In Search of Patient Zero:
Visual Analytics of Pathogen Transmission Pathways in Hospitals

T. Baumgartl*, M. Petzold*, M. Wunderlich, M. Höhn, D. Archambault, M. Lieser, A. Dalpke, S. Scheithauer, M. Marschollek, V. M. Eichel, N. T. Mutters, *HiGHmed Consortium*, and T. von Landesberger
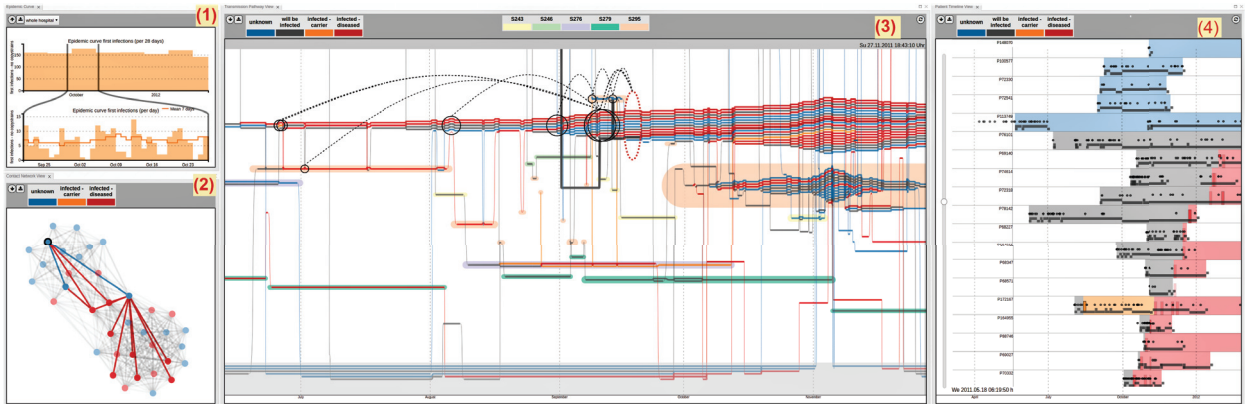
Figure 1: Transmission pathway tracing interface. 1. The *Epidemic Curve View* shows the number of infected patients. 2. The *Contact Network View* shows patient contacts as a network. 3. The *Transmission Pathway View* shows contacts and infection status. Tracing interaction shows potential infection transmission events. 4. Patient event details are shown in *Patient Timeline View*.

Abstract—Pathogen outbreaks (i.e., outbreaks of bacteria and viruses) in hospitals can cause high mortality rates and increase costs for hospitals significantly. An outbreak is generally noticed when the number of infected patients rises above an endemic level or the usual prevalence of a pathogen in a defined population. Reconstructing transmission pathways back to the source of an outbreak – the *patient zero* or *index patient* – requires the analysis of microbiological data and patient contacts. This is often manually completed by infection control experts. We present a novel visual analytics approach to support the analysis of transmission pathways, patient contacts, the progression of the outbreak, and patient timelines during hospitalization. Infection control experts applied our solution to a real outbreak of *Klebsiella pneumoniae* in a large German hospital. Using our system, our experts were able to scale the analysis of transmission pathways to longer time intervals (i.e., several years of data instead of days) and across a larger number of wards. Also, the system is able to reduce the analysis time from days to hours. In our final study, feedback from twenty-five experts from seven German hospitals provides evidence that our solution brings significant benefits for analyzing outbreaks.

Index Terms—dynamic networks, visualization applications, health, medicine, outbreak, Klebsiella, infection control

◆

1 INTRODUCTION

Pathogen transmissions are an acute problem in hospitals around the world [20, 24]. The transmission of pathogens, such as bacteria and

- * *T. Baumgartl and M. Petzold are shared first authors.*
- *T. Baumgartl, M. Wunderlich, M. Höhn and T. von Landesberger are with TU Darmstadt, Darmstadt, Germany. T. von Landesberger is now with Universität Rostock, Germany. E-mail: office@gris.tu-darmstadt.de.*
- *D. Archambault is with Swansea University, Swansea, United Kingdom E-mail: D.W.Archambault@Swansea.ac.uk.*
- *M. Lieser, M. Petzold, V. M. Eichel, N. T. Mutters are with University Hospital Heidelberg, Heidelberg, Germany, E-mail: {first.last}@med.uni-heidelberg.de.*
- *A. Dalpke is with TU Dresden, Dresden, Germany. E-mail: alexander.dalpke@ukdd.de.*
- *S. Scheithauer is with University Medicine Göttingen, Universität Göttingen, Germany. E-mail: simone.scheithauer@med.uni-goettingen.de.*
- *M. Marschollek is with L. Reichertz Institute for Medical Informatics, Hannover, Germany. E-mail: Marschollek.Michael@mh-hannover.de.*
- *This paper is contributed by HiGHmed infection control study group.*

viruses, can endanger the lives of patients, since they represent a very vulnerable group of persons. Pathogen infections in hospitals can be transmitted via patient-to-patient contact in the same room or ward [18]. Transmissions are generally difficult to detect since patients may be infected, but show no clinical symptoms. These *carriers* are invisible sources for potential transmissions. Screenings can detect carriers, but regular screening of all patients is very costly and ineffective. Thus, only high risk patients are usually screened when entering the hospital or after transfers within the hospital.

One initial infected patient, a *patient zero* or *index patient*, may transmit the pathogen to other patients [20]. These patients may change wards and infect further patients. The endemic level represents the usual prevalence of a pathogen in a defined population. When the number of infected patients rises above a certain endemic level, and a spatio-temporal context is observed, it is called *cluster* or *outbreak*. The endemic level is characterized by parameters such as the type of patients and the location within the hospital (underlying diseases), as well as the pathogen specifics: e.g., the mode of transmission, seasonality of occurrence, the resistance potential and infectiousness. As automatic outbreak detection is still an ongoing research problem [21, 83], usually, the endemic level is set manually by comparing numbers of newly infected patients within a certain time frame with the numbers of infected patients in a previously recorded time frame. As not all patients can be regularly tested or screened, depending on the pathogen,

it can take days, weeks or months before an outbreak is detected [27]. Therefore, the monitoring of infectious persons and determining their transmission pathways is the primary goal of Infection Control experts, e.g., hygienists, clinicians, hygiene experts, in order to intervene in time and prevent further pathogen spreading.

Once an outbreak has been detected (Task 1), the infection control experts need to trace the infection back to its source in order to determine if the patients are connected and belong to an outbreak. This means the identification of all potentially infected patients back to the *patient zero*, i.e., the overall source of the outbreak (Task 2). The experts need to reconstruct these *transmission pathways* – by whom, when and where did a transmission occur. This information is used to find the origin of the infectious agent: whether it is nosocomial – hospital-intern. Furthermore, patient localization (which rooms/wards – Task 3) and outbreak duration (begin/end – Task 4) need to be determined. The transmission pathway is then used to identify putative colonized or infected patients that are yet unknown and thus require testing (Task 5). Intervention procedures can then be implemented by isolating or cohorting affected patients in separate rooms, special disinfecting processes, increasing hand hygiene, screening and teaching staff affected: infected and exposed.

The identification of transmission pathways is challenging because it can occur over several contacts, involve several hospital wards and various time spans (days to years). Tracing requires integrating both the spatio-temporal patient information and microbiological test results. Outbreak pathway reconstruction needs to be fast and precise: each potential patient and their contacts need to be detected to prevent further disease spreading. Often there is uncertainty about the patient status at the time of a contact, as screening often detects an infection or colonization only at a later date. Thus, also potential infections need to be considered.

Currently, transmission pathway reconstruction is a time-consuming and potentially error-prone manual process. It may take days to weeks, using current hospital systems [27]. Visual analytics systems have the potential to support this analysis process by saving analyst time [15]. However, current solutions focus mainly on the disease evolution at a population level (see Sect. 2).

In cooperation with Infection Control experts from four German Hospitals and Infection Research Institutes, we developed a novel visual analytics system for the exploration of disease outbreaks. We used iterative, user-centered design within a common project HiGHmed over two years. The system offers several specialized views for the exploration of outbreaks and pathogen transmissions. The core contribution is a novel view for contract tracing that was inspired by the well known storyline visualization [37, 46, 52]. The visual design, layout and interactions enable to explore contacts as well as to automatically determine and highlight patients and their contacts that could transmit pathogenic organisms.

Our approach was applied to a real outbreak of *Klebsiella pneumoniae* in a large German hospital. The experts were able to effectively reconstruct the transmission pathway back to the *patient zero* in a faster and in a more comprehensive way when compared to existing methods. In our final qualitative study, we gathered feedback from twenty-five experts located in seven German hospitals. The results indicated significant added value when tracing transmission and analyzing outbreaks in hospitals.

2 RELATED WORK

Visual Analytics of Health Data. Visual Analysis of healthcare data is an active area of research. Event-based visual analytics approaches for health record data has been a main focus of this area [13, 26, 28, 41, 53, 54, 57]. Disease surveillance and epidemiology has also been of interest [49]. Visual analytics approaches for epidemiologists focus on the spatio-temporal evolution of a disease at a population level–how many people will be infected, in which geographic area(s), and the speed of the spread [12, 39, 82]. These visual analytics approaches target disease spreading over a large population at a macroscopic level. They visualize the number of people infected over time, but detailed views about individual patient contacts are not the focus. Machine learning methods working with visual analytics have been proposed [16, 34], and more specifically for disease progression pathways [33]. They have been applied to the problem of infection control [42], but in this closely related work the transmission pathway was not reconstructed. Machine learning is effective, but may not be feasible for rarely occurring pathogens, such as *K. pneumoniae*, where the numbers of items in the trained classes of (non)-infection are very biased. Thus, transmission pathway reconstruction and the identification of potentially infected patients requires interactive exploration.

