

THOVEA: Ein dreischichtiges visuelles Analysesystem für Informationsverbreitung über große Netzwerke

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Medizinische Informatik

eingereicht von

Mert Usul, BSc Matrikelnummer 01526401

an der Fakultät für Informatik der Technischen Universität Wien Betreuung: Univ.Prof. Mag.rer.soc.oec. Dr.rer.soc.oec. Silvia Miksch

Wien, 26. April 2025

Mert Usul

Silvia Miksch





THOVEA: A Three-Layer Visual Analytics System for Information Diffusion over Large Networks

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Medical Informatics

by

Mert Usul, BSc Registration Number 01526401

to the Faculty of Informatics

at the TU Wien

Advisor: Univ.Prof. Mag.rer.soc.oec. Dr.rer.soc.oec. Silvia Miksch

Vienna, April 26, 2025

Mert Usul

Silvia Miksch



Erklärung zur Verfassung der Arbeit

Mert Usul, BSc

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Ich erkläre weiters, dass ich mich generativer KI-Tools lediglich als Hilfsmittel bedient habe und in der vorliegenden Arbeit mein gestalterischer Einfluss überwiegt. Im Anhang "Übersicht verwendeter Hilfsmittel" habe ich alle generativen KI-Tools gelistet, die verwendet wurden, und angegeben, wo und wie sie verwendet wurden. Für Textpassagen, die ohne substantielle Änderungen übernommen wurden, haben ich jeweils die von mir formulierten Eingaben (Prompts) und die verwendete IT- Anwendung mit ihrem Produktnamen und Versionsnummer/Datum angegeben.

Wien, 26. April 2025

Mert Usul



Danksagung

Mit dieser Seite möchte ich mich bei meiner Betreuerin bedanken, die mich während des gesamten Prozesses hervorragend begleitet und mir stets zeitnahes und wertvolles Feedback gegeben hat.

Außerdem möchte ich meiner Familie danken, insbesondere meiner Mutter. Ohne ihre Unterstützung hätte ich dieses Studium nicht abschließen können.



Acknowledgements

With this page, I want to thank my supervisor who consistently provided valuable and timely feedback, as well as, excellent guidance throughout the whole process.

Furthermore i want to thank my family, especially my mother. I could not have completed this degree without their support.



Kurzfassung

Wir stellen *Thovea* vor, ein dreischichtiges visuelles Analysesystem für Informationsverbreitung über große Netzwerke. Mit diesem dreischichtigem Ansatz können mehrere Diffusionsprozesse gleichzeitig untersucht werden, wobei jede Schicht eine andere Abstraktionsebene darstellt. Unsere Methode unterstützt eine Top-Down-Strategie: ausgehend von einer Übersicht auf hoher Abstraktionsebene werden die einzelnen Knoten bis ins Detail untersucht. Dieses Design ermöglicht ein umfassendes Verständnis des jeweiligen Diffusionsprozesses, während es gleichzeitig eine detaillierte Untersuchung bis hin zu einzelnen Infektionsketten ermöglicht. Wir integrieren geeignete Layoutalgorithmen und Visualisierungsmethoden in jede Ebene, um eine skalierbare und flexible Erforschung von Informationsverbreitungsprozessen und Netzwerken zu ermöglichen.

Informationsverbreitung beschreibt, wie sich Information über ein engmaschiges Netzwerk verbreitet und ist für viele Bereiche relevant, z. B. für die Modellierung der Verbreitung von Fake News, Krankheitserregern oder Malware. In der Literatur gibt es zwar verschiedene Modellierungsansätze, die Visualisierung von Informationsverbreitungsprozessen ist jedoch immer noch ein wenig erforschtes Gebiet. Um einen umfassenden Überblick zu gewinnen, präsentieren wir den aktuellen Stand der Technik zu visuellen Analysesystemen, die zur Darstellung und zum Verständnis von Informationsverbreitungssysteme in Netzwerken eingesetzt werden. Wir präsentieren eine Taxonomie, die ausgewählte Beiträge kategorisiert, strukturiert und über Anwendungsbereiche hinweg verallgemeinert. Bestehende Beiträge sind oft auf Nischenprobleme gegebener Anwendungsgebiete zugeschnitten und nicht verallgemeinerbar. Die visuelle Skalierbarkeit, konkret das arbeiten mit Netzwerken, welche Knoten und Kanten im Tausenderbereich besitzten, stellt ebenfalls eine Herausforderung dar. Dies schränkt ihre praktische Anwendbarkeit ein.

Wir evaluieren unser System, mit zwei Fallstudien und der quantitativen ICE-T [WAM⁺19]-Methode. Letzteres bestätigt den Wert von *Thovea* mit einem globalen durchschnitt von 5,82, wobei jede Komponente mit über 5 Punkten bewertet wurde. Die am höchsten bewerteten Komponenten stellen *Essence* und *Insight* dar, mit jeweils 6,2 Punkten.



Abstract

We introduce *Thovea*, a **TH**ree-layer information diffusi**O**n **V**isual analytics syst**E**m for l**A**rge networks. This three-layered approach is designed to simultaneously investigate several diffusion processes—with each layer representing a different level of abstraction. Our method supports a top-down exploration strategy: starting from a high-level overview and drilling down to individual node details. This design provides a comprehensive understanding of the diffusion process(es) at hand, while enabling a detailed exploration down to distinct infection chains. We integrate suitable network layouts and representation methodologies into each level, aiming to support a scalable and agile exploration of information diffusion processes and networks.

Information Diffusion investigates how information spreads over a tightly connected network and is relevant to many domains, such as modeling fake news spreading, pathogen contagion, or malware infections. While different modeling approaches exist in the literature, the visualization of information diffusion processes still constitutes an underinvestigated problem. To gain a comprehensive overview, we present a survey and analysis of the current state-of-the-art in visual analytics techniques employed in representing and understanding diffusion processes happening over networks. We introduce a taxonomy that categorizes and structures the selected approaches while generalizing across application domains. Existing contributions are often tailored to niche domain-specific problems and lack generalizability. Visual scalability is also a challenge, as current research still struggles to effectively handle networks with thousands of nodes and edges, limiting their practical applicability.

We evaluate our system by (i) presenting two case studies and (ii) conducting a quantitative value-driven estimation using the ICE-T [WAM⁺19] methodology. The latter confirms the value of *Thovea* with a global average score of 5.82, where each component has been awarded a score greater than 5. The highest scored components are *Essence* and *Insight* with a score of 6.2 each.



Contents

Kurzfassung											
\mathbf{A}	bstract	xiii									
Co	ontents	xv									
1	Introduction	1									
	1.1 Research Questions	2									
	1.2 Contribution	$\frac{2}{3}$									
2	Related Work	5									
	2.1 Theory of Information Diffusion	5									
	2.2 Large Graph Visualization	6									
	2.3 State-of-the-Art: Visualization of Information Diffusion	6									
3	Thovea - The System	23									
	3.1 Data	23									
	3.2 Users & Tasks Analysis	24									
	3.3 Visualization Design	25									
	3.4 System Architecture	36									
4	Evaluation										
	4.1 Datasets	39									
	4.2 Case Studies	40									
	4.3 ICE-T Expert Interview	48									
5	Discussion	53									
	5.1 Revisiting Research Questions	55									
6	Summary	59									
Overview of Generative AI Tools Used											
List of Figures											
		XV									

List of Tables	67
Acronyms	69
Bibliography	71
Appendix	81
ICE-T Tasks	. 81
ICE-T Questionnaires	. 82
ICE-T Results	. 82

CHAPTER

Introduction

Information Diffusion (ID) is a research domain that investigates how information propagates across a network or a population. This phenomenon concerns various disciplines, including the analysis of epidemic transmission pathways [BPW⁺21], or the investigation of information flow through social networks [GHFZ13]. In the context of social networks, there is significant interest in understanding the dynamics of ID. A closely related combinatorial problem, known as *Influence Maximization (IM)*, focuses on identifying a small set of initial seed nodes that maximize the spread of information, i.e., the number of nodes reached, under a given ID model. This has plenty of potential applications, such as optimizing the influence or propagation [LFWT18a] of a marketing campaign [DR01]. Even under simple models, identifying an optimal seed set has been proven to be an NP-hard problem [KKT03]. IM primarily relies on stochastic ID models, such as the Independent Cascade (IC) or Linear Threshold (LT), further complicating the evaluation of individual seed sets [LFWT18a]. These challenges make scalability a significant hurdle in solving the IM problem efficiently [CWW10].

