

Analytics Visualization of User Email Activities in Corporate Networks

Background - Area of Invention

This invention relates to data summarization and visualization. We present a tool that automatically summarize email communication records into threads for each user, and visualize each user's activity records in an interactive, bio-inspired view.

Email communication records contain heterogeneous attributes such as identity, time-stamp, size, type, duration, account names of sender and receiver, etc. The dataset regarding user activities is huge. And the size is compounded by the number of users, the number of threads created by each user, and the rich content in each thread.

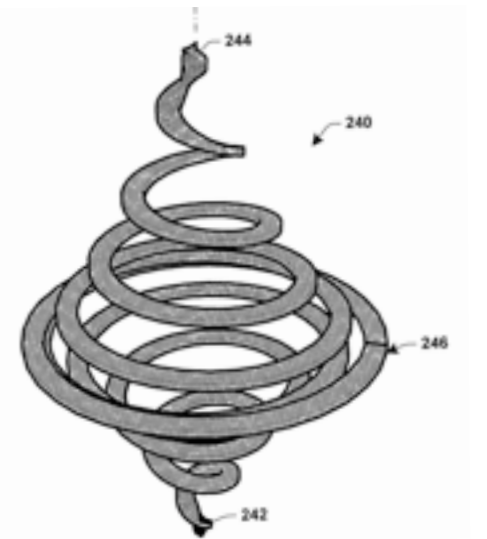
Given the complexity, interrelationships, and the sheer volume, the challenge we aim to address is to design a data transformation and visualization scheme that facilitate pattern discovery, anomaly association.

Background - Prior Art

Related Patents (search keywords: “circular timeline”, “temporal data visualization”, “activity visualization”)

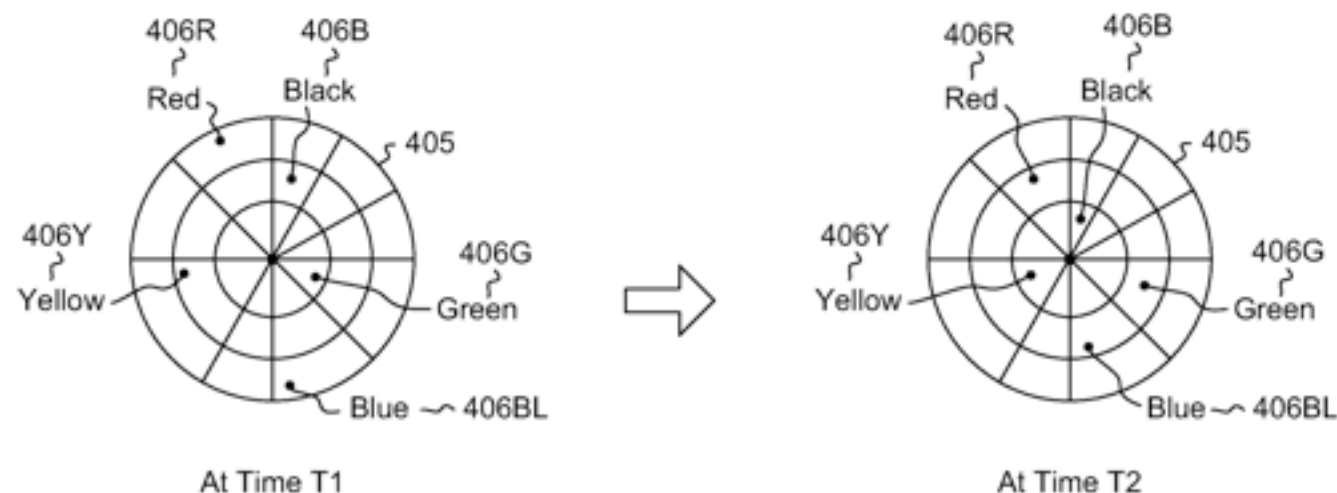
US20020126121 presents the use of a three-dimensional helical path to facilitate the visualization of data that may be unbounded in time. This visualization model is not applicable to multivariate dataset.

US 8443293 B2 proposes to visualize multiple data items with multiple attributes simultaneously. Each data item is displayed in an appropriate slice and ring, according to the data item's corresponding features and attributes. To reveal the temporal trend, it requires transition from one view to the next, instead of showing in a static image.



helix

bull's eye



Background - Prior Art

Related Publication (search keywords: “circular timeline”, “temporal data visualization”, “activity visualization”)

Time-series Bitmap (Kumar et. al.) is a tool that uses symbolic aggregation approximation to a real-valued time sequence as a bitmap icon. In their work, different periodic patterns are shown in bitmap views. It is unclear as to how to visualize various features over time as one image icon in their design.

Silva and Catarci present a survey of visualization of linear time-oriented data, while our work is a visual design for non-linear time-oriented data: we visualize the participant-thread association over time.