Disease Spread and Dynamic Networks. As many pathogens are transmitted over contacts preferentially, disease spreading is often simulated as dynamic processes over contact networks. A number of approaches have focused on the population level: the number of infected patients through line charts [1, 18, 36, 56]. They use the contact networks in the simulation, but only provide macroscopic views of disease progression. The simulations often result in a collection of thousands of dynamic graphs with very few existing approaches for this data [10, 38]. Manynets [22], GraphLandscape [29] and SOM-based clustering [75] extract graph properties and compare many static graphs and these methods could be adapted and applied. However, two very different graphs can have the same properties [14]. Network piling approaches [5, 71] allow the comparison of several static graphs in detail but do not scale well to larger graphs and have not been adapted to compare multiple dynamic graphs. Recently, a simulation and visualization of disease spreading over contact networks in hospitals has been presented [80]. It shows the disease and transmission probability on individual level, but takes only a static network as input. This simplification is done for simulation purposes. However, in real cases, patients move across wards over time. Both visualization and simulation should take into account the network dynamics over time. In sum, however, the simulations focus only on disease spreading prediction and not on the reconstruction of pathogen transmission pathways.

Visualization of Contact Dynamics. The reconstruction of pathogen transmission pathways requires analyzing patient contacts and patient infection status over time. A number of methods exist for visualizing a single dynamic graph over time in a scalable way [6, 55, 69, 73]. These methods use timeslices as a basis for the visual analytics process. The uniformity of timeslices poses a problem as the distribution of transmission events is unequal over time. Event-based methods [41, 50] for the visualization of dynamic networks [35, 63, 64, 78] are more applicable. However, they have not been designed for outbreak networks visualization in a way that fulfills the needs of the infection control experts.

The visualization of contact network dynamics for hospital data could be supported by dynamic set visualization and storyline approaches. Set visualization methods [2] for the visualization of set dynamics [45, 72] are able to show the set sizes and the number of changes between sets over time. Methods for visualizing long time series [7] and text visualizations [17, 59] have similar representations. Showing the group membership over time on individual level has also motivated the storyline research. Storyline visualizations [37, 46–48, 67, 68] devote one dimension, usually the x-axis, to time and encode each individual character in a story as a line. Lines that are placed close together to indicate the characters share a scene; when the lines separate, the scene ends. For disease spreading, this encoding can be used to show contacts in the same ward and forms a key part of our approach. Research in storyline visualization has focused on optimizing the compactness of storyline visualizations (either automatic or users-assisted) [4, 23, 37, 47, 48, 51, 61, 62, 67, 68], reducing crossings [25, 32, 70, 79], plotting approaches [60], combining storylines with event-based methods [3], genealogical data [31], streaming and dynamic data [66, 81], and contacts between living things or actors exhibiting similar behavior [52]. Reda *et al.* [52] is the closest approach to ours, but it needs to consider all contacts in the storyline. In our work, we can reduce clutter by determining and highlighting pathogen transmitting contacts. On the other hand, we need to show more specific locations inside the hospital or if the patient is discharged.

3 BACKGROUND – TASKS AND DATA

Over the course of one year, we investigated the data and tasks that are performed by infection control experts in four German hospitals during their procedures. The data and tasks were guided by the scope of the HiGHmed project. For a deep understanding of the data, tasks and current analysis methods, we conducted a structured interview following the methodology in [74]. We first used an online questionnaire that was answered by six infection control experts and epidemiologists. Moreover, we interviewed three hygienists to learn the current work methods by infection control experts in three hospitals and one infection control institute. Further details on the tasks and data were assessed during the iterative design process (see Sect. 4). Our system is designed to support one of their main tasks: tracing transmission pathways in outbreaks.

3.1 Tasks

Our work resulted in these tasks without a strict workflow order:

T1 *Detect Outbreak.* Is there an outbreak? When the number of infected patients rises above a normal level within a certain period, i.e., the endemic level, an *outbreak* occurs. Depending on the pathogen, this endemic level can be two or more patients. As patients may not be tested or screened generally, an outbreak is determined by manual inspection.

T2 *Outbreak Pathway.*

T2.1 *Determine transmission contacts.* Did contacts occur between patients that could have led to pathogen transmission? If yes, where, when and with whom?

T2.2 *Determine index patient (patient zero).* Who is the source of the outbreak? Identifying the transmission pathway should lead back to the index patient, i.e., the patient zero.

T2.3 *Distinguish between a single or multiple outbreaks.* Is the observed outbreak a single outbreak or multiple, simultaneous outbreaks of similar pathogens? Depending on the answer, there could be one or more index patients.

T3 *Outbreak Location.*

T3.1 *Determine if the outbreak is hospital-associated.* Did the transmissions occur within the hospital or have been infections brought in from outside the hospital? A within-hospital transmission, i.e. a nosocomial, points to a source of infection such as a patient/staff/device. These transmissions should be avoided as patients can have lower immune response and thus may be more vulnerable.

T3.2 *Locate ward(s) with pathogen transmissions.* For nosocomial transmissions, in which ward has the transmission occurred? This location needs to be disinfected or closely monitored.

T4 *Quantify outbreak duration.* When did the outbreak first start? How long does it last? Early detection mitigates the risk of larger outbreaks. Tracing back the time point of all transmissions to the initial index patient determines the total duration.

T5 *Identify potentially infected patients.* Who is also potentially infected? Not all patients can be screened in the hospital to determine their infection, and not all infected patients show symptoms. These potentially (i.e., putative) infected patients need to be identified via their contacts to infected patients.

3.2 Data

The pathogen transmission analysis requires to combine two types of data 1) patient locations for determining contacts, and 2) microbiological data for identifying infection status. Due to privacy reasons, collection of this data is limited (see Sect. 8). The patient location is only determined by a log record of their ward, not by tracking. The infection status is only known at the time of screening or test. These data sets can span years with events recorded down to second precision. The time between two consecutive events (e.g. transfers) can be on the order of seconds to months.

Patient Locations: The location data consists of a list of patient transfer events $TR = \{TR_k\}$. A transfer TR_k records that the patient P_k, was transferred to location L_k at time t_k^{TR} for the following reason $type_k$.
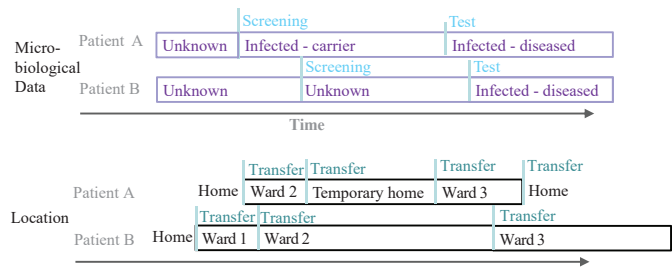

Figure 2: Data characteristics. Bottom: Patient locations determined by transfer events. Top: Infection status of patients over time determined by microbiological data events.

Thus, a transfer is $TR_k = (P_k, L_k, t_k^{TR}, type_k)$. The $type_k$, determines whether the transfer was 1) between wards, 2) from home to hospital, i.e., first hospital admission, or 3) home between hospital stays, i.e., 'temporary home', or 4) was the end of a hospital stay. The current location of the patient is the destination of the patient's last transfer (see Fig. 2). The time intervals between successive transfers are irregular and cannot be transformed to regular intervals without sampling or without causing scalability challenges. Even very short stays at a location can lead to important contacts in pathway reconstruction.

Microbiological data: MB is information about the tests and screenings of patients for pathogens. More formally, it is a set of events $MB = \{MB_j\}$. Each element MB_j records which patient P_j, was tested or screened s_j, for which pathogen ρ, at what time t_j^{MB} and the result r_j: $MB_j = (P_j, s_j, t_j^{MB}, \rho_j, r_j)$. Patients may have several microbiological data records associated with them or none at all. The analysis can take place before, during or after the hospital stay and the spacing between events is irregular.

The result r_j determines patient's *infection status* (see Fig. 2):

- *infected – carrier*: A positive result of a screening means that the patient is colonized. Due to lack of data on recovery (see Sect. 8) from this moment onwards, the patient retains this status unless he/she is later identified as diseased.
- *infected – diseased:* If a patient with symptoms is tested positively on the pathogen ρ, he/she is in a diseased state. Due to lack of data on recovery (see Sect. 8), the patient retains this status from the moment of the positive test onwards.
- *unknown:* Before the first positive screening or targeted microbiological test, the infection state is *unknown*. If the patient is not tested for the pathogen ρ, it is unknown whether they are infected/colonized. If a patient becomes infected at a later point in the data set, the patient is labeled *"unknown – will be infected"*.

4 DESIGN PROCESS AND FINAL APPROACH

Our approach was developed through iterative design with our domain experts over two years following guidelines for effective visualization design [30, 43, 58]. After the initial task and data analysis, we had regular quarterly meetings with project participants. The number of participants in the meetings varied as not all participants could attend each meeting. On average, ten of our expert participants were present in our meetings. The expertise included: infection control, hygienists, infection control data management, epidemiologists and infection control managers. Between the meetings, we communicated per email and held interactive sessions onsite.