Given these computational challenges, Visual Analytics (VA) can assist domain experts in understanding and analyzing the underlying ID process, while also facilitating the comparison of different ID models. However, visualization and VA solutions for ID remain underexplored. Existing contributions are often tailored to niche domain-specific problems and lack adaptability and generalizability for problems in other domains or for broader applications. Many approaches in social media analysis often rely on platformspecific data, such as Twitter [NYX⁺12], Weibo [CCW⁺19], and Google [VWH⁺13], where diffusion behavior is inherently represented (e.g., through reposting networks). Yet, only a few VA systems provide a flexible, scalable, and comprehensive platform for comparing diffusion simulations using different ID models [ADL⁺22]. Even in those cases, visual scalability remains a challenge, as the large quantity of nodes and edges makes it difficult to assess the role of individual activation chains among the global ID process [SGM⁺23]. This problem is amplified further when attempting to provide visual encodings for the comparison of the ID models, as currently, there is no specialized methodology for comparison within the application domain of ID $[ADL^+22]$.

1.1 Research Questions

Given the circumstances described above, we formulate our research questions that we try to answer within this thesis. The main research question is split further into 3 additional sub-questions

• How can VA be leveraged to explore and analyze diffusion processes in large networks to uncover useful, otherwise hidden insights? (Q1).

To answer **Q1** it is necessary to look at three further questions, which can be categorized as sub-questions:

- Which visualization technique or combination of techniques is suitable for exploring dynamic diffusion processes over large networks? (Q2)
- Does a layered visualization approach that uses different levels of abstraction help break down the complexity of analyzing ID over large networks (Q3)
- How can two or more diffusion models be meaningfully compared in a single view such that local node-level and broader network-wide differences can be intuitively spotted and assessed $(\mathbf{Q4})$

1.2 Contribution

In this thesis, we first explore the state-of-the-art in the visualization of ID and introduce a taxonomy that generalizes across different application domains. We identify gaps in the current literature landscape and design a system that aims to offer an answer to those identified gaps. We propose *Thovea*, a **TH**ree-layer information diffusi**O**n **V**isual analytics syst**E**m for l**A**rge networks. *Thovea* progressively reduces the level of abstraction as users delve deeper into the employed three layers, facilitating an intuitive and scalable exploration of ID processes.

- The first layer offers a *high-level aggregated overview* of the network, highlighting inter- and intra-community diffusion trends.
- The second layer provides a *density-based estimation and visualization* of user-selected communities, detailing intra- and inter-community diffusion progression.
- The last layer enables *in-depth investigation and analysis* of specific regions within communities, allowing users to explore individual diffusion paths down to the node level.

By structuring the analysis across these layers, *Thovea* enables the user to first identify and assess possible regions of interest in an overview, before delving deeper into parts of the network that are of interest. This reduces visual clutter and provides a smooth and intuitive interactive navigation between layers, while preserving the user's mental map, maintaining spatial and structural consistency across different abstraction levels. Moreover, *Thovea* supports concurrent loading and comparison of up to four different diffusion simulations across all layers, each employing a distinct visual metaphor suited to its abstraction level. This layered, multi-perspective visualization approach represents the core innovation of *Thovea*, offering a flexible and comprehensive tool for analyzing ID processes in large networks.

Finally, we showcase the use and value of Thorea via two case studies and an ICE-T $[\rm WAM^{+}19]$ evaluation

1.3 Structure

The rest of the thesis is structured as follows: Chapter 2 presents an overview of the literature related to our contribution and introduces a taxonomy that categorizes contributions that offer visualization platforms for ID-related problems. It further enables us to frame *Thovea* in the current state-of-the-art. The design and feature set of our system itself is detailed in Chapter 3. Chapter 4 presents our twofold evaluation process. In Chapter 5, we provide a summarizing discussion while elaborating on limitations and possible future research directions in the field of ID. There, we also assess whether our research questions have been answered to a meaningful extent. We give a summary of this thesis in Chapter 6.



$_{\rm CHAPTER} \, 2$

Related Work

In this section, we present and discuss the literature in the relevant fields of research. First, we provide a concise overview of the relevant ID models (see section 2.1). Then we discuss the visualization of large graphs (see section 2.2). Finally, we introduce a comprehensive taxonomy that puts structure into the field of visualization of ID (see section 2.3).

2.1 Theory of Information Diffusion

Various models exist to capture the real-world phenomenon of how information spreads across a network or a population. Significant interest in such models has been expressed in the domain of social networks [LWGZ17, GHFZ13] and also more specifically within the context of the IM problem [LFWT18b, AGB⁺21]. Classical compartmental models used in epidemiology [Bra08], such as the basic Susceptible-Infected-Removed (SIR) model, allow for subjects to recover from an infection and become immune to further infections. Besides the basic SIR model, other versions have been introduced. The Susceptible-Infected-Susceptible (SIS) model does not consider the ability to become immune after an infection, but rather enables the repeated infection of subjects. The Susceptible-Infected-Removed-Susceptible (SIRS) model assumes that a removed subject can become susceptible again with a given probability α [LWGZ17]. These compartmental models have been further developed to more accurately describe diffusion processes that are happening over social networks [LWGZ17]. Sulis and Tambuscio [BAEA11] model the spreading of fake news as a competition between "Fact Checkers" and "Believers" adopting the classical SIR model. Apart from the compartmental models that are often used in epidemiological applications [BGG⁺¹¹, YDH⁺¹⁷, MLR⁺¹¹], there exist models that are commonly used for the IM problem, namely the IC and the LT model [LFWT18b]. These stochastic models consider the probability assigned to each edge or

node, respectively. If the given threshold is passed, then the corresponding edge or node is activated, respectively. These probabilities can be uniform or set individually.

2.2 Large Graph Visualization

Defining what constitutes a large graph remains an open question, as no standardized threshold exists [vLKS⁺11]. In literature, graphs ranging from thousands to millions of nodes are commonly considered large, with edge density and connectivity playing a crucial role in this classification. This poses significant challenges for visual scalability [YAD⁺18], which refers to the ability of a visualization to effectively represent large graphs, as large graph visualizations are prone to visual clutter[vLKS⁺11]. Additionally, cognitive scalability [YAD⁺18] is a key consideration. Large graphs, especially when combined with dynamic ID processes, hold a vast amount of information and complexity that, if conveyed inefficiently, can overwhelm the cognitive capacity of the user and hinder effective analysis.

2.3 State-of-the-Art: Visualization of Information Diffusion

In this section, we will first detail our methodology for conducting the state-of-the-art on the visualization of ID (see section 2.3.1), In Section 2.3.2, we introduce our taxonomy and subsequently go through every major goal that the approaches are designed to achieve (see section 2.3.3 - section 2.3.8). We discuss our results and point to gaps in the literature in Section 2.3.9.

2.3.1 Methodology

In this section, we will in detail outline our methodology for this state-of-the-art. To gain a good understanding of the field and collect relevant papers, a thorough literature research was conducted and concluded with a total of 28 contributions, 19 of which are journal articles and nine conference papers.

The root of the research consisted of Li et al.'s Survey on Information Diffusion in Online Social Networks: Models and Methods [LWGZ17] and a novel VA approach to IM by Arleo et al. [ADL⁺22]. The search for contributions was iterative. Through a combination of backward, forward, and keyword searches, the collection of seemingly relevant papers kept expanding. For the backward search and forward search, we used the References and Cited by functionality of the corresponding library.