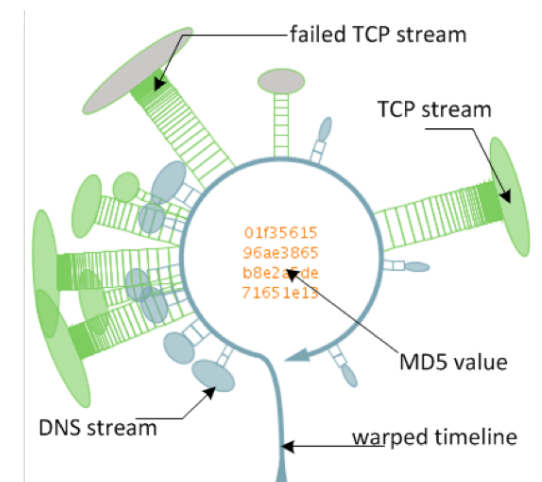
One related design is MalwareVis (Zhuo et. al.) in which each view represents the communication records, in multivariate time-series form, of a malware instance collected in the wild. This is also a type of linear time-oriented data visualization.



time-series bitmaps



perspective wall



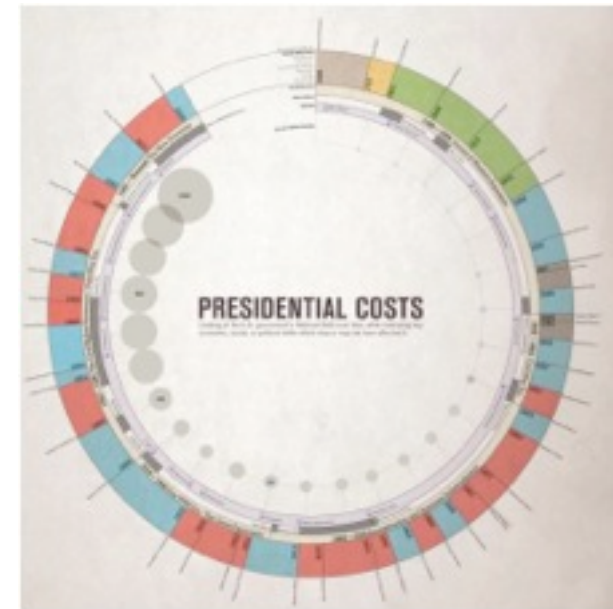
MalwareVis

Background - Prior Art

Related Material on Web (search keywords: “circular timeline”, “activity visualization”)

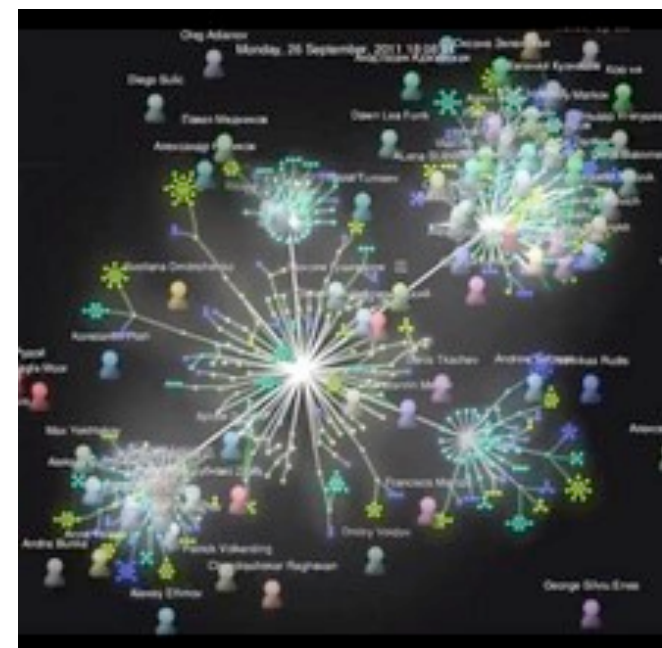
<http://www.improving-visualisation.org/vis/id=338>

is a example of circular timeline combined with bubble chart to show temporal trend of presidential costs. The visualization is on printed paper and is not computer automated.



<http://www.visualizing.org/galleries/visualization-public-activity-user-google>

of-the-shelf graph visualization tools, such as Gephi, have been used to visualize activities in online social networks. It is a network based visualization rather than based on each user.



