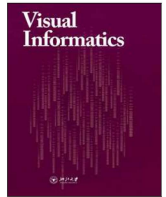京都大学
KYOTO UNIVERSITY

Title	A visual analytics system to support the formation of a hypothesis from calcium wave data
Author(s)	Umezawa, Kozen; Natsukawa, Hiroaki; Onoue, Yosuke; Koyamada, Koji
Citation	Food Science and Human Wellness (2018), 2(1): 2-13
Issue Date	2018-03-01
URL	http://hdl.handle.net/2433/243570
Right	© 2018 Published by Elsevier B. V. on behalf of Zhejiang University and Zhejiang University Press. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Type	Journal Article
Textversion	publisher

A visual analytics system to support the formation of a hypothesis from calcium wave data

Kozen Umezawa[a,*], Hiroaki Natsukawa[b], Yosuke Onoue[c], Koji Koyamada[b]

[a]*Department of Electrical Engineering, Graduate School of Engineering, Kyoto University, Kyoto, Japan*
[b]*Academic Center for Computing and Media Studies, Kyoto University, Kyoto, Japan*
[c]*Advanced Research Center for the Promotion of Interdisciplinary Education and Research, Kyoto University, Kyoto, Japan*

ARTICLE INFO

ABSTRACT

In most species, calcium waves in the oocyte are considered common phenomena in the activation of eggs. However, the mechanism of calcium waves has not yet been clarified. By collaborating with biologists studying *Caenorhabditis elegans* (*C. elegans*), which is widely used as a model organism, we observed that the following requirements must be satisfied to form a useful hypothesis based on calcium waves captured using high-speed in vivo imaging: (1) the ability to obtain an overview of how the calcium waves are propagated and (2) the ability to understand the propagation of waves in a narrow region. However, conventional visualization methods cannot satisfy these requirements simultaneously. Therefore, we propose a visual analytics system that allows users to understand and explore calcium wave images using cross-correlation analysis of the time-series data of the Ca^{2+} fluorescence intensity at each point. The interface of this system comprises an overview visualization, a detail visualization, and user interactions to satisfy these requirements and realize exploratory visualization. Some views present an overview visualization that displays the clustering results of a directed graph calculated using cross-correlation analysis. These views enable the users to understand the overview of wave propagation, thereby helping users find a region of interest. The detail visualization shows the relationship between the region of interest and other areas. Furthermore, users can use the proposed system with overview-detail and brush-link exploration to assign meaning to the region of interest and construct a hypothesis for its role. In this paper, we demonstrate how the proposed visual analytics approach works and how new hypotheses can be formed using the analysis of *C. elegans* calcium waves.

1. Introduction

It is known that the concentration of intracellular calcium ions increases during fertilization in animal oocytes and eggs (Whitaker, 2006). This phenomenon is termed as calcium waves because the increase in calcium ions propagates as a waveform from the sperm entry point. In addition, depending on the species, calcium oscillation occurs, which involves repeated increase and decrease of calcium ion concentration (Stricker, 1999). An increasing calcium ion concentration is an important signal for activating the oocyte and causing subsequent cell division. Previously, numerous studies on calcium waves have been conducted because the relationship between calcium waves and subsequent evolution can be clarified by elucidating the propagation of calcium waves.

We collaborated with life scientists who are investigating to reveal the mechanism of calcium waves, as well as the differences between measured and simulated calcium waves. Compared to simulated waves, the measured data demonstrate a complicated behavior. These scientists are investigating the mechanisms responsible for such complex behaviors and the physiological meaning of those mechanisms. However, there is lack of hypotheses on which to perform an experiment and

Corresponding author:
e-mail: umezawa@viz.media.kyoto-u.ac.jp (Kozen Umezawa)

improve simulations because the complexity of calcium waves observed in imaging data makes it difficult to extract useful information. Furthermore, these scientists hope to reveal the specificity of the phenomenon in a data-driven manner. One approach to satisfy these objectives is to support hypothesis construction using visual analytics.

In several meetings with domain experts, we determined that two design requirements must be satisfied to support hypothesis formation from calcium wave data: (1) the ability to obtain an overview of how calcium waves are propagated and (2) the ability to understand the propagation of waves in a narrow area. To enable experts to observe and interpret a region of interest (ROI), both requirements must be satisfied simultaneously. Some methods that satisfy these requirements individually have been proposed. Methods that satisfy requirement (1) include flow visualization (Afrashteh et al., 2016) using optical flow and flow visualization (Yamashita et al., 2012) using Granger causality (Granger, 1980). Methods that satisfy requirement (2) include calculating the direction of the vector of the flow field (Takagaki et al., 2011) and cross-correlation analysis within a small area of the image. However, to the best of our knowledge, no existing methods can satisfy both requirements simultaneously.

In this study, we introduce a visual analytics system for calcium wave data to realize exploratory visualization in order to find and interpret ROIs. The proposed system enables users to (1) understand an overview of the relationships among various areas in calcium wave images, (2) select an ROI, and (3) visualize the detailed relationships between the selected ROI and other areas. The proposed system comprises an overview visualization that supports finding ROIs and a detail visualization to facilitate interpretation of the ROIs. The overview visualization uses cross-correlation analysis to calculate a directed graph that indicates the relationship of increased calcium concentration at various points in the oocyte. To make this directed graph easier to understand, we cluster it using an Infinite Relational Model (IRM) (Kemp et al., 2006). In addition, we used the Sugiyama framework (Sugiyama et al., 1981) to visualize the other graph to enable users to view the relationships between clusters. The detail visualization uses the cross-correlation analysis results to shows how the area selected by the users relates to the remaining areas. By connecting these visualizations, users can repeatedly find and interpret ROIs to form a valid hypothesis. In this paper, we demonstrate the effectiveness of the proposed visual analytics system by analyzing calcium waves in *Caenorhabditis elegans* (*C. elegans*).

2. Related Work

This section presents the related works on three closely related research topics.

Calcium Wave Analysis. The propagation of calcium waves during fertilization and signal transmission by calcium waves are common phenomena in many animals. Fabrizio et al. (2014) applied Granger causality (Granger, 1980) and cross-correlation to test and visualize the relationships between the measured intercellular calcium concentrations at 11 points in

the 30-somite stage embryo of a zebrafish. Buibas et al. (2010) formed networks of neurons and glia by computing the vector field obtained from calcium signaling imaging data using an optical flow technique. Milovic et al. (2013) developed a system that employed image processing to automatically recognize calcium waves generated during signal transmission between cells. In addition, they analyzed the wave-front velocity. The results of these studies indicated that time-series analysis and image analysis algorithms are effective for the analysis of calcium waves. However, these studies did not focus on finding and interpreting ROIs. In this study, we propose a visual analysis system that supports finding and interpreting ROIs based on the method proposed by Fabrizio et al. (2014). However, our study differs in that the number of locations where calcium waves are defined is approximately 200, which is significantly greater than the 11 points examined by Fabrizio et al. To make it possible for experts to understand the cross-correlation analysis results, we have developed a visual analysis system that includes multiple views and user interactions.