In parallel, a keyword search was conducted. We compiled a list of relevant keywords into two sets, as seen in Table 2.1: The first set included all keywords that relate in some way to diffusion processes for which the root survey [LWGZ17] was helpful. The second set contained all those keywords that relate to visualization. Finally, we crafted all possible combinations of keywords between the two sets to conduct the keyword search: "Influence

Set 1	Set 2
nce Maximization	Visual Analytics

500 2
Visual Analytics
Visual Analysis
Visualization

2

Table 2.1: Selection of keywords split into two sets

Maximization" AND "Visual Analytics", "Influence Maximization" AND "Visualization", and so on. The utilized search engines were IEEEXplore¹ and Google Scholar².

During the literature research, we set specific exclusion/inclusion criteria to avoid unfitting approaches. However, we also did not want to be too specific and disqualify approaches that serve certain application domains. This reasoning led us to set the following three criteria for inclusion:

- The paper has to introduce a VA system that offers a platform to explore and analyze complex data through a set of interactive elements. We distinguish such systems from rather static visualizations that offer no or very limited interaction and do not enable an in-depth exploration of data [MBO⁺22].
- Further, the paper has to implement [ADL⁺22] or adapt [ST20] an existing model. We also consider approaches that use data that inherently describes diffusion behavior, such as reposting [YWL⁺14] or citation networks [HSS⁺20]
- Lastly, we consider approaches that use a network or graph structure to describe diffusion processes or use data that has a network or graph structure. We refer to this underlying network as a "medium" network. We also include epidemic models

¹https://ieeexplore.ieee.org/Xplore/home.jsp
²https://scholar.google.com/

that describe the diffusion process as a set of differential equations [MLR⁺11], as it is possible to model such "medium" networks as a complete graph.

We assessed the abstract and conclusion of every paper that, by its title, seemed sufficiently relevant. We preliminarily included those papers that provided a novel VA approach to an ID problem. The emphasis hereby lies in VA. We excluded simple visualizations of diffusion models and processes that did not offer any or very limited interactivity.

This thorough literature research led us to a list of 35 papers. We further examined those compiled papers in more detail and disregarded eight of them. The reasons for this exclusion were: the existence of a more up-to-date version and an insufficient presence of VA tools. This process led us to our 19 journal articles and nine conference papers. The conference papers were presented mainly at IEEE conferences (eight papers) and one in ACM CHI. The journal articles have the following distribution of publication journals:

- IEEE Transactions on Visualization and Computer Graphics 10 papers
- ACM Transactions on Intelligent Systems and Technology 2 papers
- ACM Transactions on Knowledge Discovery from Data 1 paper
- Electronic Proceedings in Theoretical Computer Science 1 paper
- Journal of Visual Languages & Computing 1 paper
- BMC Infectious Diseases 1 paper
- Information Visualization 1 paper
- Computer Graphics Forum 1 paper
- IEEE Transactions on Knowledge and Data Engineering 1 paper

Since we have collected approaches from different application domains, the categorization has to be suitable for this diversity. Ranging from popular topic detection of social media data $[NYX^+12]$ to epidemic monitoring and decision-support systems $[YDH^+17]$, the taxonomy must be able to offer a degree of generalization across multiple application domains. In our case, it is rather ineffective to craft a categorization that focuses too much on the application-specific properties of the data. Some surveys focus on specific application domains such as social networks and therefore discriminate approaches based on criteria inherent to that domain, for example, keyword-, topic-, and sentiment-based $[WCG^+16]$.

Given the above reasons, we generalize across application domains and decide to focus on the overarching abstract main goal that the VA sytsem aims to achieve. We look at what kind of tools are offered and to what extent the user can interact with the diffusion process in question to reach a specific goal. Furthermore, we also consider the type of visualizations that the approaches use, meaning what type of visualization metaphor or structure is present (maps, node-links, matrices, and so on).

To assess the main goals of the VA systems, we first had to get a good understanding of each one of them. We extracted preliminary goals from each paper and then, through an iterative process, assessed and compared them with each other. For this process, we had to focus not only on what features are offered by the system but arguably more so on the intended use case and output of it as stated by the respective authors and indicated by their evaluation. Features do overlap across different domains and systems, and often the combination of features or the lack of them plays a part in the decision process in assigning a main goal. To also represent this dimension, the category "Utility Features" is also introduced to indicate possible key features in the main goal.

2.3.2 The Taxonomy

We introduce a categorization of every approach collected in the context of this survey in Table 2.2 and distinguish three main criteria: The visualization method, the main goal that is achieved by the VA tools, and a set of utility features provided by the VA system [UA24]. Cells in dark red indicate the main visualization technique of the approach. Cells colored in light red mark visualization techniques that provide additional context, navigation functionality, or more detailed investigation. The utility features are also marked with dark red colored cells. The table is primarily sorted by the main goal. In total, the taxonomy distinguishes between:

- six main goals (Diffusion Exploration, Spread Simulation, Influence Maximization, Influence Summarization, Model Comparison, Anomaly Detection),
- eight visualization methods (Node-Link, Matrix, Map Metaphor, Map, River Metaphor, Tree, Timeline, Storyline),
- three utility features (Key-Player Detection, Comparison Tools, Model/Parameter Tweaking).

2.3.3 Diffusion Exploration

13 of our collected approaches offer an extensive toolkit for the exploration of diffusion networks to unfold patterns, analyze trends, and detect key influencers. Methods that are presented in this section focus on the overall exploration of diffusion processes via a diverse toolkit and provide a comprehensive dashboard to assist in that goal. They use datasets that inherently contain diffusion behaviour (e.g., reposting networks). All but two of those apply to social network data. The other ones address policy diffusions and epidemics. A variety of different visualization methods have been tried in this context.

2. Related Work

Approaches	Node-Link	Matrix	Map Metaphor	Map	River Metaphor	Tree	Timeline	Storyline	Goal	Key-Player Detection	Comparison Tools	Model/Parameter Tweaking
[NYX ⁺ 12]									Diffusion Exploration			
[STP ⁺ 17]									Diffusion Exploration			
[HNWC23]									Diffusion Exploration			
[CCL ⁺ 17]									Diffusion Exploration			
$[CCW^+19]$									Diffusion Exploration			
[CLCY20]									Diffusion Exploration			
[MBB ⁺ 11]									Diffusion Exploration			
[WLY ⁺ 14]									Diffusion Exploration			
$[SWL^+14]$									Diffusion Exploration			
$[DXZ^+14]$									Diffusion Exploration			
$[BPW^+21]$									Diffusion Exploration			
[YWL ⁺ 14]									Diffusion Exploration			
$[YJZ^+25]$									Diffusion Exploration			
[VLDBF15]									Spread Simulation			
[SRMV16]									Spread Simulation			
[BAEA11]									Spread Simulation			
[MLR ⁺ 11]									Spread Simulation			
[BGG ⁺ 11]									Spread Simulation			
[YDH ⁺ 17]									Spread Simulation			
[AME11]									Spread Simulation			
$[LWY^+20]$									Spread Simulation			
$[SGM^+23]$									Spread Simulation			
$[ADL^+22]$									Influence Maximization			
[LW14]									Influence Maximization			
$[\mathrm{HSS}^+20]$									Influence Summarization			
[STTL15]									Influence Summarization			
[VKPM15]									Model Comparison			
$[\mathrm{ZCW}^+14]$									Anomaly Detection			

Table 2.2: Taxonomy of all 28 contributions with three main classifications: visualization technique, goal, and utility features

10



(a) D-Map+ map metaphor with description of map [CCW⁺19].



(b) *R-Map* map metaphor with description of map [CLCY20].

Figure 2.1: Map metaphors used by D-map+[CCW⁺19] (a) and R-Map[CLCY20] (b).

Map Metaphor

Chen et al. introduce D-Map+ [CCW⁺19], a VA tool that leverages a map metaphor to convey diffusion characteristics of reposting networks. The map consists of many little hexagons that, based on their specific color and geometrical properties, convey different semantics as seen in Figure 2.1(a). It enables a rather intuitive examination of key players and communities, as both are visually highlighted. Another approach that makes use of the map metaphor is *R-Map* [CLCY20]. A set of geographical metaphors, such as *rivers*, *lakes*, and *countries*, encode specific network and diffusion properties of a reposting network. Overlaid keywords visually encode the overall sentiment toward the topic. An example can be viewed in Figure 2.1(b).