4.1 Prototype Development

Our *first prototype* had several linked views [76] (see Fig. 3a). The epidemic curve view showed the number of infected patients over a two weeks period (see Fig. 3a–part 1,2). It highlighted potential outbreaks with infections above the endemic level per selected pathogens. For identifying hospital-associated (nosocomial) infections, the patient stay along with infection status was shown (see Fig. 3a–part 3). Potential transmissions between patients were supported only by node-link

(a) Initial visualization, see also [76]
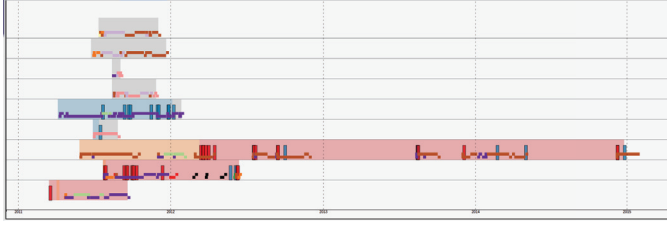
(b) Patient timeline view.
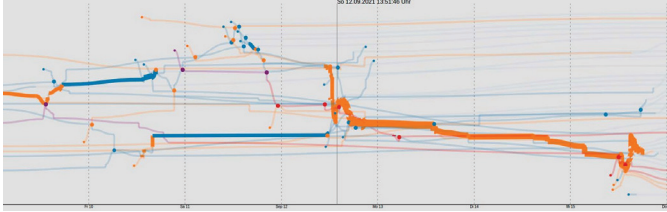
(c) Initial storyline-like design

Figure 3: Early prototypes. Interfaces were translated into English.

diagrams showing patient contacts and current infection status (see Fig. 3a–part 4). A total of thirteen domain experts responded to our call for feedback by filling out an online questionnaire [76]. The prototype demonstrated a high degree of usefulness (mode 4 and 5 respectively on a five point Likert scale). The experts especially appreciated views that were new to them: a) patient stay and the infection duration and b) the contact network. They wanted these views to scale to longer periods of time, larger sets of microbiological data, and views to support pathogen transmission pathway reconstruction.

The *second design* included a *Patient Timeline View* that showed the patient stay and his/her microbiological data (see Fig. 3b). This design leveraged the infection control expert's experience with excel-based information. Each row is one patient, and the x-axis is time. The background color shows the infection status, colored horizontal bars show the patient location, and colored vertical bars show the screening and test results. Our experts found that this view provided a good overview of patient location and infection status. However, the separation of patients into rows made it difficult to spot contacts leading to pathogen transmission. Sorting the rows by the time of the first infection did not help, as a patient can have contact with several patients.

Our *third design* used a storyline representation of the infection data (see Fig. 3c). Patients are lines and the x-axis is time. Lines are colored according to infection status. Contacts in this view are line bundles. Although the view shows all data needed for pathogen transmission analysis, the experts had difficulty relating to it. The layout that emphasized the movements made it difficult to follow patients and to determine transmissions.

4.2 Final Prototype

The *final visual interface* is shown in Fig. 1. This design combines the pathogen pathway reconstruction views of the user-selected pathogen needed for our tasks (see Sect. 3.1).

1. *Epidemic Curve View* shows the number of infected persons per day in order to support Task 1 – outbreak detection. The infection control expert can select the total number of infections or only new ones (i.e., without copystrains). To see how it relates to the endemic level, a moving average of the user-selected time period is shown. This view can show data for the hospital (see Fig. 9 top) or specific wards (see Fig. 9 bottom). It supports longer time periods via focus-and-context methods inspired by [77].
2. *Contact Network View* shows the contacts of selected patients for determining putative infected patients (Task 5).
3. *Transmission Pathway View* supports Tasks 2–4. This view shows patient contacts and possible pathogen transmissions over time in a storyline-like view. The extensions are layout, design, and highlighting of potential transmission events more directly.
4. *Patient Timeline View* shows the patient location overlaid with microbiological data. It supports the visual analytics process with infection status and location information. The data encoding is unchanged since its early prototype described in Sec. 4.1.

5 Transmission Pathway View

The *Transmission Pathway View* shows patient infection status and contacts over time and across locations (Fig. 1 (3)). Outbreak duration, potential transmission contacts between patients, and patient locations are visible (Task 2–4).

Each line represents a patient and the x-axis encodes time (Task 4). Each patient line starts with the earliest recorded admission to hospital and ends with the last recorded stay. Temporary home stays are also shown. This helps in detecting hospital-associated transmissions during previous stays (Task 3.1). Patient lines pass close to each other for every potential contact (Task 2.1). The y-axis encodes patient location. Fixed vertical positions for individual wards (as in Baling *et al.* [8]) is not scalable due to the larger number of wards and patients (hundreds), but our layout still aims at preserving the vertical position of wards [4]. Line color conveys infection status, and background color is used to encode location information. Technical details are below.

5.1 Layout

We propose a modified storyline [46] layout algorithm to support our tasks. While typical approaches to storyline drawing optimize the number of edge crossings and minimize bends, we have the constraint of patient locations, including ward (Task 3), and patient contacts (Task 2). Fast layout is required for interactive exploration, as sets of patients in the view can change when filters are applied, and data sets are loaded. Thus, we prioritize runtime over crossing optimization and minimizing bends (see Sect. 8). We build upon existing layouts and combine them and adapt them for our purposes. Given these desired goals, our algorithm is structured as follows (see also Fig. 4):

1. *Data pre-processing*: Construct patient contact graph
2. *Initialization*: Compute initial layout
3. *Constrained force-directed layout*:
   (a) *Constraint – x position*: The temporal order of patient transfers is preserved
   (b) *Constraint – home/hospital*: Patient y position for home, hospital and temporary home location is preserved
   (c) *Force – ward separation*: Different wards are spaced out on the y-axis
   (d) *Force – ward position*: Spring forces encourage keeping wards at the same y position
   (e) *Force – line straightening*: Force encourages straight lines after each iteration
4. *Finalization*:
   (a) *Line order*: For simultaneous contacts of many patients, y order of lines according to the transfer time
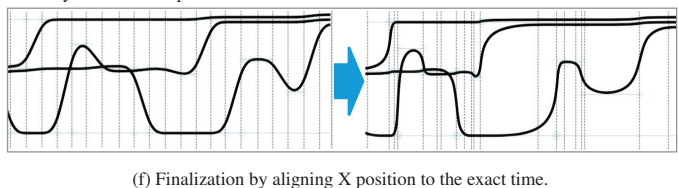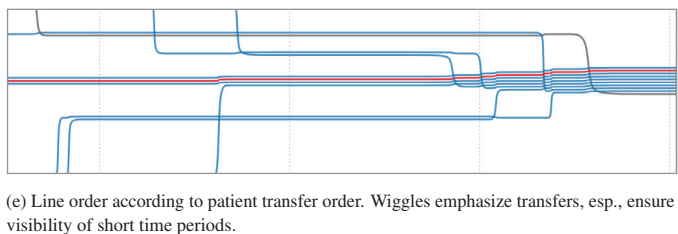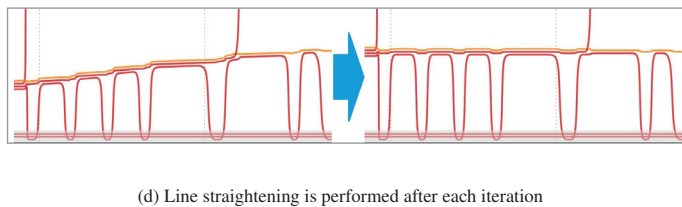   (b) *Temporal adjustment*: Align x position according to the exact time moments

(a) Pre-processing: Constructing patient contact graph

This variant leads to more clutter and lower discriminability between home and 'temporary home' transfers.

The used variant distinguishes transfers from home (lines from top) and from 'temporary home'(lines from the bottom).
(b) Constraining area for home, hospital, and temporary at home.

(c) Spring force for wards improves stability of y position.

(d) Line straightening is performed after each iteration

(e) Line order according to patient transfer order. Wiggles emphasize transfers, esp., ensure visibility of short time periods.

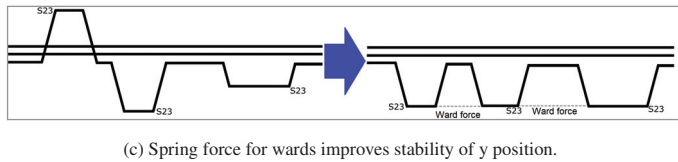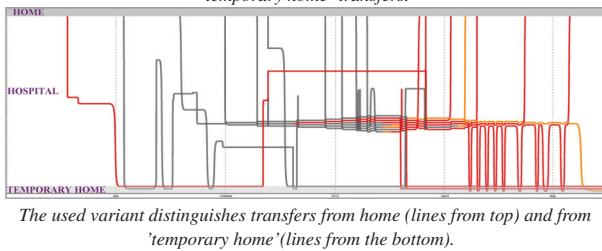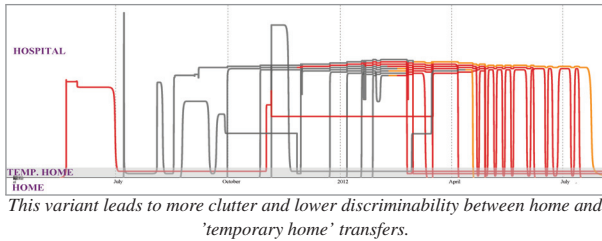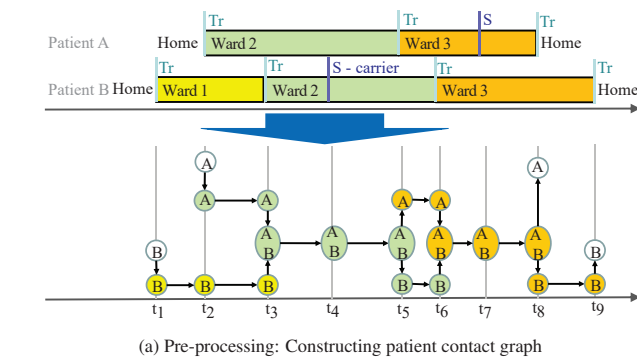(f) Finalization by aligning X position to the exact time.