Visualization of Geographic Relational Data. Visualization of the origin-destination flow is related to the proposed system in that our approach utilizes a directed graph to depict the relationship at each point in the oocyte. Both approaches require the information to be displayed along with the directional geographic information. Several studies have investigated visualization of relational data using geographic positions. For example, Guo and Zhu (2014) proposed a method used to extract major flow patterns from large geographic mobility data to understand mobility flows. Da Lozzo et al. (2015) proposed a method used to draw a georeferenced graph to explore relationships using 2.5D visualization. Von Landesberger et al. (2012) proposed a visual analytics system for categorical spatiotemporal data that can be used to understand and explore geographic temporal movement of categorical data. However, these approaches do not focus on searching for characteristic areas using geographical relational data by considering the relationship of an ROI to other areas. The proposed system is designed to help experts explore such ROIs interactively.

Visualization and Visual Analytics for Relationship Analysis. Several studies have proposed visualization and visual analytics techniques that enable users to reveal various relationships. Zhang et al. (2015) proposed a visual analytics method used to support correlation analysis between multivariate data, and Wang and Mueller (2016) extended this system to support causality analysis based on regression analysis. In addition, Wang and Mueller (2016) indicated that their method could be extended to time-series data using Granger causality (Granger, 1980). However, the method proposed by Wang et al. cannot handle numerous time-series data easily. Köthur et al. (2015) proposed a visual analytics approach to perform windowed cross-correlation analysis (Boker et al., 2002) to overcome the difficulty associated with windowed cross-correlation in two time-series ensembles when more than two ensembles are present. Frey et al. (2012) proposed a visualization method used to explore similarity using spatiotemporal data. Their visualization illustrates the clustering of similar time-series data based on correlation. These methods help users understand and

explore data; however, few studies have considered the different relationships in spatiotemporal data. The proposed system helps users explore data in terms of similarities between a given ROI and remaining areas using cross-correlation analysis.

3. Data and Design Requirements

In this section, we describe the calcium wave data employed in this study. In addition, we present the design requirements to support calcium wave analysis. These requirements were determined in several meetings with two expert biologists.

3.1. Calcium Wave Data

We used the proposed system to analyze microscopic images of calcium waves obtained by Takayama and Onami (2016). In this study, we used images of the fertilization of the *trp*-3 mutant, which is *C. elegans* observed to be deficient in the sperm TRP-3 channel. This mutant was used because both the calcium wave and rapid local Ca^{2+} increase occur in the wild type, whereas only the calcium wave occurs in the *trp*-3 mutant. Thus, using the *trp*-3 mutant simplifies the analysis of wave propagation throughout the oocyte. The calcium ion concentration is visualized using a calcium indicator that increases fluorescence intensity in a high-calcium-concentration environment. The original data were created by imaging the fertilized egg using confocal microscopy, which helped us obtain a cross section of the oocyte. The z-axis where the sperm had high probability for fertilization was selected manually as the cross section of the oocyte.

To analyze these calcium wave images, Takayama et al. generated processed images of fertilized *trp*-3 mutant eggs. First, the region of the fertilized egg was extracted using a threshold value, and a manual correction was obtained from the microscopic images. Second, registration via affine transformation was performed using the StackReg ImageJ plug-in developed by Thévenaz et al. (1998)) by considering the movement of the fertilized egg within the body during microscopic observation. Note that the ImageJ is used extensively for image processing in biology research. This registration ensures that the fertilized eggs in all images are fixed to the same position. Third, Takayama et al. set the time fertilization occurred to 0 s and calculated the average fluorescence intensities in the oocyte F_0 in the range 1.8 to 0 s. Finally, they divided the fluorescence intensity F of each pixel at each time step by F_0 to calculate the ratio of fluorescence intensity. In addition, we applied a 3 × 3 window mean filter to the images to remove noise. Figure 1 shows an example of created images of relative fluorescence intensity F. As can be seen, the calcium waves are propagated from the lower right region, which is the sperm entry point. Note that the images were obtained at 0.2-s intervals.

3.2. Interests of Domain Experts and Design Requirements

Domain experts are interested in the fact that observation data differ from simulation data obtained under the assumption that the propagation of calcium waves follows a specific equation. Here, the objective is to determine the mechanism of calcium
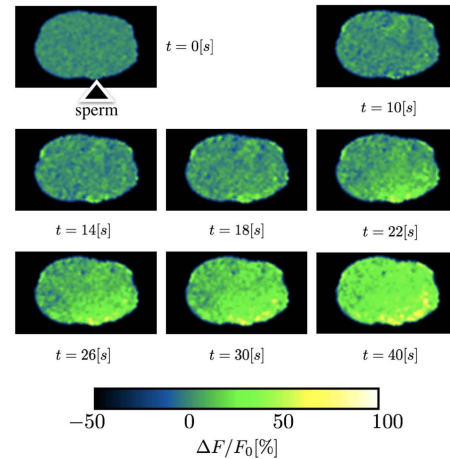

Fig. 1. Images of calcium waves in *trp*-3 mutant *C. elegans*

waves experimentally. To form a hypothesis for such experiments, it is necessary to understand the phenomena accurately. Therefore, a method that can obtain more detailed observations than those that can be obtained by simple imaging is required. After several meetings with the two biologists, we defined the design requirements to explore the observation data and form new hypotheses about calcium waves from the data. These design requirements are as follows.

R1. The ability to obtain an overview of how calcium waves propagate, thereby allowing experts to identify areas where the calcium wave flow differs from surrounding areas as a candidate ROI.

R2. The ability to understand the propagation of waves in a narrow area. To understand the behavior of the ROI, it is necessary to understand the flow in a narrow area because calcium waves occur in the chain of calcium released by the endoplasmic reticulum, which is distributed in the oocyte.

4. System Design

We designed the proposed visual analytics system based on R1 and R2 (Section 3.2). To design this system, we employed cross-correlation analysis, which is an effective method used to measure similarity and time delay between time-series data. Among existing methods used to test causal relationships, such as Granger causality (Granger, 1980), transfer entropy (Schreiber, 2000), and convergent cross mapping (Sugihara et al., 2012), we adopted cross-correlation analysis because the calcium waves in the oocyte comprise time-series data depicting a strong relation for each point.

The proposed system satisfies R1 and R2 simultaneously by depicting the relationships among the time-series data of the Ca^{2+} fluorescence intensity at each point calculated by cross-correlation analysis. The basic process flow of the proposed system is as follows.

Overview Visualization to find ROIs (meet R1)

To show clusters:
- Cluster View
- Adjacency Matrix View

To understand clusters:
- Network View
- Time Series Graph View

↓ Select ROI

Detail Visualization to interpret ROI (meet R2)

To interpret the relationship between ROI and other areas:
- Point-to-All View

To interpret the flow in a narrow area:
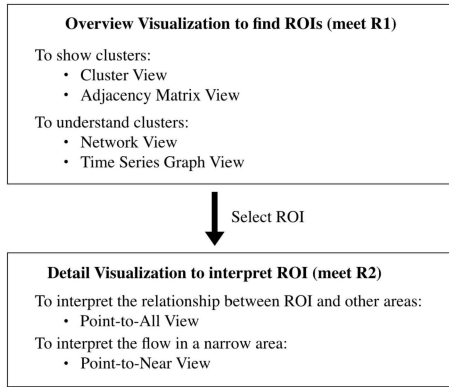- Point-to-Near View

Fig. 2. Each view in the proposed system and its role

Step 1. Extract the calcium concentration time-series data at each pixel in the oocyte from the microscopic image of the calcium waves shown in Figure 1 (Section 4.1.1).