Summary - Problem being solved

1. Data Transformation

user-based summarization and visualization

2. Visualization Design and Encoding

bio-inspired visual design and encoding

3. Use Scenario

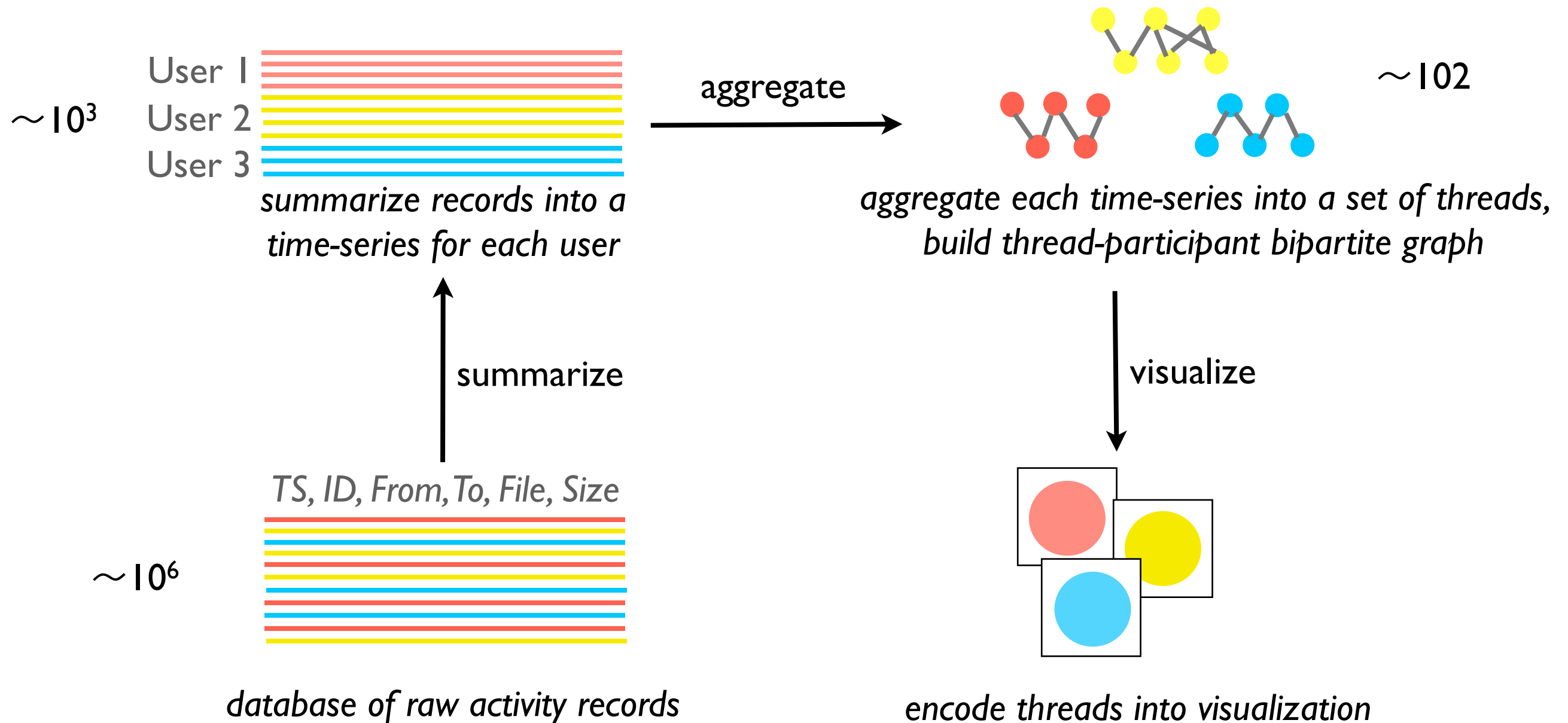