Figure 4: Stages of the layout algorithm which transform the event-based data into a final storyline layout used by the interactive system.

1. Pre-processing We construct a directed acyclic graph (DAG) from the event-based data. Nodes are sets of patients at a location (i.e., patient contacts) or infection state changes (see Fig. 4a). Edges are transitions between these states. As the transfers have a single timestamp, nodes are interpreted as instantaneous events. Edges that connect these nodes form a from-to transfer edge. This duplication also ensures the duration of the stay at a particular location, and all possible contacts between two transfer time moments. The locations, location changes, and patient contacts can be followed through paths in this graph. For changes in infection state, an infection state change node is created and passing through this node changes the patient's infection state.

2. Initialization As each node of the DAG represents a set of patients together in a location, an initial layout can be computed using dynamic set visualization based on Sankey Diagrams (D3 implementation used [11]). This initial layout takes into account the cardinalities of the sets and operates on the structure of the DAG only. The exact time of patient transfers is taken into account in the finalization.

3. Constrained force-directed layout We employ a force-directed layout for storylines [46, 62]. Our initialization helps avoid local minima, but it requires refinement subject to additional constraints (e.g., to convey patient location). After experimenting with various ways to use space to convey patient location (i.e., all on the bottom), we decided to divide the screen into home – hospital – temporary home from top to bottom (see Fig. 4b). Admission to the hospital is a vertical line descending from the top area of the screen. Temporary home transfers are vertical lines towards the bottom of the screen. When optimizing the layout with the force-directed algorithm, we use a constraint-based approach [19] to enforce these locations and use additional forces to encourage the desired properties of the layout. The first constraint preserves the temporal order of movements along the x-axis. The second constraint restricts movement outside y-axis areas for the three location types (Task 3.1). Additional forces are used to maximize the stability of the y position of a patient during a stay in a ward to help represent contact location (see Fig. 4c) (Task 3.2.). Still, the y position can vary significantly when a patient stays in a location for long periods of time. To remove this position change, we replace the individual y positions of these nodes with the average (see Fig. 4d).

4. Finalization Lines are ordered from top to bottom according to the order patients entered the ward (see Fig. 4e). Even though this may cause more edge crossings, this order helps support Task 2: tracing transmissions. Patients that were in the ward for longer periods of time are more likely to be involved in transmission events. Finally, individual nodes are placed at the precise time of their events along the x-axis, which is important for determining outbreak duration and time of possible pathogen transmissions (Task 4) (see Fig. 4f).

5.2 Visual Design and Interaction

Infection status is encoded using color (see Fig. 5 left) in order to help in tracing transmissions (Task 2). Differentiating between 'unknown' and 'unknown-will be infected' helps infection control experts track patients over long time periods, reducing the requirement to pan and zoom. Details on microbiological data are shown on demand through a tooltip. The contact location (Task 3), specifically the ward, is shown on demand by a colored background hull.

Using a process inspired by [31,68], patient lines are drawn smoothly. As short periods of contact can lead to transmission, wiggles in the line indicate new contacts (see Fig. 4e). We emphasize contacts between patients on the same ward (Task 2) by minimizing the line width for vertical lines (see Fig. 5 right). This also reduces overplotting as inspired by Tanahashi and Ma [67].

Standard pan and zoom operations are supported as well as highlighting of selected patients. When a patient is selected, its line width is increased. Non-focus information is not filtered out [31], as it is a requirement to be able to track patients in the data set. Straightening selected patient lines [37] would be an alternative, but this would interfere with our ward location encoding.

Figure 5: Left: Visual design, where line color encodes infection status. Colored background hulls display wards. Right: Emphasizing patient contacts by reducing the width of vertical lines.

(a) Contact event types for infection transmission

(b) Relevant contacts for pathogen transmission and possible transmission events.

Figure 6: The types of contacts and encoding of transmission events.

5.3 Support for Transmission Pathway Exploration

Potential pathogen transmissions must be visible [44] among many patient contacts over long periods of time (months to years). Panning and zooming over such long time periods is inefficient and may lead to missing important contacts in the data. We developed specialized tracing interactions to support this visual analysis.

Our interactive approach supports 1) *backward tracing* – finding transmission events and patients in the past that could have infected a selected patient. The interaction enables the search for an index patient, i.e., patient zero (Task 2). 2) *forward tracing* – finding patients that could be infected by a selected infected patient at a later point in time, i.e., putative infected patients (Task 5).

We now explain the necessary computation required to interactively backward trace. Computation for forward tracing is done analogously. For a selected patient P, we calculate which contacts might have lead to the patient's infection. A pathogen can be transmitted from an infected patient P_i to the patient P when the patients come into contact. Note, a contact between two diseased patients is deemed irrelevant as both patients are already diseased (see Fig. 6 top). We identify relevant contacts as shown in Fig. 6: the pathogen transmission must have happened during a *critical contact* before the infection was detected – the time moment of the earliest positive microbiological result τ. For each P_i of P and each contact location L, we determine the earliest critical contact before the infection is detected $CC_{i,L} = min_{t_s}\{P_j, t_s, t_e, t_s \leq \tau : L(P) = L(P_i) = L)\}$. We assume the starting time moment of the contact interval (t_s, t_e) where $t_t = t_s$ as the transmission event time, L_t is transmission location (Task 3) and P_t is transmission contact. Note, P can have several potential transmission events – different persons and locations and different times. This analysis is repeated, especially for contacts with *unknown* infection status, as they could be potentially infected if they had contact to an infected patient before. These critical contacts are computed using a constrained path search in the DAG used for the layout. There is currently no bound on how far back or forward in time the transmission events are searched, but user-specified bounds could be implemented, depending on the specifics of the investigated pathogen.

In the view, the selected patient P is emphasized. Relevant contact patients P_t are kept and the non-relevant patients are de-emphasized. The positive test/screening events τ are highlighted with ellipses (orange for carrier and red for diseased) and connected to potential transmission events (t_t, P_t, L_t) with flashback lines inspired by [48] (see Fig. 7 & 8). The circle color encodes the type of relevant contact (Task 2.1). We show all possible transmission events – i.e., several flashback lines. Connections indicate the length of time the potential infection was not detected by the screening or testing (Task 4).

6 CASE STUDY

The visual analytics system was applied to real data from a large German hospital by an infection control expert. The anonymized data includes hospital location data of ∼180,000 patients and ∼900,000 microbiological data over four years. The focus of the case study is

Figure 7: Forward tracing of pathogen transmissions of patient P68475. The patient potentially infected many patients at the ward S279.

a multi-resistant pathogen of special interest: the bacteria *Klebsiella pneumoniae*. *K. pneumoniae* is a commensal gut bacterium that is a dominant cause of hospital-acquired infections. It is responsible for infections in the urinary tract, respiratory tract and blood stream [40].

In the first quarter of 2012 (actual date anonymized), the infection control experts faced an outbreak of *K. pneumoniae* in an area with four wards. Staff members work on all four wards and the patients are frequently swapped between these locations. The total capacity of the area is 50 beds. The outbreak investigation performed at that time was done manually by the infection control experts, meaning the data was collected from several systems and merged into spreadsheets. The original analysis using classical epidemiological data assigned in total 12 patients to the outbreak. However, a deeper analysis of the pathogen transmission was not possible.

This use case shows how the infection control experts analyzed this outbreak retrospectively using the system. The interaction with the system was supported by the visual analytics experts. Since the *patient zero* was not found during the first outbreak investigations, the experts aimed to trace back the initial source of infection and to identify potential cases that were overlooked at that time.

First, the infection control expert consults the epidemic curve for *K. pneumoniae* of the first months of 2012 for the entire hospital to assess the number of infected patients (Task 1). The data was cleaned for copystrains (repeated samples of the same patients), allowing to focus on new positive results (see Fig. 9a). Only a small increase in the number of infections at the hospital has occurred. Thus, this outbreak is difficult to detect automatically. Based on his expertise, his focus turns to the four wards, S276, S278, S279, and S295 (see Fig. 9b), where the outbreak was initially localized. The numbers of new infections in February in the wards were: 1 patient (S276), 3 patients (S278) and 9 patients (S279) (see Fig. 9b). The number of newly infected patients on the wards was higher than the endemic level assuming an *K. pneumoniae* outbreak on the ward.