Step 2. Conduct cross-correlation analysis for each set of extracted time-series data. Among two points, cross-correlation analysis is employed to determine the point that causes earlier increase in calcium concentration. Define each point as a node, and create a directed graph by drawing an arrow from the point where the concentration initially increases to the point where the concentration increases later (Section 4.1.2).

Step 3. Cluster the directed graph such that nodes with similar relations to the remaining nodes belong to the same cluster. The clustering result, which is represented as a new directed graph, represents an overview of the flow of calcium waves (Section 4.1.3).

Step 4. Create the overview visualization, which consists of multiple linked views, to enable domain experts to find the ROI from the clustered directed graph (Section 4.2).

Step 5. Interpret the ROI using detail visualization (Section in 4.3).

The overview visualization and detail visualization (Steps 4 and 5) comprise multiple views (Figure 2) that support user understanding. We describe each step, view, and user interaction in the following.

4.1. Data Preprocessing

Steps 1-3 are performed as preprocessing steps because calculating the cross-correlation among the time-series data of the Ca^{2+} fluorescence intensity and clustering the directed graph incur high computational cost. The workflow of the preprocessing steps is shown in Figure 3.

4.1.1. Sampling Time Series from Calcium Wave Data

To conduct cross-correlation analysis, we extracted the time-series data of calcium fluorescence intensity from the calcium wave images (128 × 96 *pixels*) in Figure 1. First, we extracted

1376 time-series data by sampling the scalar value of each pixel of the image at 2-pixel intervals along the x and y axes. Note that we removed time-series data for pixels without an oocyte. As a result, we obtained 523 time-series data. Here, the time step was 0.2 s, and the data length was 200 time steps, ranging from 1.8 s before fertilization to 38 s after fertilization. Note that the graph in Figure 3b is a superposition of all generated time-series data.

4.1.2. Cross-correlation Analysis

The second preprocessing step involves calculating an adjacency matrix using cross-correlation analysis. Cross-correlation is calculated as follows:

$$C_{XY}(lag) = \frac{\frac{1}{N}\sum_{t=1}^{N}(X(t) - \bar{X})(Y(t + lag) - \bar{Y})}{\sqrt{\frac{1}{N}\sum_{t=1}^{N}(X(t) - \bar{X})^2}\sqrt{\frac{1}{N}\sum_{t=1}^{N}(Y(t + lag) - \bar{Y})^2}} \quad (1)$$

where $X(t)$ and $Y(t)$ are the time-series data extracted from the calcium wave data, $\bar{X}$ and $\bar{Y}$ are the average values of $X(t)$ and $Y(t)$, respectively, N is the length of the time-series data, and *lag* is measured by the time step. The relationships between the time-series data of the Ca^{2+} fluorescence intensity are calculated as follows.

1. Calculate the cross-correlation by substituting every time step in the range 30 to 30 by the *lag* in Eq. 1.
2. Select the *lag* with the maximum correlation.
3. When the *lag* is 0 or greater and cross-correlation is greater than 0.7, define the relationship from X to Y. We assume that increased calcium intensity is propagated in the order $X \rightarrow Y$ if the cross-correlation becomes large when the *lag* is positive. This threshold is determined empirically under the assumption that there is a relationship only when there is a strong correlation.

By applying these three steps to all time-series pairs, we can calculate an asymmetric adjacency matrix (Figure 3c) that represents the relationship of rise in calcium concentration. The rows and columns of the adjacency matrix correspond to the time-series data. For the adjacency matrix, element (i, j) is 1 if there is a relationship between the time-series data of the i-th row and the j-th column, and element (i, j) is 0 if there is no relationship. In Figure 3c, 1 is represented as black, and 0 is represented as white.

4.1.3. Clustering the Adjacency Matrix

To obtain the overview of the relationships between time-series data, we clustered the calculated adjacency matrix because interpreting the original adjacency matrix (Figure 3c) is difficult. Here, we used the IRM proposed by Kemp et al. (2006) for clustering. The IRM is a clustering algorithm that can be applied to undirected and directed graphs. The IRM estimates the cluster automatically using the relational data such that nodes with similar relationships belong to the same cluster. Here, the IRM was calculated using the Python datamicroscopes library (https://github.com/datamicroscopes). The clustering result obtained by the IRM is shown in Figure 3d. The number of clusters is 34. It was automatically determined using the infinite relational model.
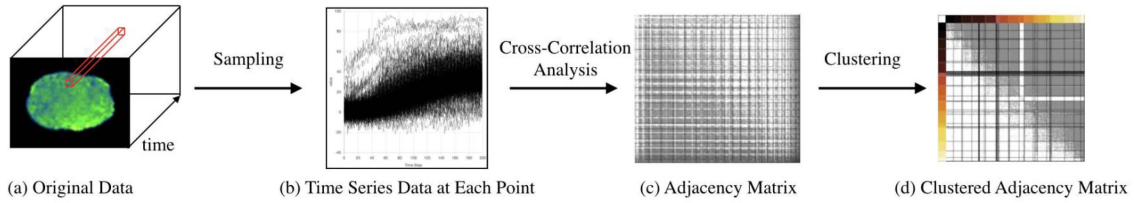
(a) Original Data   (b) Time Series Data at Each Point   (c) Adjacency Matrix   (d) Clustered Adjacency Matrix

Fig. 3. Preprocessing workflow

(a) Adjacency Matrix View

(b) Colormap of clusters

(c) Cluster View

(d) Network View

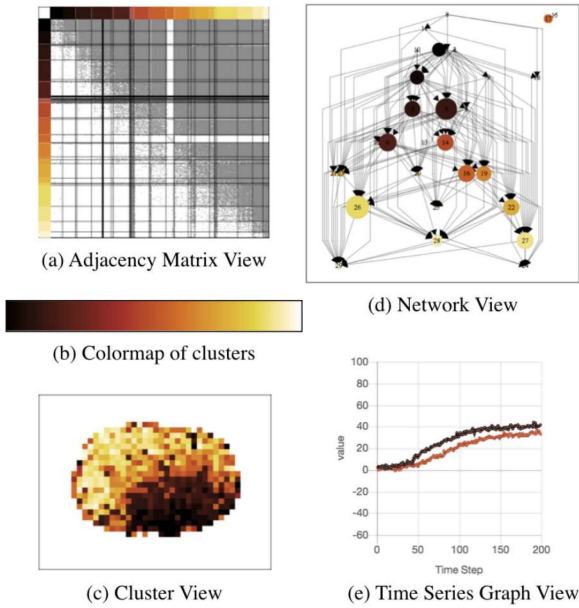(e) Time Series Graph View

Fig. 4. Overview visualization views

4.2. Overview Visualization

The overview visualization was designed to satisfy R1 (Section 3.2), and it contains four views that support data investigation and ROI determination, i.e., the Adjacency Matrix View (Figure 4a), Cluster View (Figure 4c), Network View (Figure 4d), and Time Series Graph View (Figure 4e). Note that these four views can be updated by user interactions.