E-Map by Chen et al. $[CCL^+17]$ also uses the map metaphor to convey social media diffusion behavior via geographical structure. Marcus et al. $[MBB^+11]$ introduce a method to aggregate and detect social media events in a timeline and map view. The timeline shows peaks for user-selected overarching events and labels them automatically. Geographical information and sentiment are visualized through a standard world map. While D-Map+ [CCW⁺19] and R-Map [CLCY20] conduct user studies to evaluate their systems, E-Map [CCL⁺17] only uses case studies.

River Metaphor

Sun et al. $[SWL^+14]$ introduce a novel "coopetition" model that models the interplay between competition and cooperation of social media topics, as seen in Figure 2.2(a). The temporal trend of "coopetition" is then visualized through a river metaphor. Each flow represents a different topic, which through its evolution can change its "coopetition" power. A negative value indicates that the topic competes with other topics to gain attention, and a positive value suggests cooperation between topics for attention. The visualization encodes this information by varying the height of the flow. Higher flows have a positive "coopetition" power and are also color-coded green. Lower flows are colored red and have low "coopetition" power. The visualization also integrates purple flows for



(a) *EvoRiver*'s visualization approach with its components: (a) main river-like visualization, (b) *playfair*-style chart to investigate "*coopetition*" *power*, (c) comparison view, and (d) word cloud view and topic leader flows [SWL⁺14].



(b) Interface of *OpinionFlow* with the main river-like view in the middle [WLY⁺14].

Figure 2.2: River metaphors for diffusion exploration used by $EvoRiver[SWL^+14]$ (a) and $OpinionFlow[WLY^+14]$ (b).

corresponding topic threads that represent topic leaders. This enables the identification of key influencers. A word cloud enables the viewing of keywords related to a specific topic, and a comparison view shows the detailed interplay between two selected topics.

Wu et al. $[WLY^+14]$ present *OpinionFlow* and combine a Sankey graph with a tailored density map to visualize the diffusion of opinion in social networks in a river-like way. A stacked tree describes the topic hierarchy, from which topic flows can be opened by the user, as seen in Figure 2.2(b). Those also encode the overall sentiment towards the topic. Green represents positive and red negative sentiments. Additionally, a node-link diagram can be put onto the topic flow to investigate the influence of specific users.

Both approaches conducted case studies and expert interviews to validate the usefulness of their respective system.





(a) Whisper [NYX⁺12] interface with the main visualization in θ .

(b) Interface of *SocialWave*[STP⁺17] with its spatial view (b) and temporal view (a).

Figure 2.3: Node-link diagram for diffusion exploration by *Whisper* $[NYX^+12]$ (a) and *SocialWave* $[STP^+17]$ (b).

Node-Link

Cao et al.'s Whisper [NYX⁺12] combines a node-link diagram with a world map to explain the process of ID in real-time using Twitter³ data. The design is based on disc florets of sunflowers. Inside the inner circle are yet not retweeted tweets (inactive), and on the inner circle are those that have been reposted at least once. Circular user groups are placed on the outside ring around the center. A diffusion path is drawn between an active tweet and a user group if they retweeted it. The pathway encodes the time through small perpendicular lines and the sentiment in color. The background map indicates either in longitudinal or latitudinal mode the geographic diffusion pattern. Key influencers can be traced by analyzing the number of diffusion paths leading away from a specific tweet. This can be seen in Figure 2.3(a). To evaluate their system, the authors combine a performance evaluation, a case study, and an expert interview. They conclude that their predefined design goals are met.

SocialWave, a spatio-temporal social network diffusion visualization, is presented by Sun et al. $[STP^+17]$. Taking into account linguistic and cultural proximity, a weighted node-link diagram encodes diffusion paths via edges and locations via differently sized nodes. The size is dependent on the salience of the location regarding selected hashtags. Information about the temporal trend is provided separately via a timeline view. The interface of *SocialWave* can be viewed in Figure 2.3(b). After two use cases and a user study, the authors conclude that the node-link diagram is not as intuitive to encode geospatial information as common geographical views.

Yin et al. [YJZ⁺25] introduce *BloomWind* that enables the user to explore cross-platform diffusion patterns and investigate associated and relevant social media posts. They use hierarchical glyphs that contain the most influential players at the center. Glyphs are

³https://twitter.com. formerly "Twitter", now called "X"



Figure 2.4: Interface of BloomWind [YJZ⁺25]



(a) Interface of PDViz [HNWC23] includ- (b) Interface of dashboard by Baumgartl et al. ing the main matrix view. [BPW⁺21] with the main storyline view (3).

Figure 2.5: Diffusion exploration via matrix view [HNWC23] (a) and storylines [BPW⁺21] (b).

connected, encoding the diffusion pattern, similar to a node-link visualization. The VA platform uses a specialized ID model that estimates the probability of a social media contribution spreading to different social media platforms. Yin et al. [YJZ⁺25] show the usefulness of their contribution through two case studies and expert interviews.

Matrix

Han et al. [HNWC23] introduce PDViz. A VA tool to explore the diffusion of policy among U.S. states. Figure 2.5(a) shows that the main exploration takes place in a matrix view. Each state's policy diffusion measure is listed for several different policy topics. Additionally, the diffusion between every two states is quantified by a selectable measure



(a) Simulation interface by von Landesberger et al. [VLDBF15] with the input graph in the middle, performed simulation on the left and parameter settings on the right.



(b) Comparison of multiple simulation runs and their respective time steps [VLDBF15].

Figure 2.6: Spread simulation with the system of von Landesberger et al. [VLDBF15].

in a matrix-like arrangement. Navigation and exploration are aided by a map-like view and a timeline view. After two case studies and an expert interview, the authors conclude that *PDViz* provides results that are consistent with the knowledge of domain experts and that their system, at least for experts, is easy to learn.

Storyline

In the application domain of epidemics, Baumgartl et al. $[BPW^+21]$ propose an extensive dashboard to detect epidemic outbreaks (Figure 2.5(a)(1)), trace transmission pathways (3), view patient location (4) and contact (2) information. A coordinated interface enables domain experts to effectively trace back transmission paths and identify key transmission events. The storyline view represents patients as distinct lines and illustrates their location via the y-axis and the flow of time by the x-axis. Patient lines that pass closely indicate potential contact. The lines are color-coded and express infection statuses. Their system has been evaluated on a real-world dataset from an epidemic outbreak in a hospital and provided positive results. An additional expert interview also indicates the usefulness of this approach.

2.3.4 Spread Simulation

Besides the general exploration of diffusion processes, some approaches focus on the simulation and prediction of diffusion processes. Nine of our collected papers fall into this classification. Half of them specifically apply to the domain of epidemic and pandemic response, three to social networks, and one to the financial domain. We identify two main visualization methods.

Node-Link

Von Landesberger et al. [VLDBF15] implement a VA tool to simulate the contagion process via node-link diagrams (Figure 2.6(a)). By varying seed points and parameters,



Figure 2.7: Interface of DiVA [SGM⁺23] in dual diffusion visualization mode.

one can collect a set of simulations. The approach offers a comparison view to evaluate those, as seen in Figure 2.6(b). Each colored line is a separate simulation. Each node represents the graph structure at a specific time step, indicated by the node's labeling. The distance between nodes from different simulations encodes a similarity measure. The closer the nodes, the more similar the simulation at that specific time step. A case study has been conducted on a financial network. A similar approach is presented by Skianis et al. [SRMV16]. Sulis and Tambuscio [BAEA11] simulate the spread of misinformation in social networks using a node-link diagram and the *SBFC model* ("Susceptible", "Believer", "FactChecker"). They verify their simulation results by comparing them with ones from previous work in this field. Sahnan et al. [SGM⁺23] have introduced DiVA, a domain-agnostic, scalable VA Platform for ID on large networks. It employs a simple node-link visualization and color-codes the nodes depending on their current status in the simulation. They offer a comparison view that can toggle between a side-by-side and single-view mode (see fig. 2.7). They evaluate their system through a comparative study and a case study.