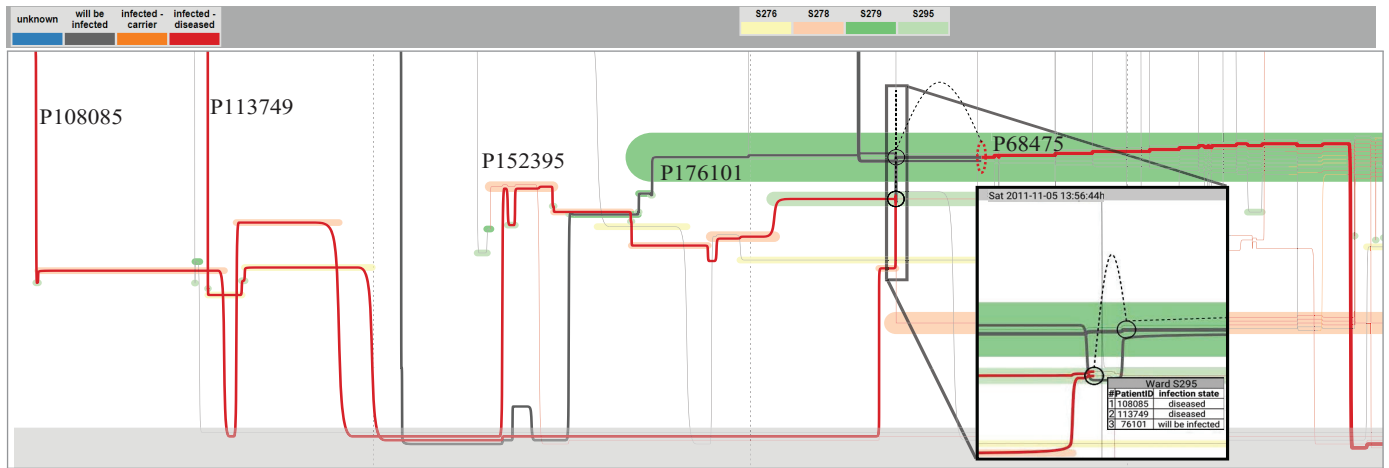The expert begins to reconstruct the transmission pathway (Task 2)

Figure 8: Backward tracing of pathogen transmissions of patient P68475 – indirectly via patient P76101's contact with two patients on ward S295.

and visualizes all positively tested patients on the wards in the *Transmission Pathway View* (see supplementary material and Fig. 7). In total, there were 23 infected patients, which is higher than the original outbreak analysis showed. A dominant cluster of patients highlights 12 infected patients on ward S279 in winter 2011/2012. It reveals the first infected patient, patient P68475, who was already tested positive in November 2011 – three months before this cluster of *K. pneumoniae* (Task 4). By using the *forward tracing interaction* from patient P68475, the infection control expert identified further patients that were in contact with this patient (see Fig. 7). Thus, patient P68475 is the cause for the spreading of *K. pneumoniae* on ward S279, but not for all patients in the outbreak.

Since there was no previous contact of patient P68475 with colonized patients (carrier) or diseased patients (with *K. pneumoniae*) in ward S279, the expert uses the *backward tracing interaction* to search for potential contacts on other wards (see Fig. 8). The visualization showed that patient P68475 had no direct contact to patients with positive laboratory tests for *K. pneumoniae*. However, the tracing lines lead to ward S95. As shown in the zoomed view of Fig. 8, patient P68475 had contact with infected patients indirectly via patient P76101, who, in turn, was in contact with two patients on ward S295 (Task 2.1). This is a novel insight. Both patients, P113749 and P108085, were tested positive for *K. pneumoniae*. Thus, a possible transmission route of *K. pneumoniae* to ward S279 could be via $P113749/108985 - P761010 - P68475$. As Fig. 8 shows, both patients are frequently returning to the hospital (many vertical lines from bottom) and usually stay for longer periods in the hospital (long horizontal lines). Forward tracing shows the potential beginning of the transmission events was before their first admission to the hospital (see the start and the length of the tracing lines in Fig. 10). *Patient Timeline View* confirms this finding (see Fig. 11), where the first positive microbiological results (red vertical lines of patient P113749 and P108985) are before the first hospital admission (Task 3.1). This is due to the limited dataset availability. The real first admission was before the start of the available data. The forward tracing shows that both patients could infect patients at ward S295 and S279 (see Fig. 10).

The *Patient Timeline View* (see Fig. 11) identifies a third infected patient, patient P152039 on ward S276, who tested positive before the increase of infected patients occurs (see red vertical line in 2011). Since all patients had contact to several patients, these three patients are regarded as *patient zero(s)* and the potential origin for the *K. pneumoniae* for this outbreak on the four wards S276, S278, S279 and S295 (Task 2& 3) with genetic analysis required for confirmation.

To provide further support and verification for the hypothesized transmission pathways, the experts also used genome information of the pathogen. Comparison of genomes of bacterial strains adds further evidence to the analysis of outbreaks. Details of this external analysis

(a) Epidemic curve for new infections of *K. pneumoniae*– sum of whole hospital.

(b) Epidemic curve for new infections of *K. pneumoniae*– suspicious wards.
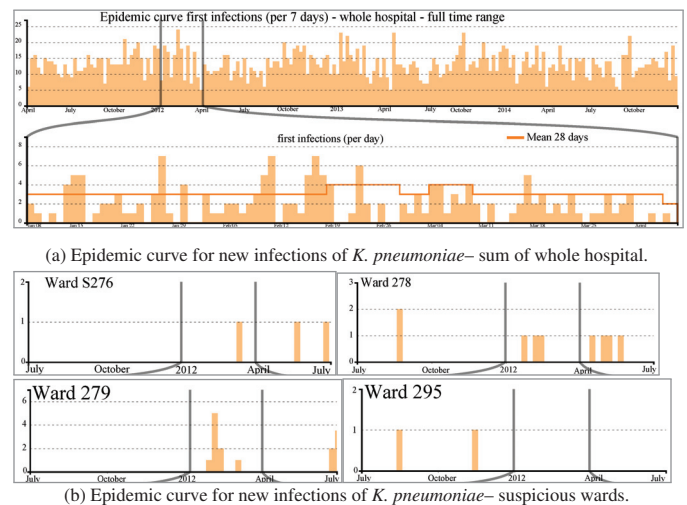
Figure 9: Epidemic curve with focus on the time period of interest: Early-2012 indicates an higher number of new infections at ward S279.

are provided in the annex in the supplementary material. The genome analysis confirms that eleven patients on the wards S279 and S295 were infected by one *K. pneumoniae* bacteria strain, including the potential *patient zero* candidate P113749. Interestingly, the two other potential *patients zero* P108085 and P152039 reveal substantial genome differences to this set of patients (see Fig. A1 in annex). Thus, the expert assigns P113749 as *patient zero* of this outbreak on the wards. The genome analysis reveals P1520395 as a further source of infection for at least one more patient on the wards. Thus, the two transmission pathways occurred simultaneously (Task 2.3). This is a novel insight.

A further question is how many undetected transmissions the *patient zero* P11349 had and whether there are potentially other infected patients (Task 5). Therefore, the infection control expert opens a *Contact Network View* for this patient in the period when the initial transmissions were detected – beginning of November 2011 (see Fig. 12). The tracing interaction in this view highlights several potential transmissions to patients that were (not) detected as infected during their stay on the wards. *Transmission Pathway View* and *Patient Timeline View* (see Fig. 1) show that the non-infected patients were screened (e.g., rectal screening, urine, catheters) regularly and tested negative.

The use case provides support that the visual analytics system enables the infection control experts to conduct a classical epidemiological analysis in a fast, reliable and comprehensive way. Highlighting potential transmission events enables to reveal transmission pathways
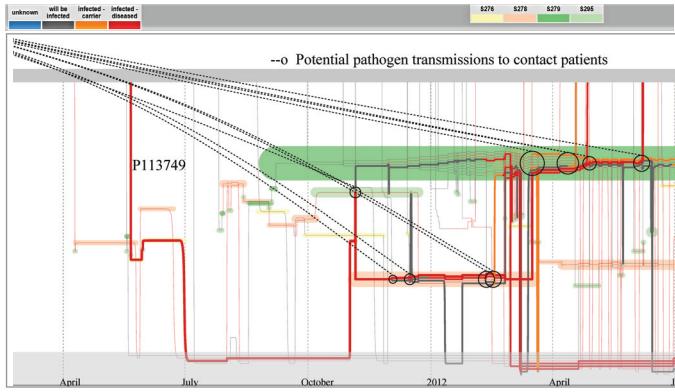
Figure 10: Forward tracing for P113749 shows potential early start of outbreak and transmissions to patients at wards S279, S295 and S278.
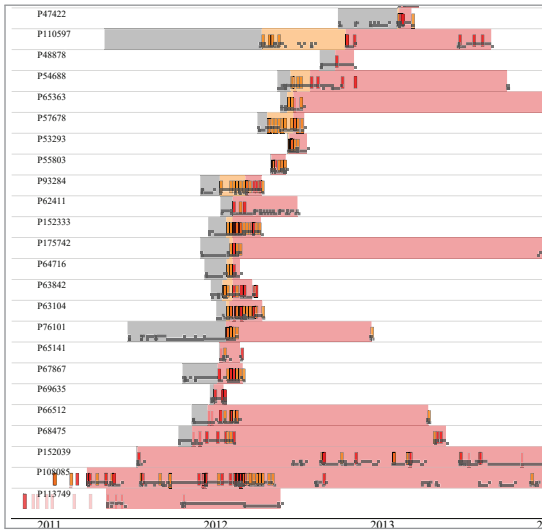

Figure 11: *Patient Timeline View* of the outbreak patients sorted by time of first infection shows first infected patients already in 2011. The red background indicates the infection status starting from the first laboratory evidence of *K. pneumoniae*–shown as red vertical bar.

within the hospital. The steps taken for this analysis with our tool took the hygienist 30 minutes. He estimated that this analysis would normally take about two working days using his previous methods. The hygienists usually determine the epidemic curves and contact networks manually through raw data from multiple information systems.

7 DOMAIN EXPERT FEEDBACK

This qualitative evaluation focused on the *Transmission Pathway View* through an online questionnaire. Twenty-five domain experts from seven institutions in Germany participated in this evaluation. All participants had several years of practical experience with infection control in hospitals: eleven were clinicians, epidemiologists or hygienists, six were medical data analysis experts, six were medical data experts, and two were healthcare managers. Overall, 60% of the participants were familiar with our interactive visualization, i.e., have participated in feedback sessions, or had seen a live demonstration of the tool. At the beginning of the experiment, a one- paragraph explanation of each view was provided with no further training.