anomaly association, pattern discovery

4. Visual Clutter Reduction

linearly interpolated layout, constraint packing

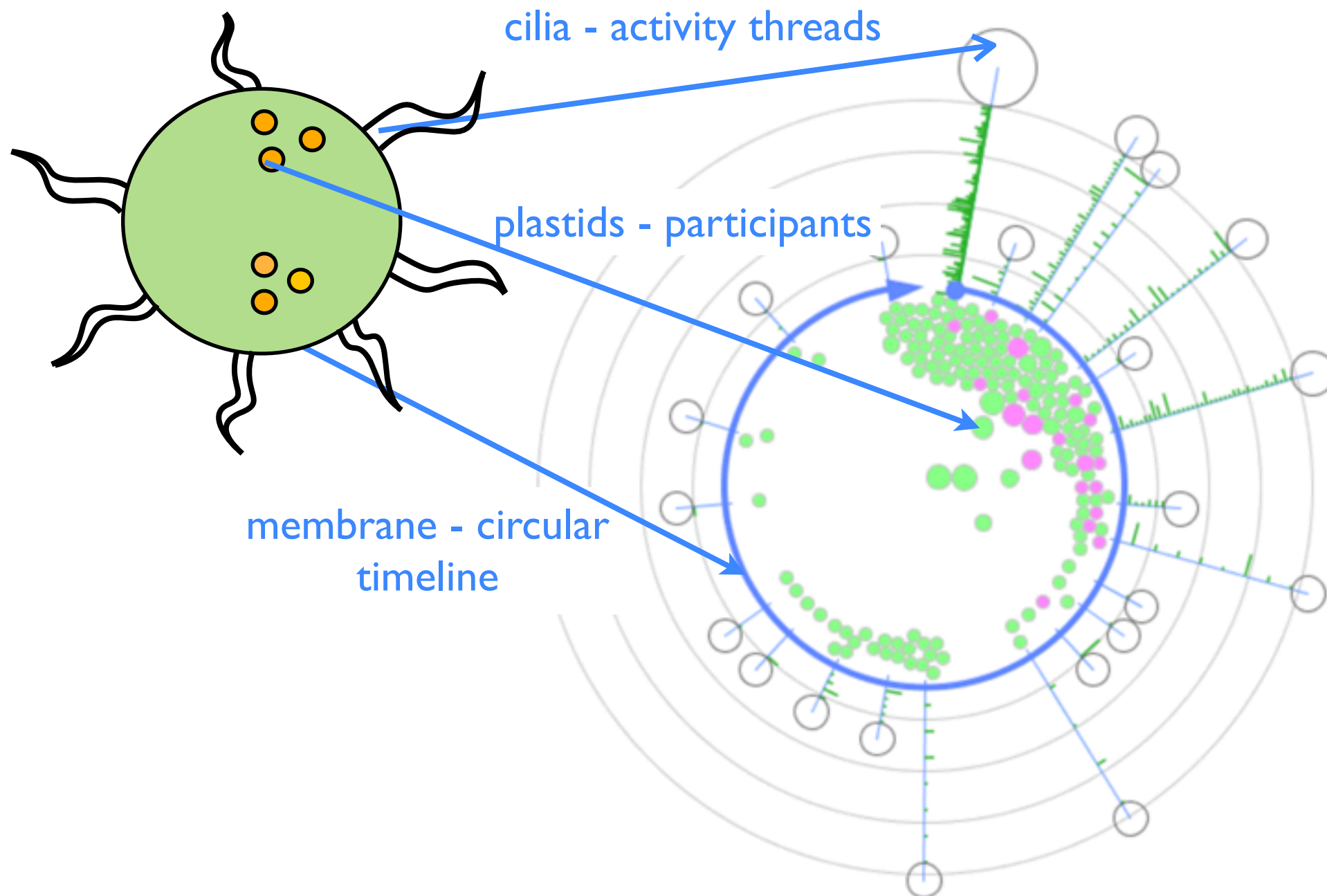
Details - Data Transformation

User-based Summarization and Visualization