Map-based

Approaches in this category use geographically accurate maps to visualize their simulated diffusion process. *PanViz* by Maciejewski et al. [MLR⁺11] offers an environment (Figure 2.8 (a)) for evaluating the impact of decision measures taken in response to a simulated pandemic. They focus on the following decision measures: *school closure*, *media alerts*, and *strategic national stockpile deployment*. A map visualizes the spread based on spatial





(b) Decision-support view to evaluate consequences of different decision measures. [AME11].

Figure 2.8: *PanViz* interface [MLR⁺11] (a), Decision history tree [AME11] (b).

diffusion. The red color indicates the severity of the selected measuring criteria. Afzal et al. [AME11] add a decision-support history tree view, as seen in Figure 2.8(b). Here, simulations are represented as lines along an x-axis that encodes time and a y-axis that plots the deviation from the baseline simulation. Utilized decision measures are marked as glyphs. These measures split the line into two scenarios.

Broeck et al. [BGG⁺11] also use a geographical map to visually investigate epidemic spread (Figure 2.9(a)). They also offer a *Model Builder* (b) with which the user can modify the underlying model of the simulation, from simple *SIR* models (*Susceptible*, *Infected*, *Recovered*) to more complex model structures. *PandemCap* [YDH⁺17] also offers a decision support environment that visually conveys the spread via a geographical map and various charts in a tile-based layout.

All these approaches use case studies to demonstrate the usefulness of their respective system. PanViz [MLR⁺11] also states that their system was used in an educational context to showcase the consequences of social distancing during a disease outbreak.

2.3.5 Influence Maximization

Given a network, we sometimes want to know how the maximal number of nodes can be activated or reached during a diffusion process, given a set of seed nodes. This algorithmic problem tries to identify a set of initial spreaders that result in the biggest possible influence spread [LFWT18a]. Arleo et al. [ADL⁺22] introduce VAIM, a VA tool for the IM problem, as seen in Figure 2.10(a). It enables a comparison of two different simulations step by step. A density matrix provides a high-level schematic view of the graph structure (B), by changing the cell color intensity based on the node count. The Diffusion Matrix (D) color codes spread distribution based on the number of likely active nodes in the cell. The node-link view (E), to explore sub-areas of the diffusion matrix, shows active and non-active nodes and edges, and color codes their corresponding probabilities. The layout of the systems makes it possible to compare four simulations



(a) *GLEaMviz* map view of epidemic spread simulation (b) Model Builder tool in GLEaMviz [BGG⁺11] [BGG⁺11]

Figure 2.9: *GLEaMviz* [BGG⁺11] map interface (a), and model builder (b)



(a) VAIM [ADL⁺22] interface: Density Matrix(B), loaded simulations(C), Diffusion Matrix(D), Node-Link view(E).



(b) Interface of *Eiffel* [HSS⁺20]: Node-Link view of summarized Influence Graph (a), timeline view with animation controller (b).

Figure 2.10: VAIM [ADL⁺22] interface (a), and *Eiffel* [HSS⁺20] interface (b).

side-by-side. By selecting the model, modifying the seed selection technique, or setting a *seed "budget"*, the user can investigate different IM strategies. The effectiveness of VAIM is indicated by two case studies that include experts and an ICE-T evaluation [WAM⁺19]. In the marketing domain, Long and Wong [LW14] leverage VA and offer a set of tools for viral marketing. As they include location information, they present a map-based approach. A demonstration explains the use case of their system.

18



(a) Valiet et al.'s [VKPM15] interface to compare different models by using graph (b) A trend glyph in #FluxFlow [ZCW⁺14]. Purplerewriting rule sets. colored circles indicate a certain degree of anomaly.

Figure 2.11: Model Comparison [VKPM15] (a), and #FluxFlow [ZCW⁺14] (b).

2.3.6 Influence Summarization

For the summarization of time-evolving influence graphs, Huang et al. [HSS⁺20] propose new edge summarization algorithms for nodes, edges, and the temporal dimension to visualize the evolution of citation influence networks. A tree-like flow map depicts the influence flow in the network (Figure 2.10(b)). Starting from a source node, the network shows the flow of influence to node groups, where each group contains a certain number of clustered nodes (a). The user can tweak the number of clusters, the similarity threshold, and the minimum citation number to filter more important work. In a timeline view, the animation controller can be used to view the temporal evolution of the influence graph (b). A user study and two case studies highlight the usefulness of this approach.

2.3.7 Model Comparison

Vallet et al. [VKPM15] lay their focus on the comparison of different diffusion models. The user can formally describe diffusion models through a set of graph rewriting rules and application strategies. The paper presents the transformation of the LT and the IC models into such as set of rules. A set of rules can be seen in Figure 2.11(a)(6). The Derivation Tree (3) keeps track of the rule application history. Each node here represents the graph structure at a certain *propagation step*. The length of the tree provides insights about the model's *propagation speed*. The authors address possible scalability issues.

2.3.8 Anomaly Detection

Zhao et al. $[ZCW^{+}14]$ use machine learning algorithms to extract anomalies from a Twitter dataset. The challenge and main focus here lie in distinguishing unconventional patterns, like the dissemination of rumors or misinformation (anomalies), from more traditional trends such as popular topics and newsworthy events, which is rather found in *Diffusion Exploration*. Via their VA interface #FluxFlow, they present each thread through a river-like method, as seen in Figure 2.11(b). Users are represented as color-coded circles, where purple indicates anomalous behavior. The evolution of time is

following the x-axis. The leftmost thread glyph summarizes the sentiment and overall anomaly in the thread. The user is provided tools to change anomaly thresholds, compare threads, and identify anomalous users by tracking them across multiple threads. A clustered tree view enables the navigation between multiple threads. The effectiveness of the model is evaluated and shown with two case studies and an expert interview.

2.3.9 Discussion

In the following subsections, we go through some discussion points that each focus on different findings and compare and relate the collected approaches in more detail. In the last subsection, we will highlight existing gaps and explain how *Thovea* fills these gaps.

High Variety in Diffusion Exploration. Among the approaches that provide novel VA tools to explore diffusion processes, there is a rather great variety of employed visualization techniques. One dimension of consistency seems to be the use of some timeline view, be it in the main view or an assistive view. This is not surprising, since all methods explore diffusion processes with temporal evolution.

SocialWave $[STP^+17]$ and Whisper $[NYX^+12]$ both make use of a node-link structure. However, they use a different approach to represent spatial information. While SocialWave distributes its nodes across the view according to the geographic location of the data, Whisper uses a more compact technique. Giving up one dimension, Whisper offers the choice between longitudinal and latitudinal representation. Combined with an underlying world map, it encodes the spatial information in a compact and, in our opinion, more elegant way.

We presented three different approaches from Chen et al. that use map metaphors: E-Map [CCL⁺17], D-Map+ [CCW⁺19], and the latest one, R-Map [CLCY20]. D-Map+extends on D-Map [CCW⁺16] and focuses on event analysis. As does E-Map. R-Map lays its focus on a single microblog and offers an in-depth diffusion analysis. One difference that can be noticed quickly is the different designs of the map metaphor. While D-Map+builds maps out of many little hexagons, the other two use more natural maps that use topographical features like countries and rivers to encode semantics. R-Map, however, uses a more schematic and colorful approach than E-Map, which, in our opinion, is more intuitive to read from.

We also take a closer look at the two approaches that use river metaphors to convey the diffusion process. Specifically, *OpinionFlow* [WLY⁺14] and *EvoRiver* [SWL⁺14]. Both were published in Issue 12 of IEEE Transactions on Visualization and Computer Graphics in 2014 and offer a VA environment to explore the diffusion of topics in social media in some way. Both use case studies to apply their approach to Twitter data. *OpinionFlow* focuses on the investigation of opinion propagation in the context of different topics and events. It intuitively integrates sentiment data into the visualization and makes it easy to analyze the flow of opinion propagation among users. *EvoRiver* approaches a different problem. It prioritizes the relationship between topics. Using a "coopetition" model,
EvoRiver intuitively depicts the flow of interplay between multiple topics over time. Color coding in both approaches helps in acquiring a quick overview of the situation.