The online questionnaire showed the views from the use case and asked participants to read these visualizations and answer the usability and understandability questions as presented in Fig. 13–15. Participants rated their answers to each question on a Likert scale: 1–5, with 5 meaning 'fully agree' and could provide free text comments.

Participants found the view useful ($Mo = 5$) and easy to understand
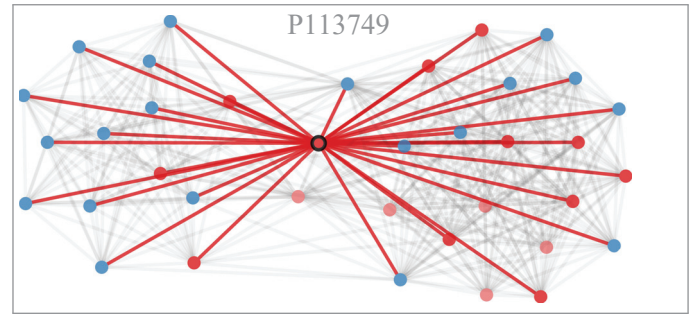
Figure 12: *Contact Network View*: Patient 113749 could have infected more patients in early November 2011.
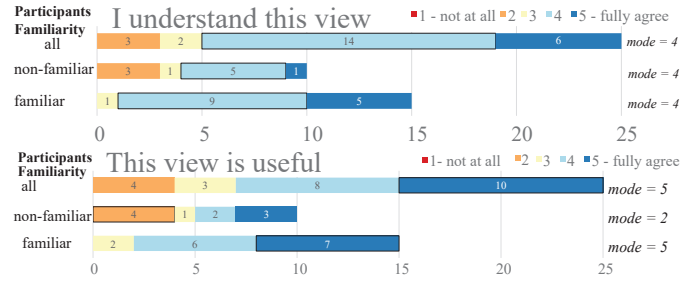
Figure 13: Understandability and usefulness of the view.

($Mo = 4$) (see Fig. 13). As expected, participants familiar with the tool gave higher scores. This result together with free text feedback indicates that training is needed for using the approach effectively.

We also assessed how well the views supported the intended tasks (see Fig. 14). Participants could identify contacts ($Mo = 4$) and understand their infection status ($Mo = 5$) (Task 2.1). They can clearly identify patient stay duration and movement in or out of the hospital ($Mo = 4$) (Task 3.1). Participants found it harder to identify the wards ($Mo = 3$) (Task 3.2). The free text showed mixed reviews of using background color with some participants preferring it while others not. Therefore, we interactively enable/disable this encoding.

Participants found the highlighting of potential transmission events very useful ($Mo = 4$) and understandable ($Mo = 4$) (see Fig. 15). They found the feature helpful in determining which contacts could lead to a transmission and spreading of pathogens ($Mo = 4, 5$). The feedback on support for long time periods when determining outbreak duration (Task 3) was also positive ($Mo = 5$). In free text, experts called for further pan and zoom support, which we revised in our visual interface.

The *comparison of the network diagram and pathway views* showed that nine participants preferred *Transmission Pathway View* and six participants preferred *Contact Network View*. Ten participants preferred a combination of the two views. The distribution of preferences was the same for both familiar and non-familiar participants. This result is expected as the views target different tasks: "I need both views, as they help find an answer to different tasks." *Transmission Pathway View* is useful for detailed spatio-temporal analysis of transmissions and the *Contact Network View* is useful for an overview of contacts without temporal focus (e.g., finding a "super-spreader").
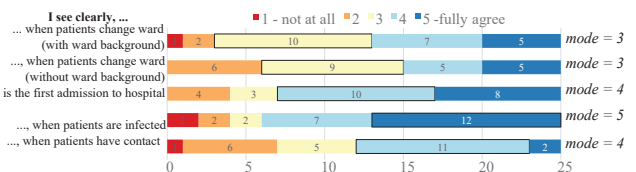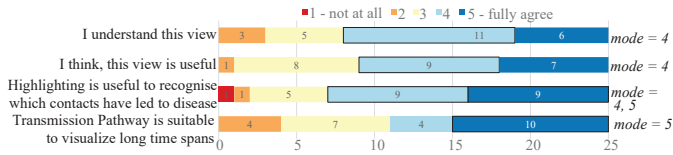

Figure 14: Results for readability of the views.

Figure 15: Results for questions on tracing interaction.

Layout	Edge crossing	Runtime (ms)	Pre-processing (ms)
Our	25^a (37^b)	197	18
[37]	14	160	N/A

Figure 16: Layout quality using Matrix dataset [65]. (^{a}) comparable to Storyflow [37], (^{b}) - including start/end lines.

The *free text feedback* was positive: "Interesting and well thought-out visualization. I was very enthusiastic about the presentation of this visualization. The visualization of such complex data is not-trivial." Participants found some training was required, but after understanding it, they found it very useful: "I find the *Transmission Pathway View* really good. Of course, one has to get used to it first and learn to use it first." One expert suggested broader applicability "I would strongly recommend this visualization to be shown to public health authority (Gesundheitsamt). It could be 'fit for purpose'".

The participants found that scalability in terms of the number of patients posed a significant challenge. They recommended a two-staged process with filtering through the contact network and then show the filtered data in the *Transmission Pathway View*. Four of the twenty-five participants wanted to change the colors (e.g., showing each patient in a different color, using more prominent colors for the three areas (home, hospital, temporary home)). One participant requested further differentiation between carrier and diseased, as it is important for certain pathogens. A number of future extensions were suggested that mainly require additional data. We discuss them in Sect. 8.

8 DISCUSSION AND FUTURE WORK

The *Transmission Pathway View* layout was specifically designed for our tasks rather than to optimize edge crossing and wiggles. Still, the approach performs well on the benchmark Matrix dataset, with 14 entities and 90 transfers [65] (see Fig. 16). We used a laptop PC with an Intel i7-9750H CPU (2.6 GHz) and 32GB memory. Pre-processing lasts 18 ms, the initialization stabilized after 30 iterations (62 ms) and force-directed layout after 10 iterations (135 ms). The layout of use case with 27 entities and 704 transfers and the same number of iterations lasted 2517ms in total (190ms + 1265ms + 1062ms). The computation of tracking interaction for all critical patients lasted 37ms in total.

Our data sets pose visual scalability challenges in terms of the time spans of the data sets, number of patients, number of locations, number of patient transfers, and the number of screenings and tests. Interaction is leveraged to allow the approach to scale to larger data sets, combining several views and automatically highlighting transmission pathway events. They allow to explore pathways of up to hundred patients (see Annex). Explicitly visualizing the number of forward and backward connections of a patient would help the system scale when tracing contacts. Other methods for increasing the scalability of our approach should be investigated.

The visual design of *Transmission Pathway View* relies heavily on colors to encode infection status and wards, limiting the number of wards that can be represented. Automatic detection of suspicious wards (i.e., a sudden increase in the number of cases) would help with this scalability along with forms of automated support. In order to scale a larger number of wards, interactive highlighting could help and hulls could improve space efficiency [67]. The expert feedback indicated a need for enabling annotations to the visualizations with genome data or measures taken during the outbreak, which could be implemented.

The data collection methodology and privacy considerations pose limitations on what can be analyzed. A core challenge is localizing contacts accurately, which is currently handled via electronic health records. More accurate data on patient mobility through the hospital, locations of rooms and beds, as well as information where procedures occur (e.g., surgery, endoscopy, radiography) could be helpful for certain visual analytics tasks. Technologies, such as RFID, could be used, but would incur significant privacy considerations. Also, current records do not confirm when a patient has been recovered. Therefore,
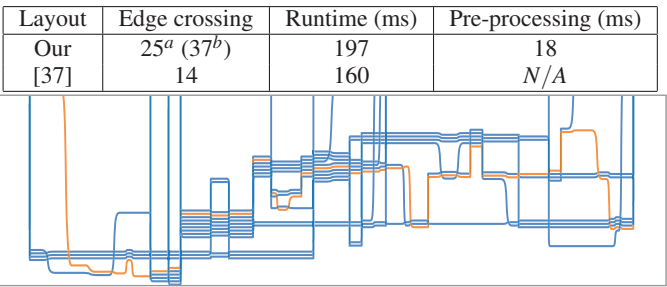
patients remain infected for the remainder of the data set, which is often not the case. These data need to be extracted from the current raw data by new algorithms. These algorithms have to be adjusted to the ward, the patient-type, the type of pathogen and the tests. Additional features, like the comparison and matching of antimicrobial profiles simultaneous with the integration of genomics results of the pathogens would help the infection control expert automatically cluster patients according to strains to better identify transmission pathways. We used an external tool to perform genome analysis to verify the suggested transmission routes. The genome analysis visualization is not part of our system. However, since laboratories have access to near real-time genome analysis, this data can be extracted and used for matching specific pathogens to help verify the transmission pathway hypotheses. Moreover, the visual interface could be extended to support provenance of analysis and interaction. In the future, it would be interesting to encode the mode of transmissions (e.g., airborne or contact) or other potential risk factors (e.g., diabetes, immunotherapy).

Our approach was designed and implemented for transmission pathway reconstruction in hospitals. A modified version of the system has now been deployed in several hospitals for analyzing COVID-19 hospital-associated transmissions (see video in German at https://youtu.be/HAsb3dnUKyI). More generally, our approach could be used to visually analyze contact tracing graphs of localized outbreak clusters and in closed environments, such as cruise ships and buildings. The approach could be applicable to different definitions of spread, such as financial or information contagion. The tracing interaction could help with this reconstruction, but it needs to be adapted to the specifics of the application. For example for information spread, the model would take the beliefs about the information, i.e., whether a person supports or opposes the information [9] into account.