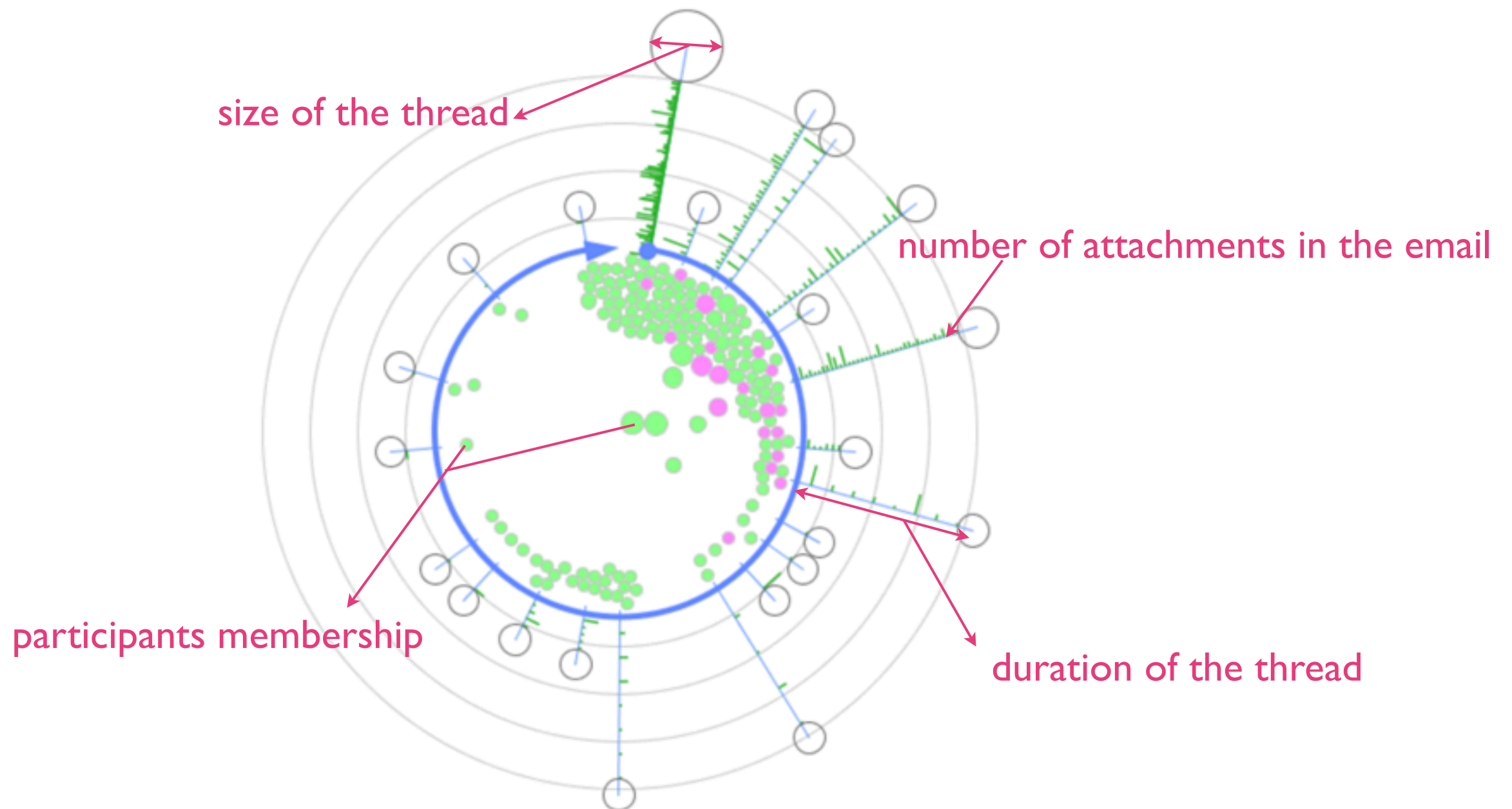
Details - Visualization Design

Each user is represented as a cell



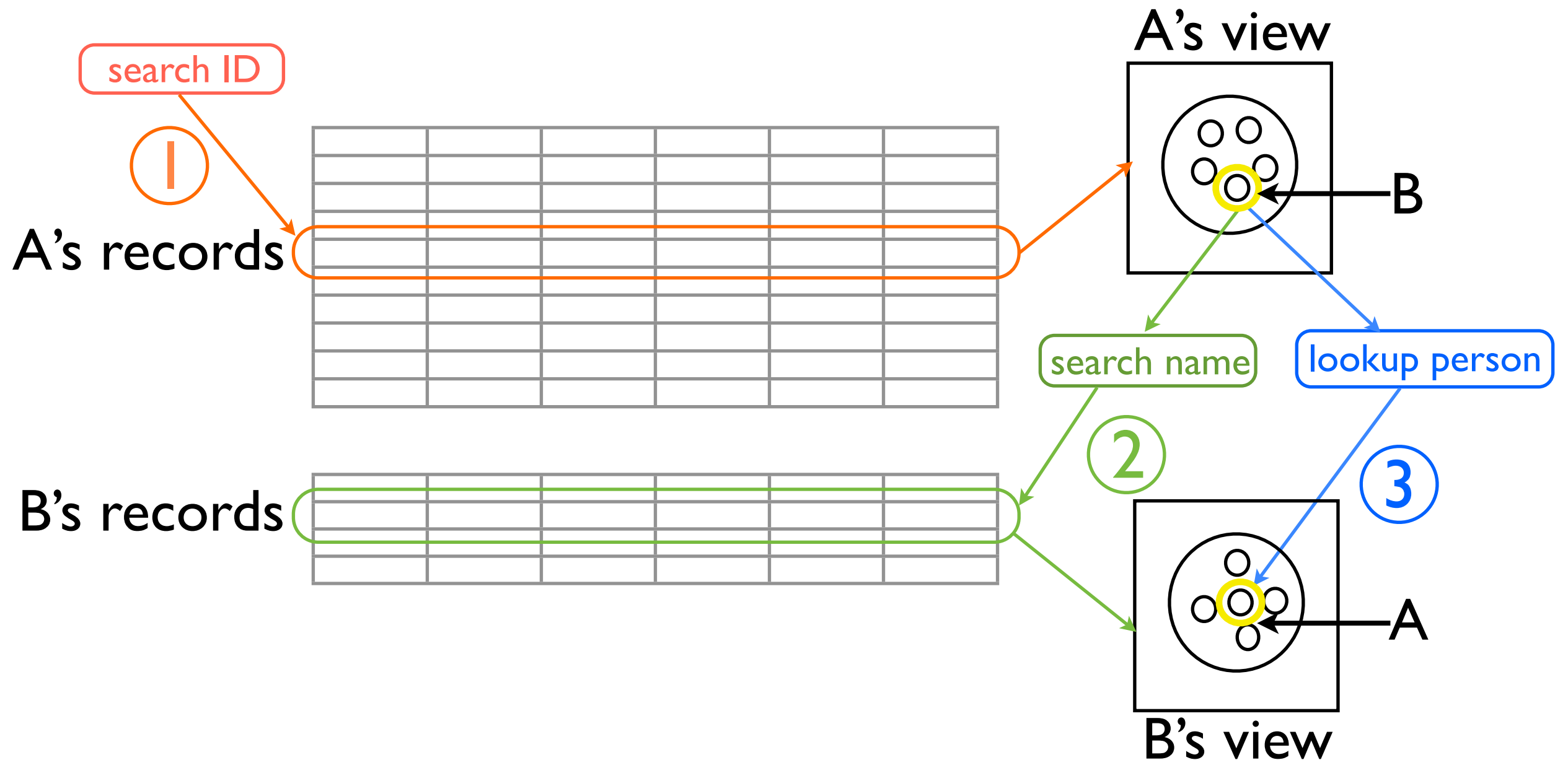
Details - Visual Encoding

Encode various features into geometric dimensions