4.2.1. Adjacency Matrix View

The Adjacency Matrix View (Figure 4a) shows the adjacency matrix calculated during preprocessing. This view is used to confirm the clustering result and the number of time-series data included in each cluster. To determine the order of the adjacency matrix, we introduced a criterion value δ based on the number of out-edges from node (δ^+) and in-edges from node (δ^-). Here, δ is determined as follows:

$$\delta = \delta^+ - \delta^- \qquad (2)$$

We determined the order of the adjacency matrix by sorting δ in descending order. Furthermore, we assigned colors to the clusters to identify corresponding clusters in other views. We selected the black-body radiation colormap (Borland and Taylor, 2007) shown in Figure 4b to represent the order of clusters

by color and to make each area easy to distinguish. The labels on the left and top of the Adjacency Matrix View represent the color of each cluster. Here, clusters are divided by black lines. In this view, users can select a cluster by clicking the label. The positional information of the selected cluster is highlighted in the Cluster View (Section 4.2.2).

4.2.2. Cluster View

In the Cluster View (Figure 4c), the color of each cluster is based on its coordinates. Note that the color of each node is the same as the cluster label in the Adjacency Matrix View.

With this view, users can evaluate the region belonging to the same clusters. Here, the positions of selected cluster are highlighted in blue when the user selects clusters in the Adjacency Matrix View and Network View. In this view, if the user selects (clicks) a point of interest, the detail visualization (Section 4.3) is displayed. Note that the selected point is highlighted in green.

4.2.3. Network View

As mentioned previously, it is difficult to interpret the relationships among clusters in the Adjacency Matrix View. Thus, the Network View (Figure 4d) visualizes these relationships using a directed graph. To define the relationships between clusters, we use the adjacency matrix calculated during preprocessing. The concept of the calculation is shown in Figure 5. If the connection between the cluster of the i-th row of the matrix and that of j-th column is greater than the threshold, we define a relationship from the i-th cluster to the j-th cluster. In the proposed system, we set the threshold to 90% of the maximum number of connections. This threshold is determined empirically to draw an edge only when there are many connections between clusters. We then calculate all relationships between clusters to generate a directed graph.

We use a graph layout algorithm to visualize this directed graph. Among various graph layout algorithms, we adopted the Sugiyama framework (Sugiyama et al., 1981) to visualize a hierarchy to help users understand the relationship between clusters. In addition, the Sugiyama framework was selected so that users can interpret the flow of calcium waves as an edge. Note that the direction of the flow of the calcium wave is reflected in the graph layout. Healy and Kuusik (2013) summarized the aesthetics of the Sugiyama framework as follows.

- "Edges should point in a uniform direction."

- "Short edges are more readable."
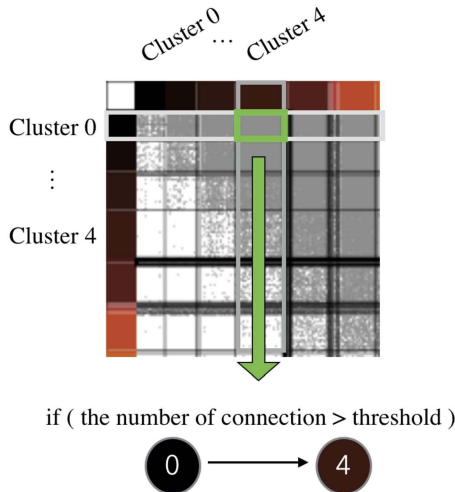
- "Uniformly distributed nodes avoid clutter."

Fig. 5. Calculation example of the relationship between clusters

- "Edge crossing obstructs comprehension."

- "Straight edges are more readable."

We implemented the Sugiyama framework using the Open Graph Drawing Framework (Chimani et al., 2014) in JavaScript.

In the Network View, each cluster is numbered from zero in descending order using a criterion value used to determine the order of the adjacency matrix. This cluster number is displayed on each node in the Network View. In addition, the color of each node is the same as the cluster label in the Adjacency Matrix View. The node size is proportional to the number of points in the clusters. In the Network View, it is possible to confirm a connection via a user interaction, such as changing the size and position of the nodes.

4.2.4. Time Series Graph View

The Time Series Graph View displays the mean time-series data of Ca^{2+} fluorescence intensity belonging to the selected clusters in the Network View or Adjacency Matrix View. The mean time series is calculated by estimating the mean value of the time-series data at each time step at all points in the cluster. Figure 4e shows the Time Series Graph View produced when cluster numbers 3 and 14 are selected in the Network View. Users can understand the features of the time-series data in the selected clusters with this view. Note that we employ Chart.js (http://www.chartjs.org/) to draw the graph.

4.3. Detail Visualization

In the Detail Visualization, the system visualizes the detailed relationships from the ROI to the remaining areas. The Detail visualization contains two views to satisfy R2 (Section 3.2). The two views (i.e., Point-to-All View and Point-to-Near View) shown in Figure 6 are displayed when a point in the Cluster View is selected. These views allow the user to assign meaning to the ROI and construct a hypothesis for its role.
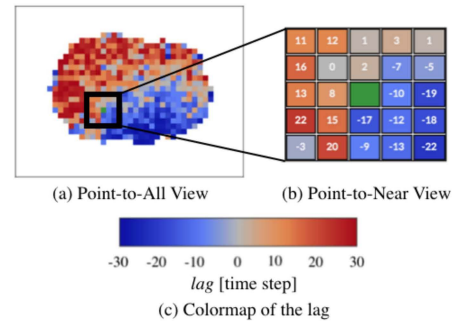

Fig. 6. Views in the detail visualization

4.3.1. Point-to-All View

The Point-to-all view displays the *lag* that maximizes the cross-correlation from the selected point to all other points. This *lag* is calculated using the step described in Section 4.1.2, and the *lag* size is indicated by the colormap shown in Figure 6. The red areas in Figure 6 indicate that this *lag* is a positive area, which suggests that an increase in calcium concentration occurs after the selected point. The blue region indicates that the increased calcium concentration occurs earlier. This view helps users evaluate the ROI by displaying the relationship between the ROI and the entire area as a *lag*.

In this view, a single point selected in the Cluster View is shown in green. Moreover, the user can select this view to set a new selected point. If the user selects a new point in this view, the corresponding points in the Cluster View are displayed in green.

4.3.2. Point-to-Near View

The Point-to-Near view shown in Figure 6 visualizes the relationships from a selected point to the surrounding 24 points. This is an enlarged view of the area near the point selected in the Cluster View or Point-to-All View. This view helps the user understand the flow in a narrow area (R2). The number of surrounding points to be displayed in the Point-to-Near View is determined empirically. The center green cell of this view indicates the selected area. Note that the *lag* that maximizes the cross-correlation is represented by both color and numbers in the Point-to-Near View.

4.4. User Interaction

In this section, we summarize the user interaction in the proposed system (Figure 7). In Figure 7, the solid line arrows indicate user operations, and the dotted line arrows indicate an operation that is calculated automatically and displayed by a user interaction.