Key-Player Detection in Diffusion Exploration. When we look at the utility features defined in Table 2.2, it can be seen that tools for the detection of key players appear in all but two approaches. Depending on the actual application domain, the tools enable the identification of topic leaders [SWL⁺14], key U.S. states in policy diffusion [HNWC23], and patient zero in a hospital epidemic situation [BPW⁺21].

Interactive Parameter Tweaking and Comparison in Spread Simulation. When simulating a spreading process, there is a need to adjust and tweak parameters and models such that the simulation becomes more accurate and fulfills the user's requirements. This is also apparent when we look at our categorization. Every approach offers a set of parameters that can be adjusted to achieve a more suitable simulation. When it comes to comparison tools, the situation looks different. Von Landesberger et al. [VLDBF15] enable a side-by-side comparison of multiple simulation runs. Their system quantifies the similarity between simulations for each time step and intuitively visualizes the results. With a glance, the user can identify similar simulations.

In the domain of pandemic and epidemic response, decision-support features play an important role $[MLR^+11][YDH^+17][AME11]$. These decision measures are applied to the simulation at different time points and different combinations. To evaluate the most effective decision measures, some sort of comparison tool would be useful. Afzal et al.'s [AME11] Decision History Tree covers this need. Through a rather compact timeline view, the baseline simulation line is split once a decision is made. This can be repeated for a series of measures and results in a tree where the distance to the baseline and alternative pathways encodes insightful information about the effectiveness of the measures. *PandemCap* [YDH⁺17] also offers comparison tools, however, they are rather simple. Here, various scores of different simulations can be plotted in the same line chart. Rather than having one graph where information about the effectiveness of decision measures can be extracted, the user has to view multiple separate graphs.

Comparing with Different Goals in Mind. Here, we want to briefly look at VAIM's [ADL⁺22], DiVA's [SGM⁺23], and Vallet et al.'s [VKPM15] comparison environments. Even though all of them offer a comprehensive interface that is geared toward comparing multiple models and simulation runs, they have notable differences, as, looking at the taxonomy, we notice that they strive for different goals. VAIM implements popular IM algorithms and models, which can then be compared to maximize the propagation in a network with features like "budgeting" seeds and modifying the seed selection technique. Vallet et al.'s system, different from VAIM, facilitates the creation and application of a set of common graph rewriting rules with which different diffusion models can be expressed and compared. They do not focus on IM but rather create a formal common language to describe and compare different ID models over networks while giving the user the ability to modify the defined models. DiVA [SGM⁺23], similarly to VAIM, offers a dual visualization mode where two simulations can be compared side by side, as seen in Figure 2.7. In contrast to Vallet et al. [VKPM15], they enable either a

scalable intra-model or inter-model comparison, where differences in hyperparameter configurations or selected models are assessed, respectively.

Evaluation Methods. When we look at the evaluation methods that are used by the 28 collected approaches, we see that case studies seem to be a popular way of demonstrating the different use case scenarios and interaction possibilities. Some approaches also test their systems by letting people interact with them. Here, we can differentiate between user studies and expert interviews/ expert feedback. For expert interviews, respective domain experts are consolidated, while for normal user studies, no experts are needed. Approaches that also introduce a novel diffusion model or algorithm tend to utilize quantitative performance evaluation [STTL15] [LWY⁺20]. Arleo et al. [ADL⁺22] evaluate their visualization by making use of an existing value-driven heuristic called ICE-T [WAM⁺19].

Literature Gap. Most VA platforms that explore different facets of the diffusion process are designed to address specific problems within their respective application domain. [CCW⁺19, CLCY20, CCL⁺17, BPW⁺21, HNWC23, WLY⁺14, SWL⁺14], limiting their generalizability—an area where *Thovea* excels. Domain-agnostic VA applications, while more flexible, often exhibit significant shortcomings. Looking at DiVA [SGM⁺23], we notice that especially while exploring large graphs, visualization techniques that solely utilize node-link diagrams can suffer from visual clutter that makes tracking and identifying distinct diffusion pathways difficult, especially when they change over time. Unlike Those, DiVA keeps edges static and does not encode their role in the simulation, preventing users from exploring distinct activation pathways. Similarly, VAIM [ADL⁺22] does not provide visual encodings for this information, but rather encodes the probability of that edge to activate within the context of the respective simulation. While DiVA supports the comparison of two simulations, Thovea matches VAIM $[ADL^+22]$ and offers the concurrent comparison of up to four simulations. However, both DiVA and VAIM struggle with visual scalability, particularly when analyzing networks with thousands of nodes. Additionally, Vallet et al. [VKPM15] specialize in the assessment of ID models, but support networks with only a few hundred nodes and edges. By addressing these limitations, *Thovea* offers a more scalable and comprehensive approach, providing clear, interactive visual encodings for diffusion pathways while maintaining usability across large networks.

TU Bibliothek Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar WIEN Vourknowledge hub The approved original version of this thesis is available in print at TU Wien Bibliothek.

CHAPTER 3

Thovea - The System

The design of *Thovea* follows Munzner's [Mun09] nested model for design and evaluation (see fig. 3.1). It entails four layers by which we, first, define the data and tasks within the context of the problem domain, in our case ID (see section 3.1). From those, we generate high-level operations and data types (see section 3.2). Subsequently, meaningful visual encodings as well as interaction methods are identified (see section 3.3). Within the innermost layer, we implement those design decisions (see section 3.4).

3.1 Data

We model entities that are related to one another via a Network (or Graph) G. A graph G = (V, E) consists of a set of nodes (or vertices) V and edges E that are vertex pairs [BM08]. An edge consist of an unordered pair $(u, v) \in E$ if it is undirected, and an ordered pair $(u, v) \neq (v, u)$ if it is directed. Within the context of Thovea, we support graphs that either consist of just undirected or directed edges. Within an ID process, a directed edge signifies that only one entity has the ability to activate the other, whereas an undirected edge represents a reciprocal relationship, where both entities can activate each other.



Figure 3.1: Munzner's [Mun09] nested model for design and evaluation

We consider a static Graph G = (V, E) as the structure over which we simulate an ID process via a suitable diffusion model. We refer to G as a "medium" graph, meaning that this represents the base network over which we simulate an ID process. We simulate the ID process over G for a total of T time steps, where $0 \le j \le T$. We define the set of vertices $V_j \subseteq V$ and edges $E_j \subseteq E$ that contain all nodes and edges that are active, or facilitate an activation at step j, respectively. Thus, we can consider the diffusion process that is happening over G as a "diffusion" dynamic graph $D = (G_0, G_1, ..., G_T)$ where each $G_j = (V_j, E_j)$ is a static graph that represents the current status of the diffusion process at time step j. The visualization of such a combination of a static "medium" graph G and a "diffusion" dynamic graph D constitute the main challenge in our design process, as we not only have to find meaningful visualizations for the underlying base network G but also visually encode a dynamic process (D) on top of it.

We use ID models to encapsulate real-world dynamics of a diffusion process over a "medium" graph G. During this process, a vertex V becomes *active*, a term that generalizes across different domains. For example, in epidemiology, activation may represent an individual becoming infected, while in a social network, it can indicate a user being influenced by another. Accordingly, an edge (whether directed or undirected) becomes *active* if it facilitates an activation from one of its vertices to another. For undirected edges, the ID simulation assigns a direction to an edge to encode the specific flow of activation between the two vertices.

We use the NDLib [RMR18] to compute the simulation over a given network. We support the IC and LT models, as well as compartmental models, such as the SIR, which introduce the concept of node deactivation (e.g., gaining immunity after infection) [LWGZ17]. However, some problems in ID require a comparison to assess the infectiousness, reach, or effectiveness of a given diffusion simulation. This could be the need to evaluate an optimal seed node selection for the problem of IM [LFWT18a] or to investigate whether or to what degree disease prevalence varies based on countermeasures taken [BGG⁺11]. The investigation of such differences needs to be accommodated.