9 CONCLUSION

In this paper, we have presented a visual analytics system for visualizing transmission events and outbreaks in hospitals. The approach focuses on methods for reconstructing the transmission pathways for a certain pathogen back to a *patient zero*. This visual analytics approach was developed through iterative, user-centered design with seven German Hospitals and Infection Control Institutes or Units. As part of the design process, storyline visualizations were adapted to visualize temporal contact networks. Our approach was applied to a real outbreak of *Klebsiella pneumoniae* in a large German hospital. Infection control experts were able to effectively unravel the outbreak and reveal two distinct transmission pathways. Tracing back the initial source, *patient zero* was conducted in a faster and in a more comprehensive way when compared to existing methods. In our final qualitative user study of twenty-five users, we found significant value in the approach when making sense of outbreaks in hospitals.

REFERENCES

[1] S. E. Abdelhamid, C. J. Kuhlman, M. V. Marathe, and S. S. Ravi. Interactive exploration and understanding of contagion dynamics in networked populations. In *Int. Conf. Behavioral, Economic and Socio-cultural Computing*, pp. 1–6, 2016.

[2] B. Alsallakh, L. Micallef, W. Aigner, H. Hauser, S. Miksch, and P. Rodgers. The state-of-the-art of set visualization. *Computer Graphics Forum*, 35(1):234–260, 2016.

[3] D. Arendt and M. Pirrung. Too much time not enough space: Multiscale event sequence visual analytics with m-SVEN. In *The Event Event: IEEE VIS 2016 Workshop on Temporal & Sequential Event Analysis*, 2016.

[4] D. Arendt and M. Pirrung. The "y" of it matters, even for storyline visualization. In *IEEE Conference on Visual Analytics Science and Technology*, pp. 81–91. IEEE, 2017.

[5] B. Bach, N. Henry-Riche, T. Dwyer, T. Madhyastha, J.-D. Fekete, and T. Grabowski. Small multipiles: Piling time to explore temporal patterns in dynamic networks. *Computer Graphics Forum*, 34(3):31–40, 2015.

[6] B. Bach, E. Pietriga, and J.-D. Fekete. GraphDiaries: Animated transitions and temporal navigation for dynamic networks. *IEEE Transactions on Visualization and Computer Graphics*, 20(5):740–754, 2013.

[7] B. Bach, C. Shi, N. Heulot, T. Madhyastha, T. Grabowski, and P. Dragicevic. Time Curves: Folding time to visualize patterns of temporal evolution in data. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):559–568, 2016.

[8] J. T. Balint, D. Arendt, and L. M. Blaha. Storyline visualizations of eye tracking of movie viewing. In *IEEE Workshop on Eye Tracking and Visualization*, pp. 35–39. IEEE, 2016.

[9] Y. Bao, C. Yi, Y. Xue, and Y. Dong. A new rumor propagation model and control strategy on social networks. In *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 1472–1473, 2013.

[10] F. Beck, M. Burch, S. Diehl, and D. Weiskopf. A taxonomy and survey of dynamic graph visualization. *Computer Graphics Forum*, 36(1):133–159, 2017.

[11] M. Bostock, V. Ogievetsky, and J. Heer. D^3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, 2011.

[12] C. Bryan, X. Wu, S. Mniszewski, and K.-L. Ma. Integrating predictive analytics into a spatiotemporal epidemic simulation. In *IEEE Conference on Visual Analytics Science and Technology*, pp. 17–24. IEEE, 2015.

[13] H. S. G. Caballero, A. Corvo, P. M. Dixit, and M. A. Westenberg. Visual analytics for evaluating clinical pathways. In *IEEE Workshop on Visual Analytics in Healthcare*, pp. 39–46. IEEE, 2017.

[14] H. Chen, U. Soni, Y. Lu, R. Maciejewski, and S. Kobourov. Same stats, different graphs. In *Graph Drawing and Network Visualization*, pp. 463–477. Springer, 2018.

[15] M. Chen, L. Floridi, and R. Borgo. What is visualization really for? In L. Floridi and P. Illari, eds., *The Philosophy of Information Quality*, pp. 75–93. Springer, 2014. doi: 10.1007/978-3-319-07121-3_5

[16] E. Choi, M. T. Bahadori, J. Sun, J. Kulas, A. Schuetz, and W. Stewart. Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. In *Conference on Neural Information Processing Systems*, pp. 3504–3512, 2016.

[17] W. Cui, S. Liu, L. Tan, C. Shi, Y. Song, Z. Gao, H. Qu, and X. Tong. Textflow: Towards better understanding of evolving topics in text. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2412–2421, 2011.

[18] T. Donker, J. Wallinga, R. Slack, and H. Grundmann. Hospital networks and the dispersal of hospital-acquired pathogens by patient transfer. *PloS one*, 7(4), 2012.

[19] T. Dwyer, Y. Koren, and K. Marriott. Ipsep-cola: An incremental procedure for separation constraint layout of graphs. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):821–828, 2006.

[20] S. Engler-Hüsch, T. Heister, N. T. Mutters, J. Wolff, and K. Kaier. In-hospital costs of community-acquired colonization with multidrug-resistant organisms at a german teaching hospital. *BMC health services research*, 18(1):1–7, 2018.

[21] D. G. Enki, P. H. Garthwaite, C. P. Farrington, A. Noufaily, N. J. Andrews, and A. Charlett. Comparison of statistical algorithms for the detection of infectious disease outbreaks in large multiple surveillance systems. *PloS one*, 11(8):e0160759, 2016.

[22] M. Freire, C. Plaisant, B. Shneiderman, and J. Golbeck. Manynets: an interface for multiple network analysis and visualization. In *SIGCHI Conf. on Human Factors in Computing Systems*, pp. 213–222. ACM, 2010.

[23] T. Fröschl and M. Nöllenburg. Minimizing wiggles in storyline visualizations. In *Graph Drawing and Network Visualization*, pp. 585–587. Springer, 2017.

[24] C. Geffers and P. Gastmeier. Nosocomial Infections and Multidrug-resistant Organisms in Germany. *Detsches Ärzteblatt International*, 108(6):87–93, 2011. doi: 10.3238/arztebl.2011.0087

[25] M. Gronemann, M. Jünger, F. Liers, and F. Mambelli. Crossing minimization in storyline visualization. In *International Symposium on Graph Drawing and Network Visualization*, pp. 367–381. Springer, 2016.

[26] S. Guo, K. Xu, R. Zhao, D. Gotz, H. Zha, and N. Cao. Eventthread: Visual summarization and stage analysis of event sequence data. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):56–65, 2018.

[27] B. Haarbrandt, B. Schreiweis, S. Rey, U. Sax, S. Scheithauer, O. Rienhoff, P. Knaup-Gregori, U. Bavendiek, C. Dieterich, B. Brors, et al. Highmed–an open platform approach to enhance care and research across institutional boundaries. *Methods of information in medicine*, 57(S 01):e66–e81, 2018.

[28] A. Hakone, L. Harrison, A. Ottley, N. Winters, C. Gutheil, P. K. Han, and R. Chang. Proact: Iterative design of a patient-centered visualization for effective prostate cancer health risk communication. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):601–610, 2016.

[29] A. Kennedy, K. Klein, A. Nguyen, and F. Y. Wang. The graph landscape: using visual analytics for graph set analysis. *J. Visualization*, 20(3):417–432, 2017.

[30] N. Kerracher and J. Kennedy. Constructing and evaluating visualisation task classifications: Process and considerations. *Computer Graphics Forum*, 36(3):47–59, 2017.

[31] N. W. Kim, S. K. Card, and J. Heer. Tracing genealogical data with timenets. In *International Conference on Advanced Visual Interfaces*, pp. 241–248, 2010.

[32] I. Kostitsyna, M. Nöllenburg, V. Polishchuk, A. Schulz, and D. Strash. On minimizing crossings in storyline visualizations. In *International Symposium on Graph Drawing*, pp. 192–198. Springer, 2015.

[33] B. C. Kwon, V. Anand, K. A. Severson, S. Ghosh, Z. Sun, B. I. Frohnert, M. Lundgren, and K. Ng. DPVis: Visual analytics with hidden markov models for disease progression pathways. *IEEE Transactions on Visualization and Computer Graphics*, 2020.

[34] B. C. Kwon, M.-J. Choi, J. T. Kim, E. Choi, Y. B. Kim, S. Kwon, J. Sun, and J. Choo. Retainvis: Visual analytics with interpretable and interactive recurrent neural networks on electronic medical records. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):299–309, 2018.

[35] A. Lee, D. Archambault, and M. Nacenta. Dynamic network plaid: A tool for the analysis of dynamic networks. In *CHI Conference on Human Factors in Computing Systems*, p. paper 130, 2019.

[36] Y. Liang, Z. Shi, S. I. Sritharan, and H. Wan. Simulation of the spread of epidemic disease using persistent surveillance data. *Simulation*, p. 1, 2010.

[37] S. Liu, Y. Wu, E. Wei, M. Liu, and Y. Liu. Storyflow: Tracking the evolution of stories. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2436–2445, 2013.

[38] Z. Liu, S. H. Zhan, and T. Munzner. Aggregated dendrograms for visual comparison between many phylogenetic trees. *IEEE Transactions on Visualization and Computer Graphics*, 2019. doi: 10.1109/TVCG.2019.2898186

[39] R. Maciejewski, P. Livengood, S. Rudolph, T. F. Collins, D. S. Ebert, R. T. Brigantic, C. D. Corley, G. A. Muller, and S. W. Sanders. A pandemic influenza modeling and visualization tool. *Journal of Visual Languages & Computing*, 22(4):268–278, 2011.