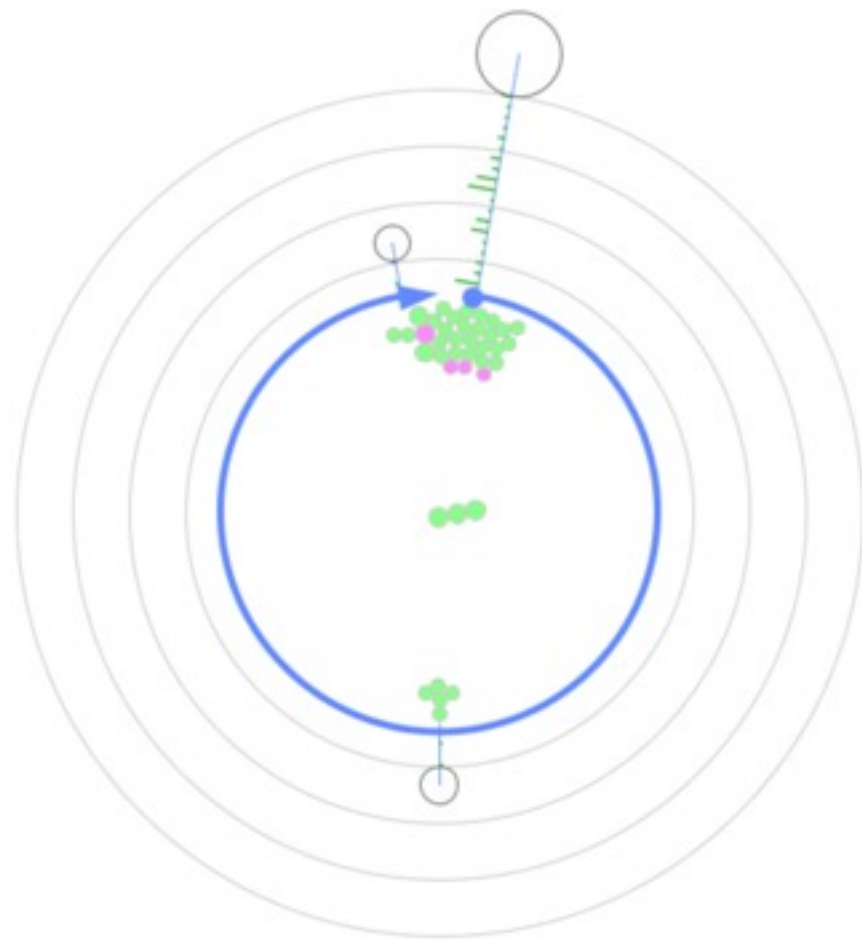
Details - Usage Scenario

Anomaly association

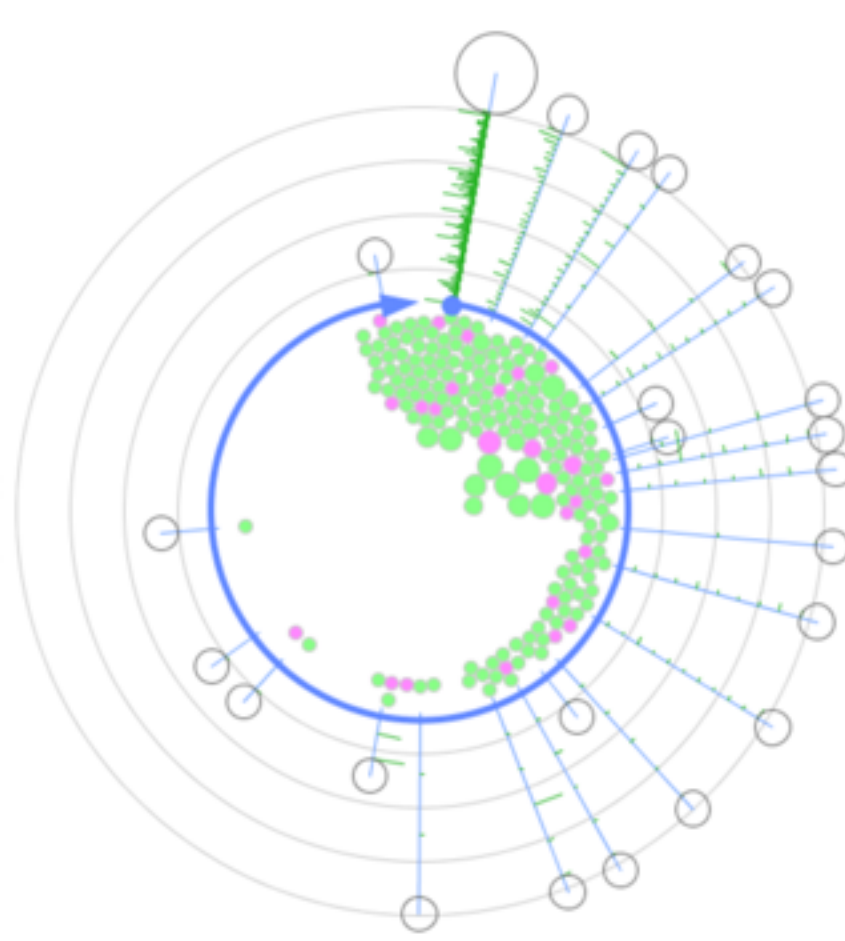


Details - Usage Scenario

Pattern discovery



inactive



multi-tasking



state-switching

Details: Visual Clutter Reduction

