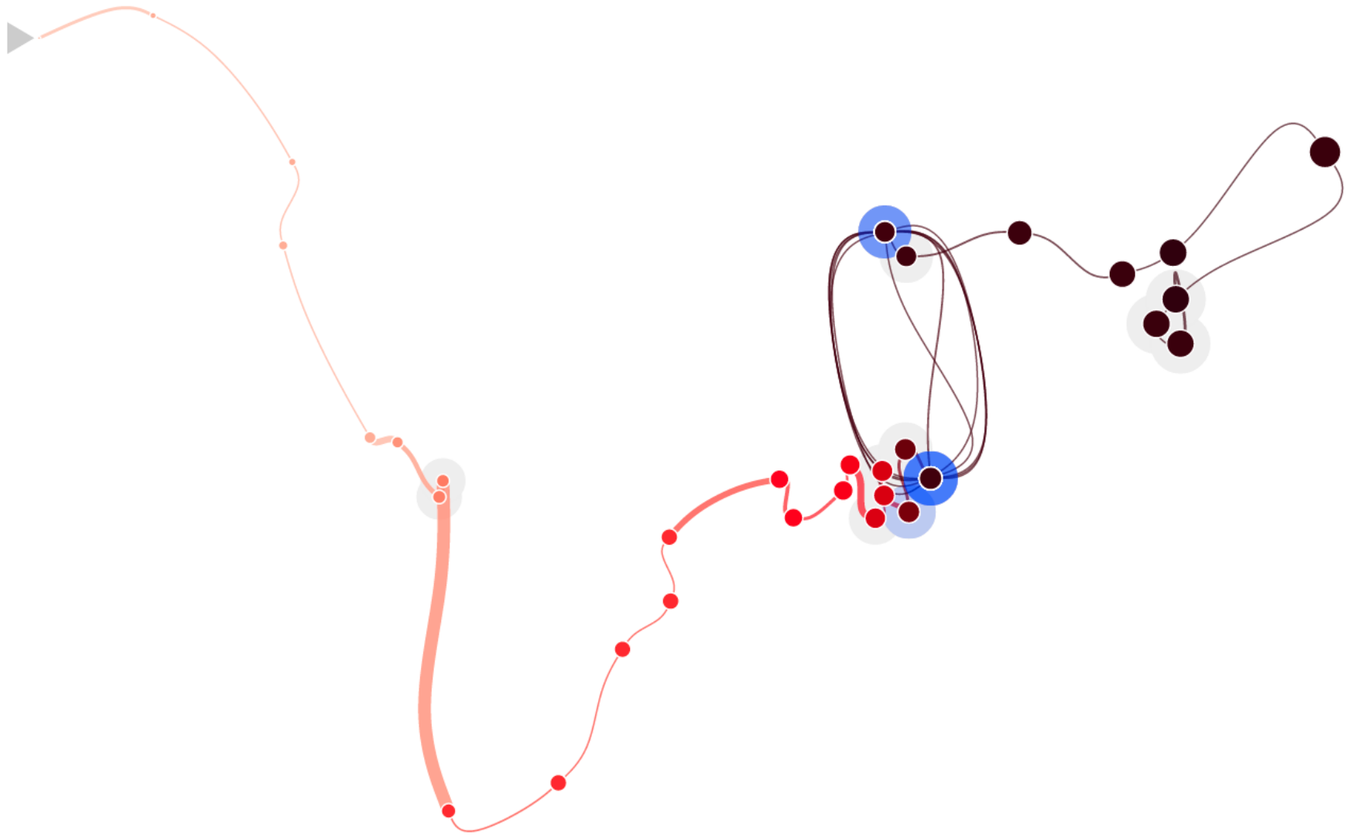
Important choices

1. Similarity metric
2. Dimensionality reduction method, e.g. MDS
3. Curve drawing algorithm, e.g. Catmull-Rom
4. Extras:
   - Remove overlap
   - Coloring
   - Node size