In the proposed system, user interactions are divided into selecting the cluster to explore ROIs and selecting the ROI to interpret its role. In Figure 7, "①Click the label" and "②Click nodes" subtracted from User are operations used to select clusters in respective Views. When a cluster is selected in these views, the corresponding area is highlighted in blue in the Cluster View, and the mean time-series data of Ca^{2+} fluorescence intensity of the selected clusters is displayed in the Time Series

Graph View. Figure 7 shows the Cluster View and Time Series Graph View when cluster numbers 3 and 14 were selected in the Network View or Adjacency Matrix View. With these interactions, the user can explore the ROI in the Cluster View. "③Click one point" drawn from the User to Cluster View and Point-to-All View is used to select the found ROI. When the user selects the ROI in these views, the Point-to-All View and Point-to-Near View are displayed. These views facilitate interpretation of the ROI.

5. Application Example

In this section, we demonstrate how the proposed system supports hypothesis formation through analysis examples using calcium wave data (Section 3.1). Figure 8 shows a screenshot of the proposed system.

The proposed system juxtaposes the visualization results of actual and simple simulation data modeling calcium waves in each view because this functionality was required by domain experts. In each view shown in Figure 8, the actual data are on the left, and the simulation data are on the right. This juxtaposition helps an expert confirm differences between the actual data and the expert's knowledge because the simulation data reflect the experts existing knowledge. Moreover, original imaging data are displayed at the top left. The original imaging data can be used to check the original data. In this section, we first introduce the simulation data utilized in our analysis example. We then show the analysis results obtained using the proposed system.

5.1. Calcium Wave Simulation Data

We generated a simulation of calcium waves (Sneyd et al., 1998) using Nagumo's equation (Nagumo, 1962):

$$\frac{\partial c(x,y,t)}{\partial t} = \nabla \cdot (D \nabla c(x,y,t)) + c(x,y,t)(1 - c(x,y,t))(c(x,y,t) - \alpha) \quad (3)$$

where $c(x,y,t)$ is the calcium concentration at time t at coordinate (x,y), α is the threshold value, and D is the diffusion coefficient. The simulation result obtained using $D = 20\,\mu m^2/s$ and $\alpha = 0.1$ is shown in Figure 9. In this simulation, the diameter of the box in Figure 9 is $50\,\mu m$, $\Delta x = \Delta y = 1.0\,\mu m$, and time step $dt = 1.0 \times 10^{-4}$. In this simulation, the point at the upper left corner is the origin, right is positive along the x-axis, and down is positive along the y-axis. As the initial value of the simulation, the calcium concentration was set to 5 in the region satisfying $20 < x \leq 30$, $45 < y \leq 50$, and in other respects, it was 0. This high-calcium-concentration region represents the sperm entry point. It simulated the first rise in calcium concentration at the sperm entry point. This region is white in Figure 9 when $t = 0$. In this simulation, calcium waves propagate from the bottom. We created images of 300 time steps by sampling the simulation results at 0.1-s intervals. We extracted the time series data for each pixel from the simulation data, and preprocessing (Sections 4.1.2 and 4.1.3) was performed to create input data for the proposed system. Using the preprocessing discussed in Section 4.1, 289 time-series data were extracted from the simulation data.

5.2. Analysis Results Using Proposed System

In this section, we discuss the analysis results obtained using the proposed system (Figure 8). Note that the proposed system was developed as a web-based application. A domain expert used the system on the expert's own computer. To explain the system, a description of each view and the user interactions were provided (approximately 1.5 hours). Moreover, visualization researchers (authors of this paper) analyzed. We then discussed the analysis results. As a result, we found new characteristic points that the expert had not noticed previously. Moreover, new hypotheses suggest that these characteristic points comprise an "area where calcium release is occurring" or an "area where calcium is sucked by something." Consequently, points to pay attention to when analyzing the calcium wave became clear, and new experimental plans were established.

The analysis flow followed to obtain the above results using the proposed system is shown below.

5.2.1. Comparison between Actual and Simulation Data

The first analysis question from the expert was whether the calcium wave of the actual data flows in the same manner as that of the simulation data. We show an analysis result relative to this question using the proposed system.

In the proposed system, the size of each node in the Network View is proportional to the number of data in the cluster. Therefore, the Network View has nodes of various sizes, as shown in Figure 4d. Moreover, nodes corresponding to areas where the calcium concentration increases before other areas are placed at the top of the Network View. To compare the overview of the calcium wave flow obtained with the actual data and simulation data using the proposed system, we selected large nodes among those located at the top of the Network View and highlighted the corresponding areas. The Cluster View (Figure 10b, left) shows the view when cluster numbers 0 and 1 from the actual data (red circle at the left of Figure 10a) were selected. The Cluster View on the right of Figure 10b shows the view produced when cluster numbers 0 and 1 of the simulation data were selected. By comparing the Cluster View of the actual and simulation data (Figure 10b), it can be seen that the vicinity of the sperm entry point is blue in the actual data, and the lower part near the area with the high initial value is blue in the simulation data. Furthermore, the Cluster View visualized when cluster numbers 3 and 4 of the actual data and cluster numbers 2 and 3 of the simulation data (green in Figure 10a) were selected is shown in Figure 10c, and the Time Series Graph View is shown in Figure 10d. By comparing Figures 10b and 10c, it can be seen that, in addition to the area highlighted in Figure 10b, the area further from the sperm entry point is highlighted in Figure 10c. Moreover, from the time series graph shown in Figure 10d, the calcium concentration increased earlier with cluster number 0 (black line) than cluster 4 (brown line; actual data) and cluster 3 (red line; simulation data).

The following results were observed from these user interactions.

1. The proposed method can properly extract an overview of the flow of calcium waves near the sperm entry point.
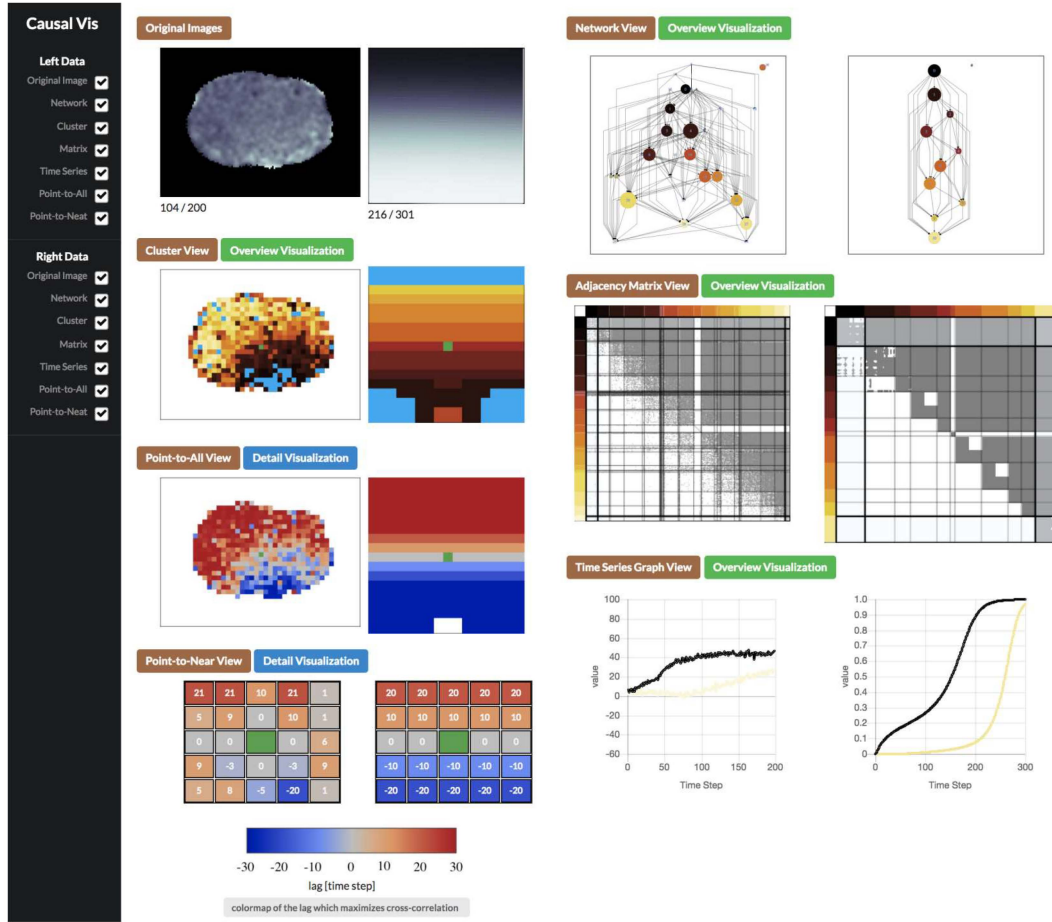
Fig. 7. Summary of user interactions in the proposed system. The Cluster View and Time Series Graph View show visualization results when cluster numbers 3 and 14 were selected in the Network View or Adjacency Matrix View