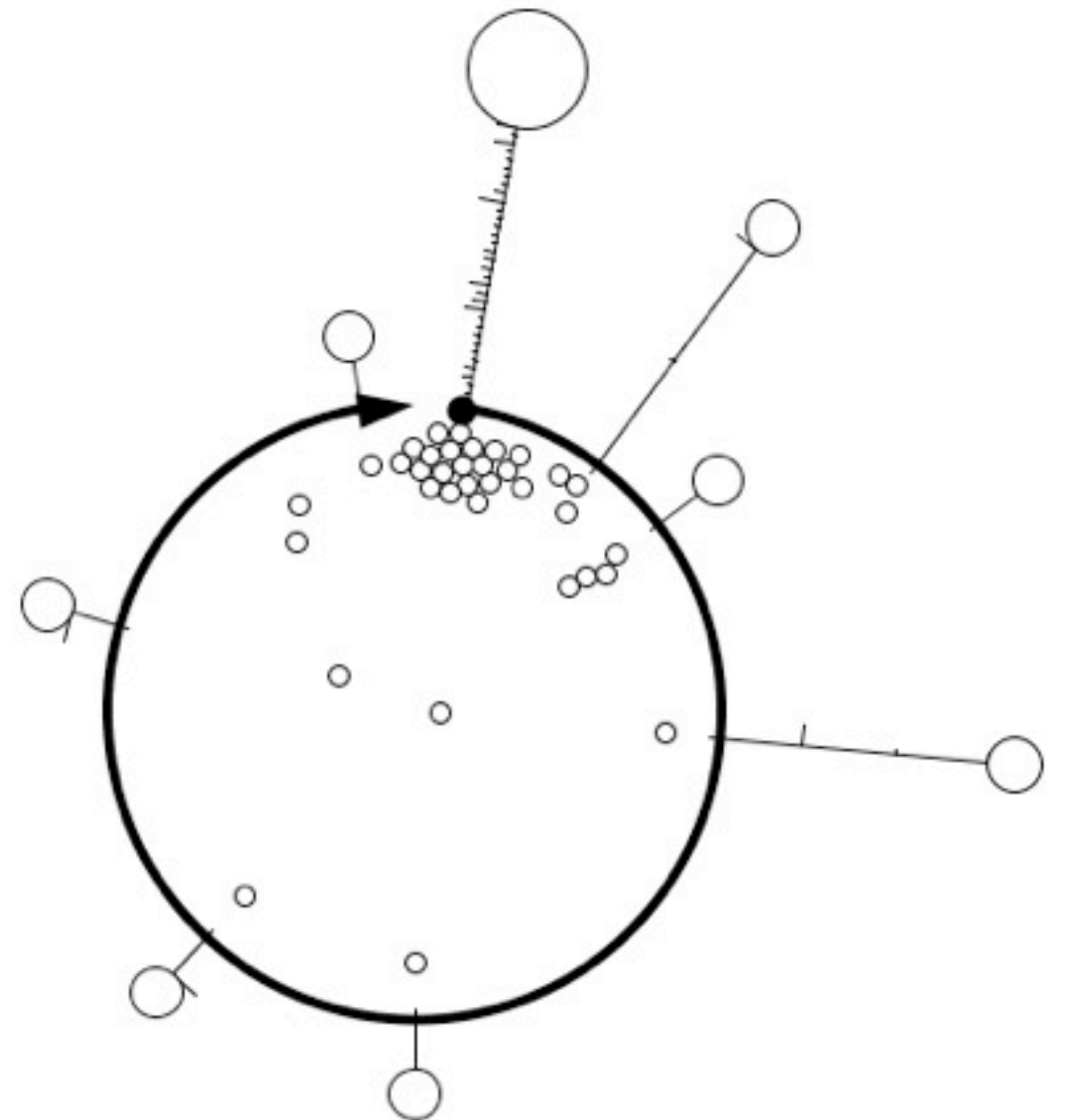
Linearly Interpolated Layout

The position P_i of each node

$$P_i = \sum_j w_{ij} C_j$$

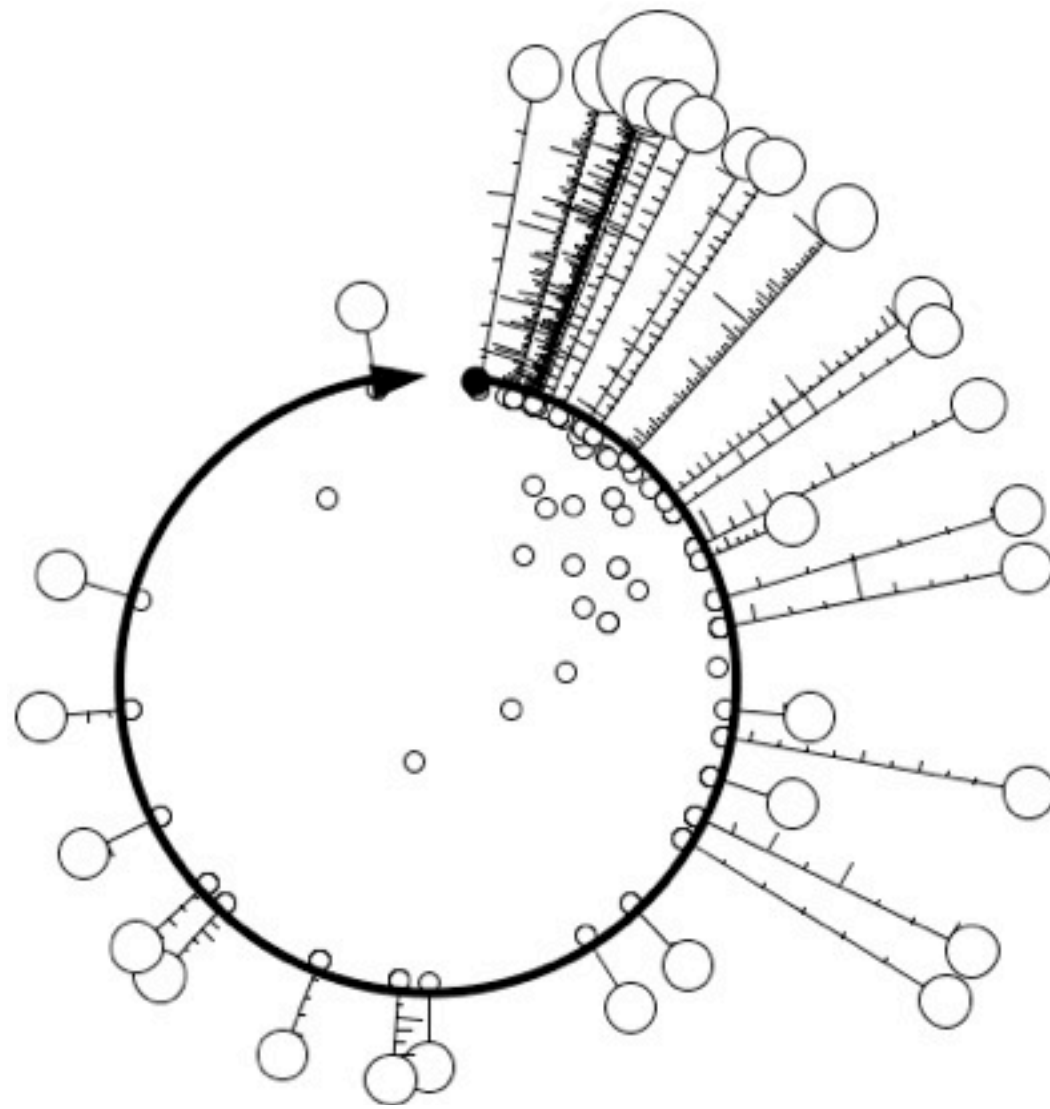
$$w_{ij} = \delta(i, j) / \sum_k \delta(i, k)$$