3.2 Users & Tasks Analysis

The target users of this system are knowledgeable about information visualization and able to interpret modes of network visualization. Our approach is not tailored to any specific research domain, which would typically require specialized tools or features. For instance, in epidemiology, geographical maps are often used in visualizing infection processes [PL20], which may not be applicable to other areas. Instead, we aim to provide a domain-agnostic platform designed to address a wide range of problems within the ID research field. Thus, our target users do not need knowledge specific to application domains, however, they are knowledgeable in network visualization and interested in diffusion visualization and algorithms.

By reviewing the literature in the research field of ID in visualization, as described in Section 2.3, and considering our research questions stated in Section 1.1, we collect

a set of common tasks across domains of application where ID processes are relevant. Ultimately, *Thovea* is designed to support the following tasks:

- **T1 Explore** the "medium" network's topology, community structures, inter-/intracommunity relationships, and the "diffusion" dynamic ID process.
- **T2 Compare** multiple diffusion simulations concurrently, to analyze differences and similarities in diffusion patterns.
- **T3 Navigate** the temporal evolution of the diffusion processes, tracking changes over time.
- **T4** Gain a comprehensive **overview** of the "medium" network, its communities, as well as the "diffusion" network and its diffusion trends—globally, i.e. across the "medium" network, and locally, i.e. within the individual communities.

As an example, in the domain of epidemiology, **T1** involves analyzing disease spread across a population, identifying community structures and diffusion patterns, while **T2** allows for the comparison of disease simulations under different interventions, such as vaccination versus quarantine. **T3**, instead, tracks the disease's temporal evolution, observing peak infection periods and community impacts, and **T4** provides a comprehensive view of the disease's spread across the population, with detailed insights into specific communities, like age groups or geographic regions. Similarly, in social networks, **T1** investigates how information or influence spreads, uncovering community structures and inter-community diffusion, while **T2** enables the comparison of diffusion simulations for different content types, like news versus memes. **T3** tracks the evolution of influence or information over time and, finally, **T4** offers a global view of information spread, with the ability to focus on localized trends within specific communities.

3.3 Visualization Design

As the number of nodes and edges increases, graph drawing algorithms alone cannot keep up. Therefore, graph filtering and aggregation have become a prominent direction in large graph visualization [vLKS⁺11]. *Thovea* follows this direction by dividing the problem into smaller problems. As a consequence, we do not need to calculate a fitting graph layout for the whole "medium" network, but only for sub-parts of the network. We achieve this by introducing a three-layer visualization approach, where each layer provides a characteristic and integrated visual encoding to support an intuitive comparison of multiple ID processes across the different layers of abstraction.

Design Challenges. The main challenge when designing the system entailed dealing with the computational and visual complexity associated with large networks. Our first design iteration followed a single-layer Level-of-Detail (LoD) approach, where individual nodes and edges would be aggregated into cluster nodes and edge bundles, respectively.

This results in an abstract representation of the network, providing an overview of the "medium" network structure. Selecting a cluster node or edge bundle would reveal the individual elements inside this group, allowing for the interactive exploration of regions of interest while keeping the overall context of the network. In our design, the basis for the groups would be formed by the coordinates, calculated by the continuous force-directed graph drawing layout ForceAtlas2 [JVHB14]. ForceAtlas2 enables the usage of the LinLoq energy model [Noa07], which causes nodes to repulse each other like charged particles and edges to act as springs pulling together the nodes adjacent to it. It iteratively applies such forces until the node positions converge into a balanced state with minimum energy. This tends to reveal local group structures within the network visually [JVHB14].

However, we encountered two main challenges: (i) achieving a readable layout and visually distinguishing clusters from the network proved to be time-consuming, and (ii) grouping nodes into clusters based solely on visual separation was difficult. As a result, we opted to use the Louvain method [BGLL08] to assign each node to a cluster. However, this dual clustering approach introduced ambiguity and inconsistencies, making it impractical to calculate the size and position of the corresponding high-level cluster nodes.

These problems prompted us to take a step back and reevaluate our design. We tackled the scalability problem by dividing the computational load into smaller problems. Computing a layout of a network with thousands of nodes was simply too time-inefficient. Therefore, we decided to first determine the communities within the "medium" network using the Louvain method [BGLL08]. Once assigned, we could then limit the layout calculations to consider only the individual communities. Thereby, we address the problem of scalability and avoid ambiguous community assignments by only relying on one method. This approach was inspired by Onoue and Koyamada's group-in-a-box layout [OK17]. Their graph drawing algorithm is designed to highlight the predefined group structures within a graph and reorder those groups to avoid overlaps in the drawing, however, it struggles with larger networks.

Three-Layer Approach. The primary view of *Thovea* employs a three-layer visualization approach. Figure 3.2 showcases the concept of this approach: The first view is called the Overview Layer (1) and operates on the highest level of abstraction. It divides the loaded "medium" network into non-overlapping communities. The main purpose of this layer is to provide an aggregated, high-level overview of both the underlying "medium" and the "diffusion" dynamic network. The second layer enables the investigation of topology and local diffusion trends inside selected communities. It is accordingly referred to as the Local Layer (2). Here, after the initial assessment in the Overview Layer, one can delve deeper into specific communities and explore local diffusion patterns. The *Detail* Layer (3) constitutes the final step of this three-layer approach and contains low-level information about nodes and their relationships to each other. Here, distinct diffusion pathways can be revealed and explored. This approach is in line with Schneiderman's mantra, which states "Overview first. Zoom and filter. Details on demand" [Shn03]. Each layer contains its own set of interactions and visualizations, designed to fit the respective level of abstraction.

26



Figure 3.2: Flowchart of *Thovea's* layers: *Overview Layer* (1), *Local Layer* (2), and *Detail Layer* (3). One Node is highlighted throughout the layers.

Across all three layers, zooming and panning can be used to navigate within the view. We offer different ways of highlighting nodes of interest, e.g. they are marked in pink and are easier to track when switching between layers (T1) (see fig. 3.2).

Overview Layer. The main purpose of this layer is to provide an overview **(T4)** of the "medium" network and the ID process (see fig. 3.3-D). Therefore, we do not intend to show any individual nodes at this layer. Instead, the Louvain method allows us to detect community structures inside networks [BGLL08] and assigns each node of the network exactly one community. As seen in Figure 3.3-D, each community is represented via a rectangular Community Tile. The size of this tile is proportional to the number of nodes inside the respective community, allowing for a quick assessment of the communities' sizes relative to each other. We connect two Community Tiles via an Edge Bundle. An Edge Bundle represents the collection of all edges that connect two nodes from different communities. The size of such an *Edge Bundle* is proportional to the interconnectivity of the related communities and is scaled relative to the minimum and maximum observed instances to ensure a normalized representation. As a measure against edge clutter, we only show the largest Edge Bundle. We refer to such an Edge Bundle as the dominant one. Only when hovering over or clicking on a Community Tile, all the Edge Bundles are shown (see fig. 3.2(1), Community 2). If a *Community Tile* contains a highlighted node, it is visually emphasized via a pink border (see fig. 3.2(1), Community 2).

Once a diffusion simulation is loaded, there are additional visual representations that encode the diffusion trends inside and between the communities of the "medium" network for each specified time step. Each *Community Tile* contains n colored bars, where n is equal to the number of loaded simulations **(T4)** (see fig. 3.3-D; green, blue, magenta



Figure 3.3: *Thovea* interface: (A) Setup view containing drawing parameters, node search & highlight, simulation control panel, and loading tools; (B) Community-wise stats and metrics; (C) Global stats and metrics; (D) Primary visualization view—currently showing the *Overview Layer* for three simulations, (E) Panel for customization of primary visualization including the Time Control Panel.

bars). This constitutes a multivariate network visualization where the topology serves as a base representation with embedded ID-related attributes [HSS15]. The height of each bar encodes the activation rate of the respective simulation within that community. This simple juxtaposition enables the user, via a glance, to assess relative differences in activation rates between communities, as well as to compare intra-cluster reach between loaded diffusion simulations via a local side-by-side comparison (**T2**).