Applications

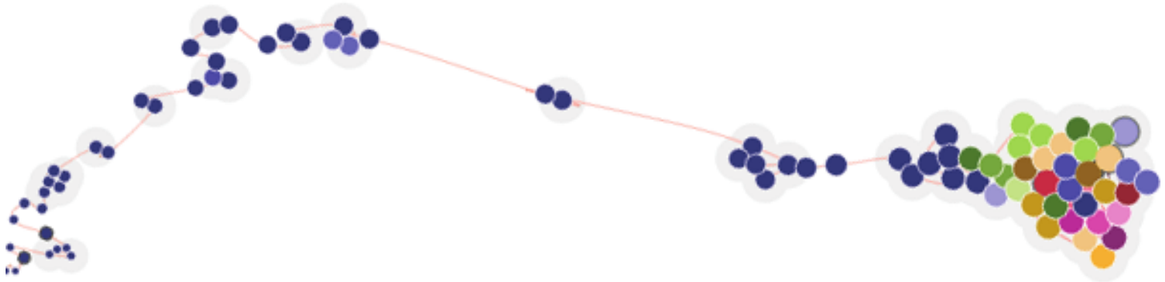
Document Histories

The Wikipedia chocolate page edit war



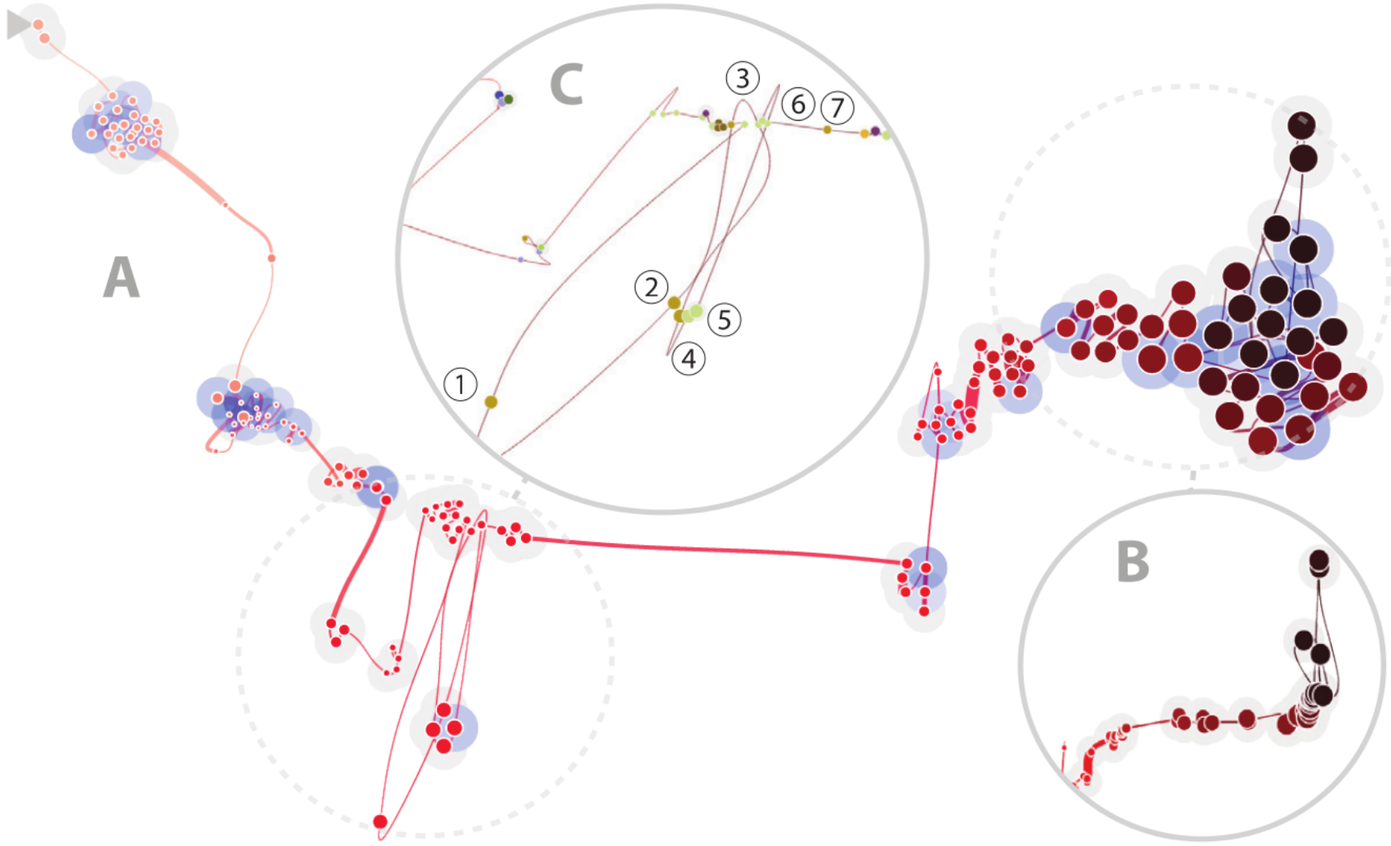
Document Histories

Alternative coloring highlights user groups



Document Histories

Palestine Wikipedia Page (Caution: 2016)