$\delta(i, k) = 1$ if user i participate
in thread k , 0 otherwise

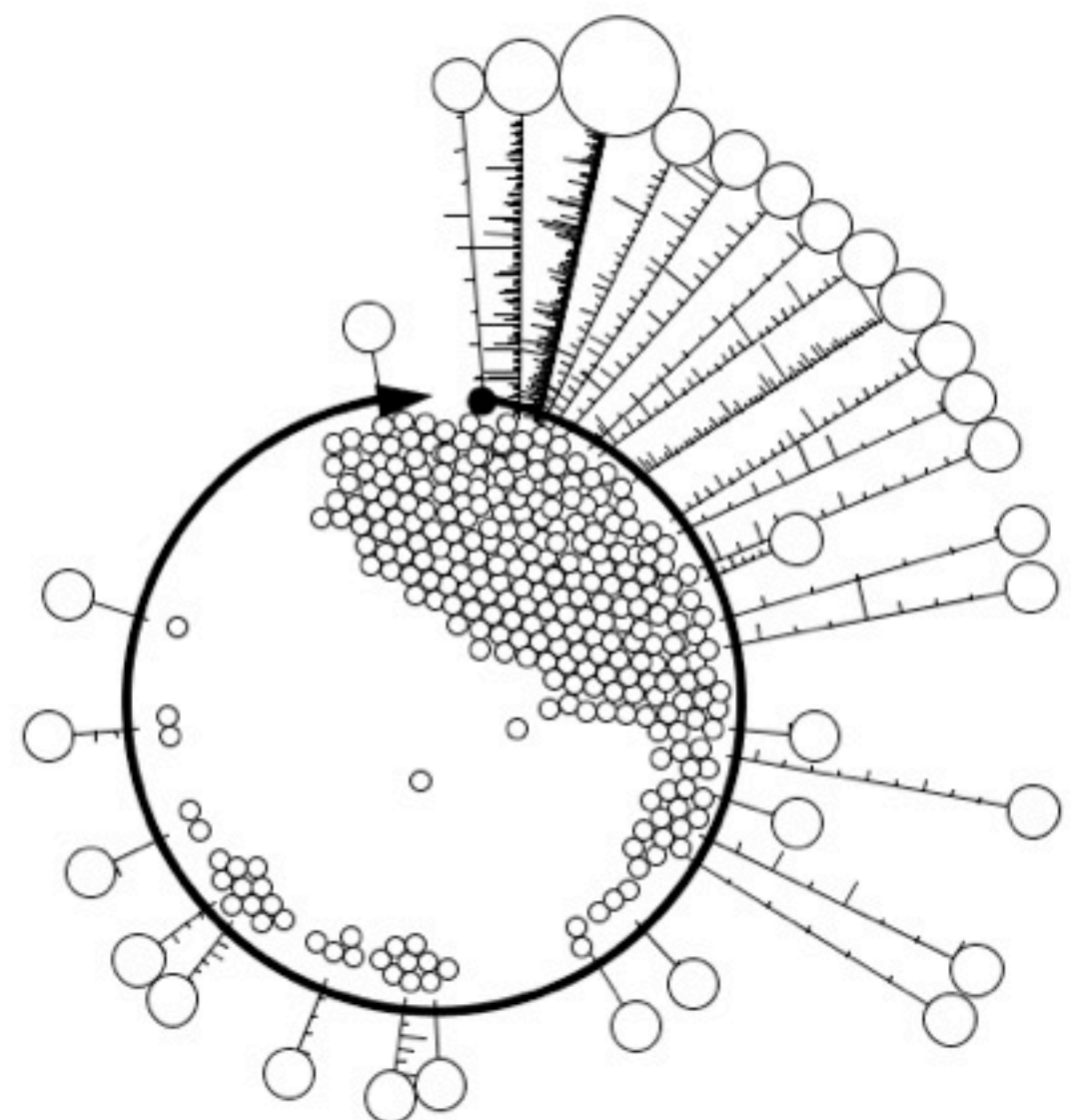


Details: Visual Clutter Reduction

Constraint Packing



before packing



after packing