2. Compared to the simulation data, the highlighted areas of the real data are non-uniform.

From these results, we can conclude that the flow of calcium waves of the actual data is more complicated than that of the simulation data. The results of this analysis support the hypothesis that the actual data are non-uniform compared to the simulation data. Note that this result is consistent with the experts expectation.

5.2.2. Exploration of ROIs using the Proposed System

The second analysis question from the expert was where the characteristic points were. Characteristic points are where the flow of the actual data differs compared to simulation. Moreover, the expert was interested in the feature of those points. We show the result of exploring and interpreting these characteristic points (i.e., the ROIs) using the proposed system, as well as the hypothesis formed using the detail visualization.

We first explored ROIs using the Cluster View (Figure 10c). As can be seen in Figure 10c, there is an uncolored region near the sperm entry point in the actual data. This indicates that the behavior of this point differs locally. Thus, this area can be considered an ROI. Figure 11 shows the Detail Visualization when point ($x = 76$, $y = 69$) was selected. Here, the coordinates of the upper left corner of the image are $x = 0$, $y = 0$, and down is positive along the y-axis. Figure 11 shows an example of the detail visualization when we selected point ($x = 33$, $y = 25$) in the simulation data to compare the actual and simulation data. In the detail visualization of the simulation data, the Point-to-All View shows a flow from the bottom to top by color gradation (blue to red). Furthermore, the Point-to-Near View of the simulation data suggests that there exists a flow with a time difference of 10 time steps. On the other hand, in the actual data example, we interpret that the increased calcium concentration at the selected area came later than that in the surrounding area. As a result, we formed a new hypothesis that calcium is sucked by something for some reason in this characteristic point.

Next, characteristic point was explored by confirming the flow by further selecting nodes in the Network View of the actual data. To confirm the flow, we focused on nodes that meet both near the nodes selected in the Network View of the real data in Figure 10a (left side) and many connections from those nodes. Figure 12 shows the Cluster View when cluster number 14 was selected in addition to the nodes selected in Figure 10a. By comparing this Cluster View to the Cluster View in Figure 10c, it can be seen that the area near the sperm entry point was newly highlighted. On the other hand, some areas distant from it are also highlighted. The increased calcium concentration in these areas distant from the wave-front of the calcium waves is a phenomenon that has not been considered conventionally. Therefore, these points were investigated as ROIs. Among them, the points indicated by the arrows ($x = 40$, $y = 41$) in the Cluster View in Figure 12 are described as follows. The Point-to-Near View in Figure 12 was displayed by selecting the point ($x = 40$, $y = 41$). By interpreting this Point-to-Near View, in contrast to the Point-to-Near View in Figure 11, the periphery of the selected point is red, and most

Fig. 8. Screenshot of the proposed system. Each view is juxtaposed. Original images are displayed at the top left. This image was captured when the point (x = 64, y = 48) in the Cluster View (left side) and point (x = 25, y = 25) in the Cluster View (right side) were selected. In addition, the top and bottom rows in the Adjacency Matrix View were selected
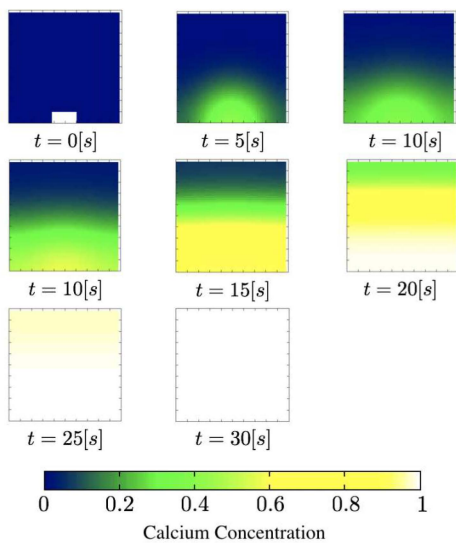

Fig. 9. Images of the simulation data of calcium waves

of the cross-correlation is maximized when the *lag* is greater than 0. Therefore, the increased calcium concentration occurs earlier than that in the areas surrounding this point. From this result, we formed the hypothesis that calcium release occurred for some reason in this area such that the increased calcium concentration occurred earlier than that in the surrounding areas.

As a result of this analysis, new ROIs were found and interpreted by exploring all views available in the proposed system (except for the Adjacency Matrix View). Moreover, new hypotheses were constructed using the Point-to-Near View. Prior to this analysis, hypotheses could not be formed sufficiently; thus, the expert could not develop an effective experimental plan. However, as hypotheses were formed using the proposed system, a new experimental plan was drafted to verify them. The new plan included examining whether the cell structure and characteristic points match using a marker that only highlights a specific organ in the cell.