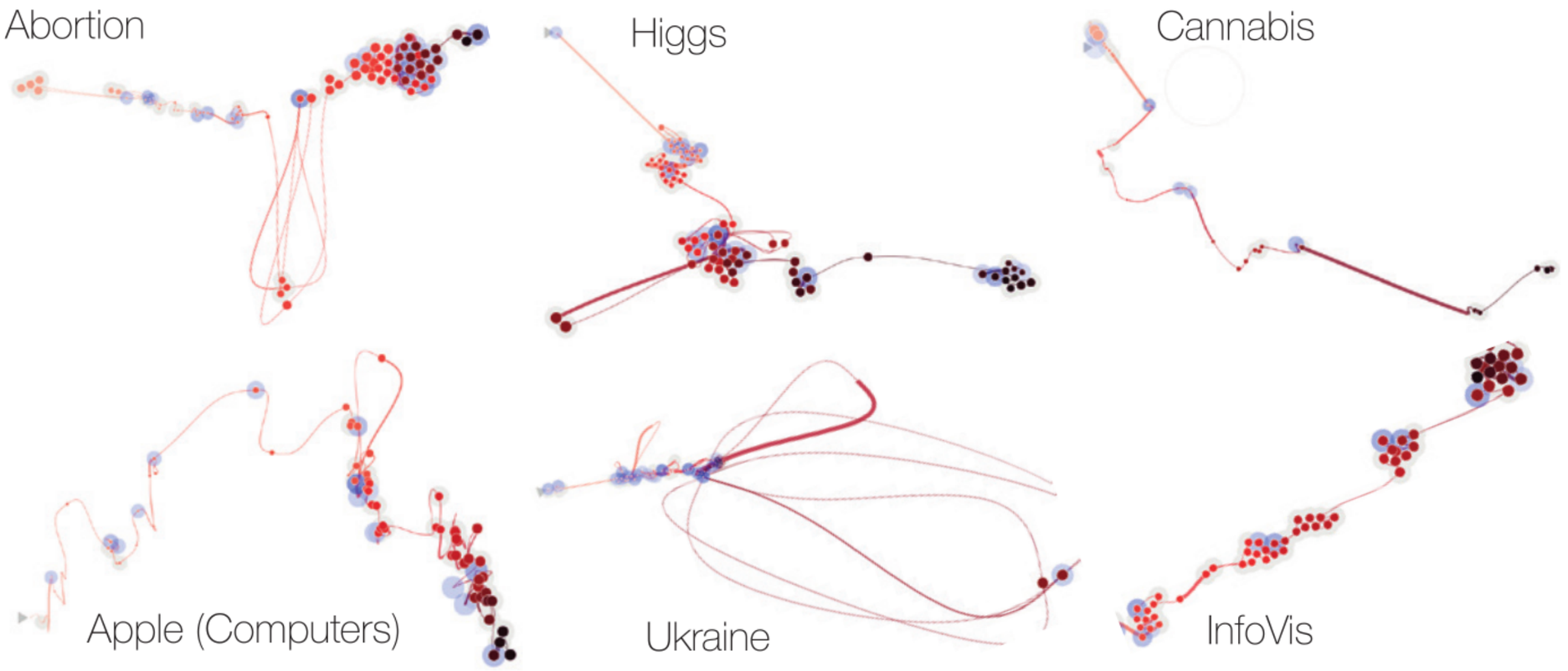


Document Histories

Vandalism

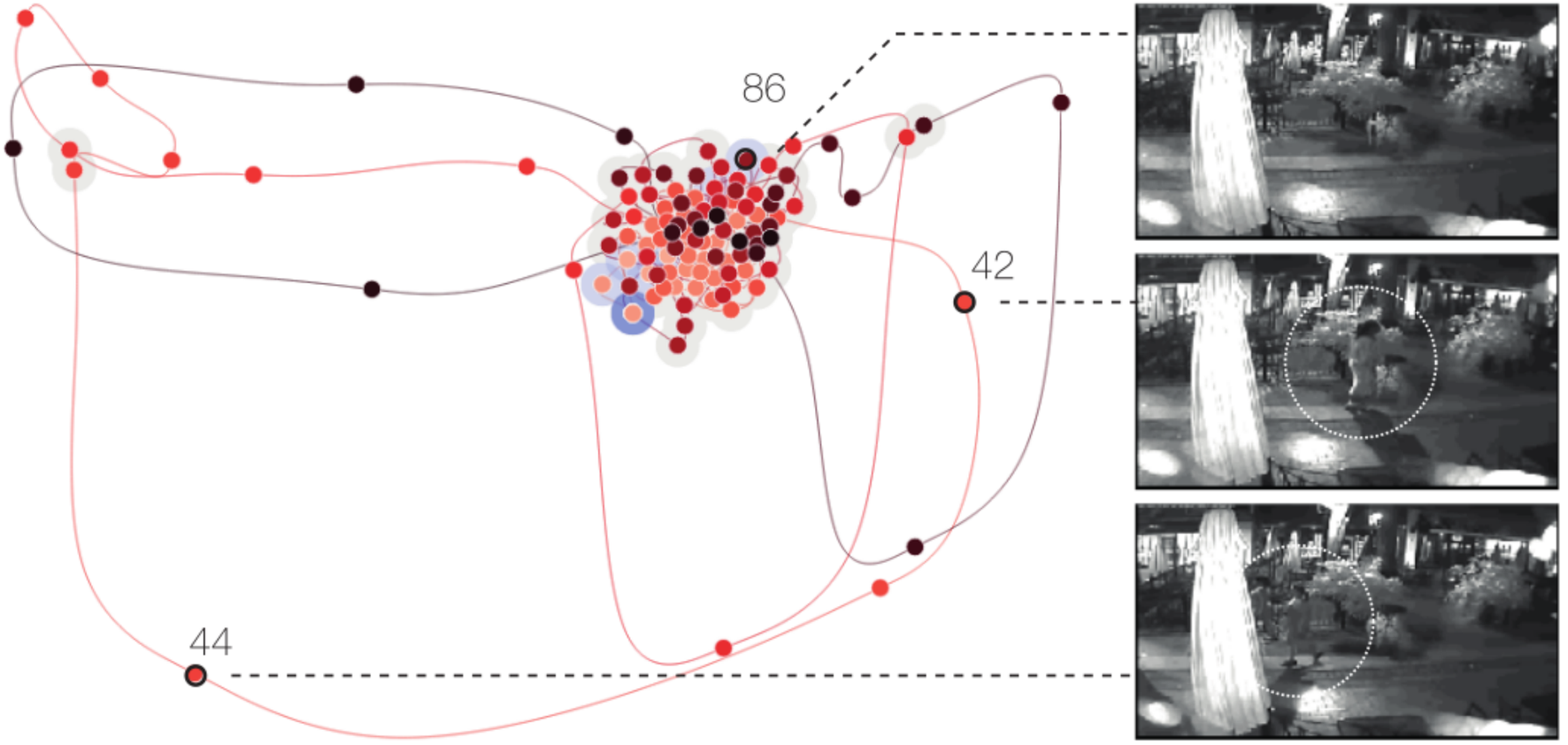
Document Histories

Time Curves as fingerprints



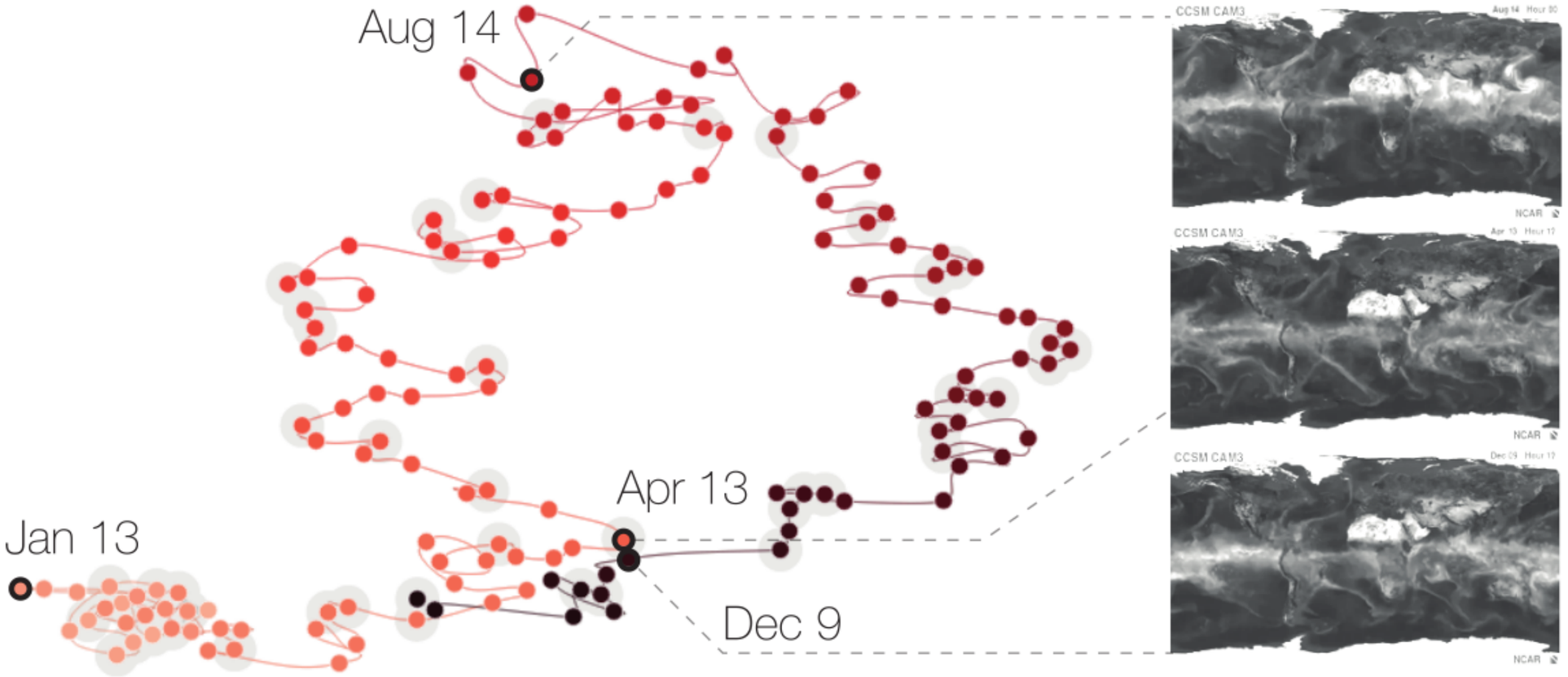