[40] S. Magill, J. Edwards, W. Bamberg, Z. Beldavs, G. Dumyati, M. Kainer, R. Lynfield, M. Maloney, L. McAllister-Hollod, J. Nadle, S. Ray, D. Thompson, L. Wilson, S. Fridkin, and the Emerging Infections Program Healthcare-Associated Infections & Antimicrobial Use Prevalance Survey Team. Multistate point-prevalence survey of health care-associated infections. *N Eng J Med*, 370(13):10, 2014.

[41] M. Monroe, R. Lan, H. Lee, C. Plaisant, and B. Shneiderman. Temporal event sequence simplification. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2227–2236, 2013.

[42] M. Müller, M. Petzold, M. Wunderlich, T. Baumgartl, M. Höhn, V. Eichel, N. Mutters, S. Scheithauer, M. Marschollek, and T. von Landesberger. Visual analysis for hospital infection control using a rnn model. In *EuroVA*, pp. 073–077. EuroGraphics, 2020.

[43] T. Munzner. A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):921–928, 2009.

[44] T. Munzner, F. Guimbretière, S. Tasiran, L. Zhang, and Y. Zhou. TreeJuxtaposer: Scalable tree comparison using focus+context with guaranteed visibility. In *ACM SIGGRAPH 2003*, p. 453–462, 2003. doi: 10.1145/1201775.882291

[45] P. H. Nguyen, K. Xu, R. Walker, and B. W. Wong. Timesets: Timeline visualization with set relations. *Information Visualization*, 15(3):253–269, 2016.

[46] M. Ogawa and K.-L. Ma. Software evolution storylines. In *International Symposium on Software Visualization*, p. 35–42. ACM, 2010. doi: 10.1145/1879211.1879219

[47] K. Padia, K. H. Bandara, and C. G. Healey. Yarn: generating storyline visualizations using HTN planning. In *Graphics Interface*, pp. 17–24. ACM, 2018.

[48] K. Padia, K. H. Bandara, and C. G. Healey. A system for generating storyline visualizations using hierarchical task network planning. *Computers & Graphics*, 78:64–75, 2019.

[49] B. Preim and K. Lawonn. A survey of visual analytics for public health. *Computer Graphics Forum*, 39(1):543–580, 2020.

[50] J. Qi, C. Liu, B. C. Cappers, and H. van de Wetering. Visual analysis of parallel interval events. In *EuroVis (Short Papers)*, pp. 31–35, 2018.

[51] L. Qiang and C. Bingjie. Storycake: A hierarchical plot visualization method for storytelling in polar coordinates. In *International Conference on Cyberworlds*, pp. 211–218. IEEE, 2016.

[52] K. Reda, C. Tantipathananandh, A. Johnson, J. Leigh, and T. Berger-Wolf. Visualizing the evolution of community structures in dynamic social networks. *Computer Graphics Forum*, 30(3):1061–1070, 2011.

[53] A. Rind, P. Federico, T. Gschwandtner, W. Aigner, J. Doppler, and M. Wagner. Visual analytics of electronic health records with a focus on time. In *New Perspectives in Medical Records*, pp. 65–77. Springer, 2017.

[54] A. Rind, T. D. Wang, W. Aigner, S. Miksch, K. Wongsuphasawat, C. Plaisant, and B. Shneiderman. Interactive information visualization to explore and query electronic health records. *Foundations and Trends in HCI*, 5(3):207–298, 2013.

[55] M. Rosvall and C. T. Bergstrom. Mapping change in large networks. *PLoS One*, 5(1):e8694, 2010.

[56] F. D. Sahneh, A. Vajdi, H. Shakeri, F. Fan, and C. Scoglio. Gemfsim: a stochastic simulator for the generalized epidemic modeling framework. *Journal of Computational Science*, 22(1):36–44, 2017.

[57] C. Schmidt, M. Röhlig, B. Grundel, P. Daumke, M. Ritter, A. Stahl, P. Rosenthal, and H. Schumann. Combining visual cleansing and exploration for clinical data. In *IEEE Workshop on Visual Analytics in Healthcare*, pp. 25–32, 2019.

[58] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2431–2440, 2012.

[59] D. Shahaf, C. Guestrin, and E. Horvitz. Trains of thought: Generating information maps. In *International Conference on World Wide Web*, p. 899–908, 2012. doi: 10.1145/2187836.2187957

[60] A. Shrestha, B. Miller, Y. Zhu, and Y. Zhao. Storygraph: Extracting patterns from spatio-temporal data. In *SIGKDD Workshop on Interactive Data Exploration and Analytics*, pp. 95–103. ACM, 2013.

[61] S. Silvia. Variantflow: Interactive storyline visualization using force directed layout, 2016.

[62] S. Silvia, J. Abbas, S. Huskey, and C. Weaver. Storyline visualization with force directed layout. In *IEEE VIS Posters*, 2015.

[63] P. Simonetto, D. Archambault, and S. Kobourov. Drawing dynamic graphs without timeslices. In *Graph Drawing and Network Visualization*, pp. 394–409. Springer, 2018.

[64] P. Simonetto, D. Archambault, and S. G. Kobourov. Event-based dynamic graph visualisation. *IEEE Transactions on Visualization and Computer Graphics*, 26(7):2373–2386, 2018.

[65] Y. Tanahashi. Movie interaction dataset. `http://vis.cs.ucdavis.edu/~tanahashi/data_downloads/storyline_visualizations/story_data.tar`, 2013.

[66] Y. Tanahashi, C.-H. Hsueh, and K.-L. Ma. An efficient framework for generating storyline visualizations from streaming data. *IEEE Transactions on Visualization and Computer Graphics*, 21(6):730–742, 2015.

[67] Y. Tanahashi and K.-L. Ma. Design considerations for optimizing storyline visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2679–2688, 2012.

[68] T. Tang, S. Rubab, J. Lai, W. Cui, L. Yu, and Y. Wu. istoryline: Effective convergence to hand-drawn storylines. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):769–778, 2018.

[69] S. van den Elzen, D. Holten, J. Blaas, and J. J. van Wijk. Reducing snapshots to points: A visual analytics approach to dynamic network exploration. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):1–10, 2015.

[70] T. C. van Dijk, M. Fink, N. Fischer, F. Lipp, P. Markfelder, A. Ravsky, S. Suri, and A. Wolff. Block crossings in storyline visualizations. In *International Symposium on Graph Drawing and Network Visualization*, pp. 382–398. Springer, 2016.

[71] A. Vogogias, J. Kennedy, D. Archambault, B. Bach, V. A. Smith, and H. Currant. BayesPiles: Visualisation support for bayesian network structure learning. *ACM Transactions on Intelligent Systems Technology*, 10(1), 2018.

[72] T. von Landesberger, S. Bremm, N. Andrienko, G. Andrienko, and M. Tekušová. Visual analytics methods for categoric spatio-temporal data. In *IEEE Conference on Visual Analytics Science and Technology*, pp. 183–192. IEEE, 2012.

[73] T. von Landesberger, S. Diel, S. Bremm, and D. W. Fellner. Visual analysis of contagion in networks. *Information Visualization*, 14(2):93–110, 2015.

[74] T. von Landesberger, D. W. Fellner, and R. A. Ruddle. Visualization system requirements for data processing pipeline design and optimization. *IEEE Transactions on Visualization and Computer Graphics*, 23(8):2028–2041, 2016.

[75] T. von Landesberger, M. Gorner, and T. Schreck. Visual analysis of graphs with multiple connected components. In *Visual Analytics Science and Technology*, pp. 155–162. IEEE, 2009.

[76] T. von Landesberger, M. Wunderlich, T. Baumgartl, M. Höhn, M. Marschollek, and S. Scheithauer. Visual-Interactive Exploration of Pathogen Outbreaks in Hospitals. In *EuroVis 2019 - Posters*. The Eurographics Association, 2019. doi: 10.2312/eurp.20191133

[77] J. Walker, R. Borgo, and M. W. Jones. Timenotes: a study on effective chart visualization and interaction techniques for time-series data. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):549–558, 2015.

[78] Y. Wang, D. Archambault, H. Haleem, T. Moeller, Y. Wu, and H. Qu. Nonuniform timeslicing of dynamic graphs based on visual complexity. In *IEEE Visualization Conference (VIS) Short Papers*, pp. 241–245. IEEE, 2019. doi: 10.1109/VISUAL.2019.8933748

[79] A. Wolff. Computing storyline visualizations with few block crossings. In *Graph Drawing*, vol. 10692, p. 365. Springer, 2018.

[80] M. Wunderlich, I. Block, T. von Landesberger, M. Petzold, M. Marschollek, and S. Scheithauer. Visual analysis of probabilistic infection contagion in hospitals. In *Vision, Modeling, and Visualization*, pp. 143–150. Eurographics Association, 2019.

[81] S. Yagi, T. Itoh, and M. Takatsuka. A layout technique for storyline-based visualization of consecutive numerical time-varying data. In *International Symposium on Visual Information Communication and Interaction*, pp. 156–157. ACM, 2015.

[82] A. Yanez, J. Duggan, C. Hayes, M. Jilani, and M. Connolly. Pandemcap: Decision support tool for epidemic management. In *IEEE Workshop on Visual Analytics in Healthcare*, pp. 24–30. IEEE, 2017.

[83] B. Zacher and I. Czogiel. Supervised learning improves disease outbreak detection. *arXiv preprint arXiv:1902.10061*, 2019.