6. Discussion

Although the proposed method facilitates forming hypotheses from calcium wave data, several issues must be discussed. The first is related to the advantages of the proposed system
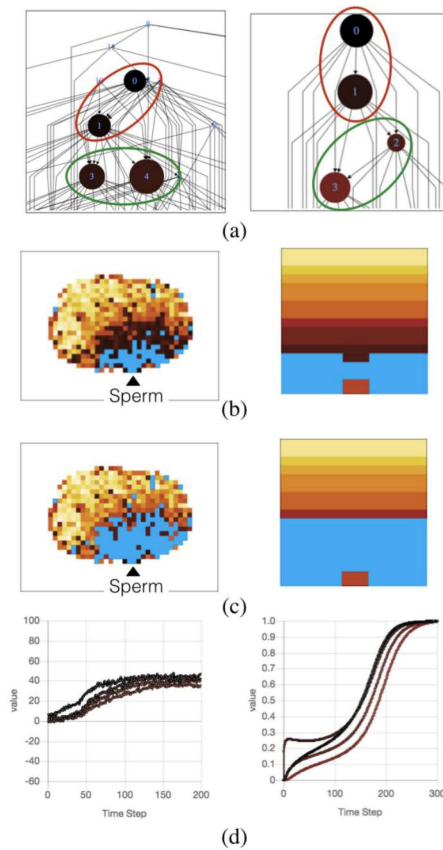
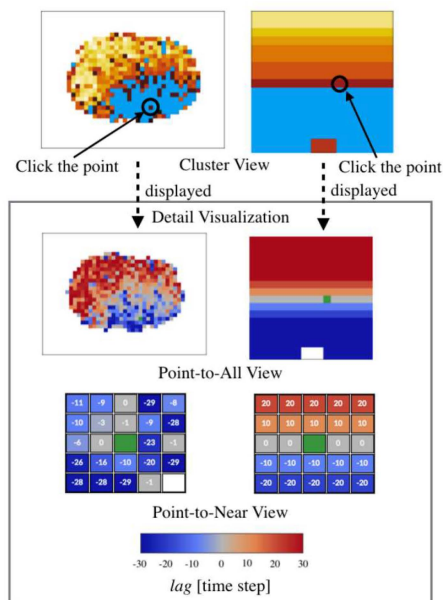Fig. 10. Analysis of calcium waves near sperm entry point to compare actual (left) and simulation data (right)
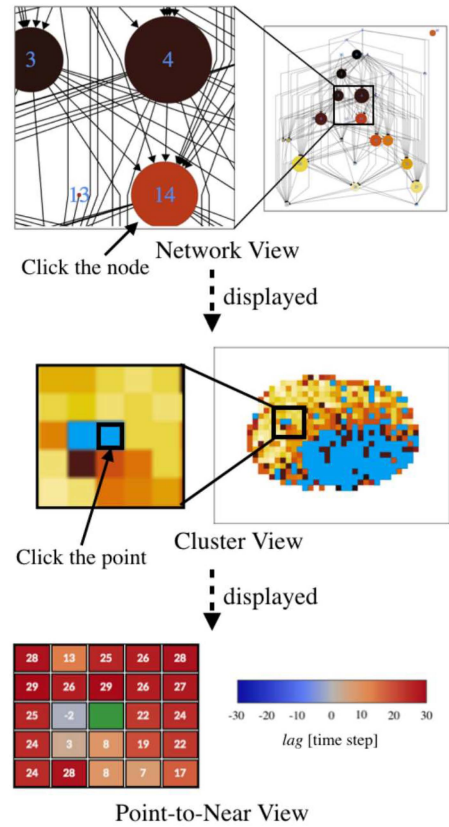

Fig. 12. Analysis example of finding a new ROI using the Network View. The Cluster View shows the result when cluster number 14 is selected after cluster numbers 0, 1, 3, and 4 were selected. The point indicated by the arrow in the Cluster View was selected as an ROI, and this ROI is interpreted using the Point-to-Near View.

Fig. 11. Analysis example of finding an ROI using the proposed system. The left column shows the visualization of the actual data, and the right shows the simulation data. The Detail Visualization is displayed when the point indicated by the arrow in Cluster View is selected.

compared to existing methods. The second is related to the effectiveness of the various views available in the proposed system. The third relates to the possibility of extracting incorrect characteristic points with the clustering algorithm.

A previous study that analyzed calcium waves applied cross-correlation to test and visualize the relationships between the measured intercellular calcium concentrations of 11 points (Fabrizio et al., 2014), and another study visualized the flow of calcium wave using Optical Flow (Buibas et al., 2010). Compared to these studies, the advantage of the proposed system is that it can be performed to find and interpret an ROI simultaneously by performing interactive analysis. Moreover, it is conceivable that the measured calcium wave data contains measurement error and noise. We consider that visual analytics in which analysis is performed while interpreting data is an effective way to prevent misinterpretation. Experimental plans can be improved based on the hypotheses formed using the proposed system by analyzing the results of a new experiment using the system. Therefore, the proposed system can promote the research cycle of forming and testing hypotheses.

In the application example described in Section 5, the Adjacency Matrix View of the proposed method was not used in the analysis. The comment regarding the Adjacency Matrix

View from the expert was, "Since there was no information on what kind of adjacency matrix will be shown when complicated waves occur, it is not useful for exploring ROI." Thus, to use the Adjacency Matrix View to explore characteristic points, it is necessary to investigate how waves with complicated behaviors are reflected in the adjacency matrix. The expert's comment about the Network View was, "If the vertical axis has meaning in the hierarchical graph displayed in the Network View, it is useful to explore ROI by assigning the meaning to each hierarchy. In particular, I am interested in what kind of phenomena the source and sink vertices in network view reflect." In this study, the Sugiyama framework was used to represent the flow of calcium waves in a hierarchical graph. To handle the problem indicated by the expert, we plan to implement a method (Onoue et al., 2017) that constrains the source and sink vertices in the Sugiyama framework such that they are located in the same layer. Moreover, in the current Network View, many edge crossings occur; therefore, visibility is degraded. To address this problem, edge bundling technology (Pupyrev et al., 2011) is considered effective.

In this study, we utilized the IRM (Kemp et al., 2006) as the clustering algorithm for the adjacency matrix (Section 4.1.3). Since the IRM performs clustering using a probabilistic approach, errors may occur in the clustering results. Thus, points that are not actually characteristic points may appear as characteristic points. In the proposed system, we think it is possible to prevent misinterpretation of the clustering results and characteristic points by interpreting the ROI using the Point-to-All View. On the other hand, to realize more accurate analysis, it would be effective to prevent misinterpretation by displaying all time-series data and the standard deviation of the selected clusters in the Time Series Graph View.

7. Conclusion

In this paper, we have proposed a visual analytics system to understand calcium wave data and form hypotheses. The proposed system was designed to meet the design requirements for finding and interpreting ROIs. We have demonstrated how the proposed system, which comprises multiple linked views, works using an analysis example. Moreover, we have demonstrated hypotheses relative to ROIs discovered using the proposed system. In summary, our exploratory visualization flow helps in achieving the following.

1. Observe an overview of calcium wave propagation through graph visualizations, time-series plots, and the position corresponding to the cluster.
2. Explore areas of interest.
3. Investigate the relationships between an ROI and other areas.

By exploring and interpreting ROIs using the above procedure, new hypotheses were formed to attach meaning to those points through discussion of the visualization results by domain and visualization experts. The hypotheses are that one is "there are areas where calcium release is occurring" and the other is "there are areas where calcium is sucked by something."

In the future, we are planning to evaluate the views and interactions required to form new hypotheses. By conducting an evaluation experiment using an eye tracker, we would like to clarify how the characteristic points are discovered by investigating how the proposed systems views and user interactions are used, and we expect to improve the proposed system according to the results of that investigation. In addition, by clarifying the effectiveness of each view and interaction, if the proposed method is utilized for time-series data from other fields, we consider that effective system design can be accomplished by selecting necessary views and user interactions.