Video Time Curves

Surveillance

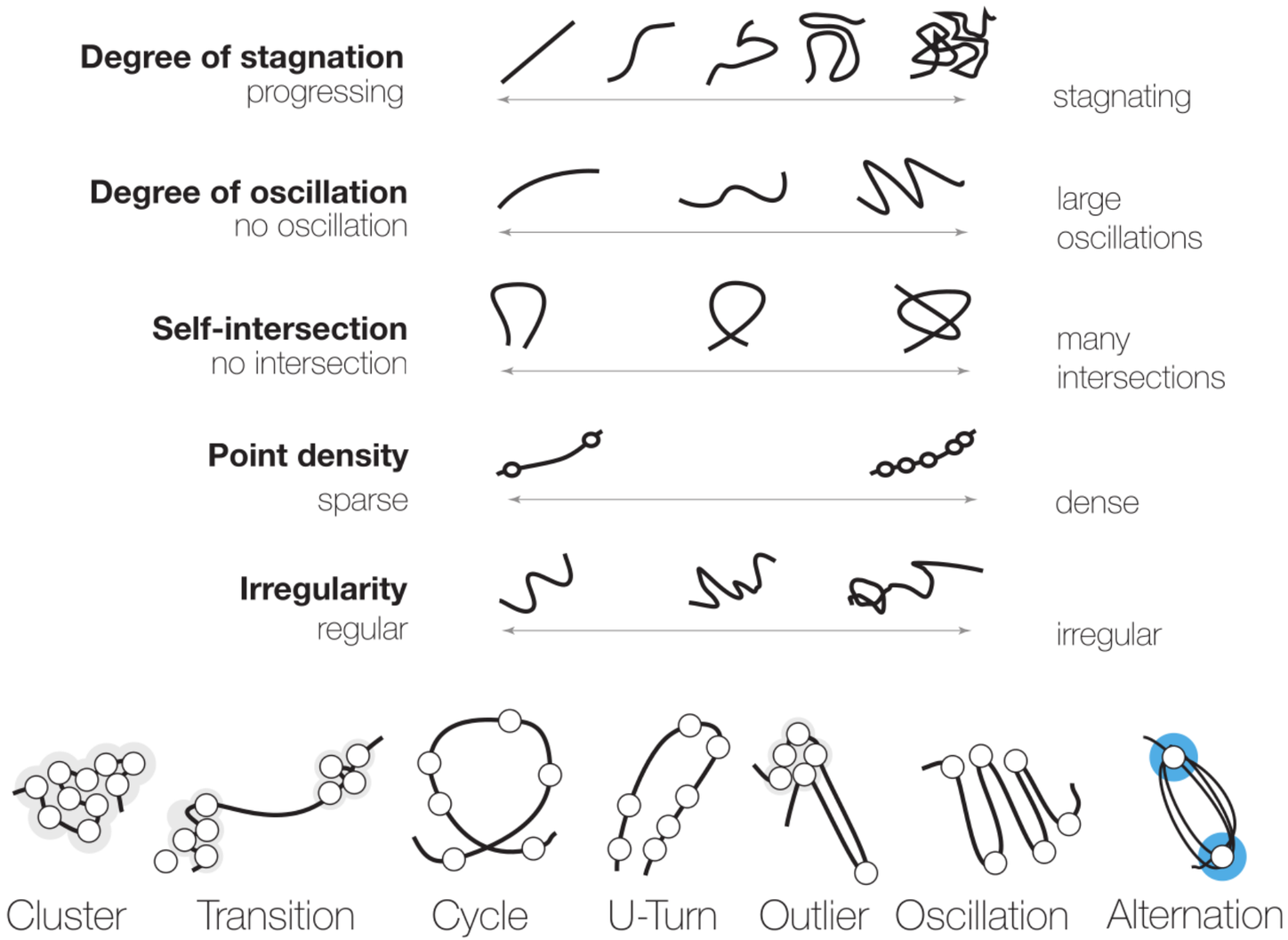


Video Time Curves

Precipitation



Taxonomy of patterns and characteristics



Limitations

Limitations

- Quantitative aspect of time is lost.
- High dependence on distance metric.
- MDS expensive to compute.
- Curve might not be legible due to high complexity.
- Not resistant to noise.

Why did I choose this paper?

Because it might be relevant for your projects!

- PRESENT
- Hereditary
- A+CHIS
- OpenReassembly: Possible Puzzle Wars?

Can you think of possible use cases?

Could you show an "edit war" in some datasets provided in your projects?