Acknowledgments

References

Afrashteh, N., Inayat, S., Mohsenvand, M., Mohajerani, M.H., 2016. Optical-flow analysis toolbox for characterization of spatiotemporal dynamics in mesoscale optical imaging of brain activity. NeuroImage 153, 58–74. doi:10.1101/087676.

Boker, S.M., Rotondo, J.L., Xu, M., King, K., 2002. Windowed cross-correlation and peak picking for the analysis of variability in the association between behavioral time series. Psychological Methods 7, 338–355. doi:10.1037//1082-989X.7.3.338.

Borland, D., Taylor, R.M., 2007. Rainbow color map (still) considered harmful. IEEE Computer Graphics and Applications 27, 14–17. doi:10.1109/MCG.2007.323435.

Buibas, M., Yu, D., Nizar, K., Silva, G.A., 2010. Mapping the spatiotemporal dynamics of calcium signaling in cellular neural networks using optical flow. Annals of Biomedical Engineering 38, 2520–2531. doi:10.1007/s10439-010-0005-7, arXiv:0912.0265.

Chimani, M., Gutwenger, C., Michael, J., Klau, W.G., Klein, K., Mutzel, P., 2014. The Open Graph Drawing Framework (OGDF), in: Tamassia, R. (Ed.), Handbook of Graph Drawing and Visualization. chapter 17, pp. 543–569.

Da Lozzo, G., Di Bartolomeo, M., Patrignani, M., Di Battista, G., Cannone, D., Tortora, S., 2015. Drawing georeferenced graphs: Combining graph drawing and geographic data. IVAPP 2015 - 6th International Conference on Information Visualization Theory and Applications; VISIGRAPP, Proceedings , 109–116doi:10.5220/0005266601090116.

Fabrizio, D.V.F., Corazzol, M., Sternberg, J., Wyart, C., Chavez, M., 2014. Hierarchy of Neural Organization in the Embryonic Spinal Cord: Granger-Causality Graph Analysis of Calcium Imaging Data. Neural Systems and Rehabilitation Engineering, IEEE Transactions on PP, 333–341. doi:10.1109/TNSRE.2014.2341632.

Frey, S., Sadlo, F., Ertl, T., 2012. Visualization of temporal similarity in field data. IEEE Transactions on Visualization and Computer Graphics 18, 2023–2032. doi:10.1109/TVCG.2012.284.

Granger, C.W.J., 1980. Testing for causality: A personal viewpoint. Journal of Economic Dynamics and Control 2, 329–352. doi:10.1016/0165-1889(80)90069-X.

Guo, D., Zhu, X., 2014. Origin-destination flow data smoothing and mapping. IEEE Transactions on Visualization and Computer Graphics 20, 2043–2052. doi:10.1109/TVCG.2014.2346271.

Healy, P., Kuusik, A., 2013. Hierarchical Drawing Algorithms, in: Handbook of Graph Drawing and Visualization. CRC Press, pp. 409–454.

Kemp, C., Tenenbaum, J.B., Griffiths, T.L., Yamada, T., Ueda, N., 2006. Learning systems of concepts with an infinite relational model. Proceedings of the 21st AAAI conference 21, 381–388.

Köthur, P., Witt, C., Sips, M., Marwan, N., Schinkel, S., Dransch, D., 2015. Visual Analytics for Correlation-Based Comparison of Time Series Ensembles. Computer Graphics Forum 34, 411–420. doi:10.1111/cgf.12653.

Milovic, C., Oses, C., Villalón, M., Uribe, S., Lizama, C., Prieto, C., Andia, M.E., Irarrazaval, P., Tejos, C., 2013. Calcium (Ca2+) waves data calibration and analysis using image processing techniques. BMC Bioinformatics 14, 1. doi:10.1186/1471-2105-14-162.

Nagumo, J., 1962. An Active Pulse Transmission Line Simulating Nerve Axon, in: Proceedings of the IRE, pp. 2061–2070.

Onoue, Y., Kukimoto, N., Sakamoto, N., Misue, K., Koyamada, K., 2017. Layered graph drawing for visualizing evaluation structures. IEEE Computer Graphics and Applications 37, 20–30. doi:10.1109/MCG.2016.40.

Pupyrev, S., Nachmanson, L., Kaufmann, M., 2011. Improving layered graph layouts with edge bundling. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 6502 LNCS, 329–340. doi:10.1007/978-3-642-18469-7_30.

Schreiber, T., 2000. Measuring information transfer. Physical Review Letters 85, 461–464. doi:10.1103/PhysRevLett.85.461.

Sneyd, J., Dale, P.D., Duffy, A., 1998. Traveling Waves in Buffered Systems: Applications to Calcium Waves. SIAM Journal on Applied Mathematics 58, 1178–1192. doi:10.1137/S0036139996305074.

Stricker, S.A., 1999. Comparative biology of calcium signaling during fertilization and egg activation in animals. Dev Biol 211, 157–176. doi:10.1006/dbio.1999.9340.

Sugihara, G., May, R., Ye, H., Hsieh, C.h., Deyle, E., Fogarty, M., Munch, S., 2012. Detecting Causality in Complex Ecosystems. Science 338, 496–500. doi:10.1126/science.1227079.

Sugiyama, K., Tagawa, S., Toda, M., 1981. Methods for Visual Understanding of Hierarchical System Structures. IEEE Transactions on Systems, Man and Cybernetics 11, 109–125. doi:10.1109/TSMC.1981.4308636.

Takagaki, K., Zhang, C., Wu, J.Y., Ohl, F.W., 2011. Flow detection of propagating waves with temporospatial correlation of activity. Journal of Neuroscience Methods 200, 207–218. doi:10.1016/j.jneumeth.2011.05.023.

Takayama, J., Onami, S., 2016. The Sperm TRP-3 Channel Mediates the Onset of a Ca2+ Wave in the Fertilized C. elegans Oocyte. Cell Reports 15, 625–637. doi:10.1016/j.celrep.2016.03.040.

Thévenaz, P., Ruttimann, U.E., Unser, M., 1998. A pyramid approach to subpixel registration based on intensity. IEEE Transactions on Image Processing 7, 27–41. doi:10.1109/83.650848.

Von Landesberger, T., Bremm, S., Andrienko, N., Andrienko, G., Tekušová, M., 2012. Visual analytics methods for categoric spatio-temporal data. IEEE Conference on Visual Analytics Science and Technology 2012, VAST 2012 - Proceedings , 183–192doi:10.1109/VAST.2012.6400553.

Wang, J., Mueller, K., 2016. The Visual Causality Analyst: An Interactive Interface for Causal Reasoning. IEEE Transactions on Visualization and Computer Graphics 22, 230–239. doi:10.1109/TVCG.2015.2467931.

Whitaker, M., 2006. Calcium at fertilization and in early development. Physiological Reviews 86, 25–88. doi:10.1152/physrev.00023.2005.

Yamashita, Y., Harada, T., Kuniyoshi, Y., 2012. Causal flow. IEEE Transactions on Multimedia 14, 619–629. doi:10.1109/TMM.2012.2191396.

Zhang, Z., McDonnell, K.T., Zadok, E., Mueller, K., 2015. Visual correlation analysis of numerical and categorical data on the correlation map. IEEE Transactions on Visualization and Computer Graphics 21, 289–303. doi:10.1109/TVCG.2014.2350494.