The user also has the option to explore the role of a community in propagating information to its neighbors. To do that, the user has to hover over a simulation bar inside the desired *Community Tile* (**T1**), which unfolds a set of two *Diffusion Flux Circles* for each *Edge Bundle* (**T4**). The term *Diffusion Flux* refers to the flow of information from one community to the other, one circle for each direction. As seen in Figure 3.2(1), hovering over Community 2 reveals green circles traveling from Community 2 to its neighbours and vice versa. The speed of this circle is determined by the activation rate of available diffusion paths in their respective directions. We calculate the speed by assigning each direction (eg. Community 1 to Community 2) a speed value S where

 $S = MIN_SPEED + ((infected/connectivity) * (MAX_SPEED - MIN_SPEED))$

 MAX_SPEED and MIN_SPEED refer to the set boundary values that the speed S cannot exceed. *infected* contains the number of inter-community edges that facilitate an

activation from the source to the target community (eg. Community 1 to Community 2). We divide this value by the *connectivity* of these communities. For undirected graphs, both directions (e.g. Community 1 to 2 and 2 to 1) have the same connectivity value, namely the total number of inter-community edges between two communities. For directed graphs, *connectivity* only considers the inter-community edges pointing in the respective direction. If, for a direction, none of the available paths are activated. the corresponding circle remains idle and is therefore omitted. This is due to the MIN SPEED variable being set to 0 within the implementation. The visual encoding regarding the Diffusion Flux Circles follows the idea behind unfolding edges [BDT23] and incorporates multivariate edge attributes dynamically. Initially, the choice of considering a motion-based approach was inspired by Ware's contributions [WB04, WB05] and its use in dynamic network visualization [VAA24] (T1). The decision was made due to the following considerations: (i) we wanted to avoid overloading the color channel, as the system already uses four colors for the visual encoding of the simulations, and (ii) the high-popup effect it would introduce. To avoid user distraction, we only show circles that are traveling from or towards the currently hovered *Community Tile*. During initial iterations, we concluded that displaying the circles of each community leads to too much clutter and distraction.

There are three customizable parameters special to this layer, as follows: (i) The Louvain method [BGLL08] accepts a *resolution* parameter that determines the size of the retrieved communities. The higher the *resolution*, the more communities will be produced. To find an optimal split, the method optimizes *modularity* [New06], which scores the relative density of intra-community edges to inter-community ones. The *Louvain Resolution Slider* (see fig. 3.3-A) enables the user to reconfigure the Louvain method to split large communities into smaller ones, so a more concise selection can be investigated in the subsequent layers; (ii) To customize the set of visible *Edge Bundles*, the user can set a relative threshold via the *Edge Visibility Threshold* (see fig. 3.3-A). Only edges with a size above a threshold will be shown. Especially for larger networks with a high number of communities, one might want to focus on the more tightly connected ones and hide the smaller bundles that would otherwise lead to edge clutter; (iii) The *Maximum Oscillation Speed* (see fig. 3.3-A) slider controls the maximum travel speed of the *Diffusion Flux Circles*. By pressing the pause symbol, the user can turn off this feature.

To investigate the local topology and diffusion propagation of a subset of communities, the user can select them by clicking the corresponding *Community Tiles* and proceed to the *Local Layer* by pressing the "Investigate" button (T1) (see fig. 3.3-E).

Local Layer. The goal of the *Local Layer* is to provide the local context of the selected communities by integrating topological and local diffusion information, while omitting detailed information about node attributes and, optionally, position. The local node layout is calculated using the force-directed *ForceAtlas2* [JVHB14] algorithm, due to its efficacy for larger networks, ease of use through existing implementations [Pli25], and good performance on networks with sizes that we also aim to accommodate [JVHB14]. *Thovea* pre-calculates the layouts for every single community and stores them in the



Figure 3.4: The *Local Layer* with two communities selected (indicated within the dashed blocks) and nodes and edges toggled on. The dominant *Edge Bundle* between the communities is shown with the thick, connecting line. Two simulations (blue vs. green) are loaded, shown at time step 0.

browser's memory. This enables seamless exploration of many communities without having to wait for loading times during the exploration (T1). Only when selecting two or more communities, we need to calculate their aggregated layout on demand, before switching to the corresponding *Local Layer*, as pre-calculating every combination of communities would be too inefficient.

As seen in Figure 3.4, the visualization method employs a Kernel Density Estimation (KDE) [and17] using the calculated positions of all nodes within the selected communities of the "medium" graph and draws the result as a contour plot, which we call *base contours*. We choose a density-based approach [ZBDS12] to accommodate the visualization of a large quantity of nodes to deal with possible clutter and overplotting (T4). After all, the selected communities could easily contain up to thousands of nodes. However, there are situations or selections where additional context in the form of node position or edges might be useful. To facilitate this situational need for additional context, we enable toggling on or off the rendering of individual nodes and edges (T1) (see fig. 3.4, bottom). Each node highlighted by the user can be identified by its pink halo (see fig. 3.2(2)).



Figure 3.5: *Local Layer* representation of one community: (a) without and (b) with additional context through nodes and edges.

We employ Wallinger et al.'s Edge Path Bundling [WAA⁺23] to reduce edge clutter (see fig. 3.4). Their implementation also specializes in eliminating *independent edge ambiguities*, which describes the problem where two independent edges are bundled together, such that two non-adjacent vertices are visually connected (see fig. 3.6). We divide the edges into two categories: inter-community edges and intra-community edges. We visually emphasize the inter-community edges, since they can be critical in identifying jumping points between communities (see fig. 3.4) (T1). Intra-community edges, depending on the loaded network, can provide additional context to the density contours and help avoid incorrect assumptions about the underlying topology (see fig. 3.5).

Once an ID simulation is loaded, we take the set of activated nodes from the "diffusion" dynamic graph at the current time step and draw *simulation contours* around them, according to a KDE calculation [and17]. We do this for each loaded simulation. As seen in Figure 3.4, this results in a superimposed representation that highlights activated sub-regions of the selected communities (**T4**), while also enabling quick assessment of regional differences of multiple simulations (**T2**) (see fig. 3.4). From a visualization perspective, we aim for an "ink drop" effect using node splatting [NBW14]. If nodes are toggled on, they are rendered in the respective colors of each simulation where it is activated at the current time step (**T2**) (see fig. 3.4). While navigating through time, the *simulation contours* dynamically evolve, continuously shifting in size and shape as they adapt to the underlying changes in node activation status. Thereby, we aim for an intuitive visual metaphor of a diffusion process evolving through time (**T3**). We allow the adjustment of the KDE *bandwidth* parameter to either sharpen or smooth the resulting density field. This provides independent control over both the *base contours* and the *simulation contours*. A higher bandwidth results in smoother contours by blending



Figure 3.6: Independent Edge Ambiguity as described by Wallinger et al. [WAA⁺23]

nearby points, while a lower bandwidth captures finer structures at the cost of potentially introducing noise.

Once the user has identified a region of interest, further investigation can be conducted. For this purpose, they can enter the *Detail Layer*. There are two interaction techniques to enable the following layer **(T1)**: (i) The user can hover over and click on a *simulation contour*. This selects every node inside the contour. (ii) Alternatively, the *selection circle* can be opened by right-clicking (see fig. 3.2(2)). The user can make a more fine-tuned selection by moving the mouse and using the scroll wheel to modify the radius. The number of nodes encircled can always be viewed at the top of the circle.

Detail Layer. The *Detail Layer* is designed to accommodate the detailed investigation of no more than a few hundred nodes. At this point, we expect the user to be fully aware of the context in which the diffusion takes place, which communities they are in, and which sub-regions they are currently investigating (e.g., see fig. 3.2(1)-(3)). Here, by right-clicking the desired node, we enable the viewing of node-specific metadata via the *node profile* tooltip window (see fig. 3.7 (b), (c)). It contains the node's label, attributes like (in/out)degree, closeness centrality, and community, as well as information regarding the loaded ID simulations. Each simulation is represented via its associated color in a column, containing: the time of activation (*infectedAt*), the source of its activation (*infectedBy*), and the time it is removed (*removedAt*). If the *node profile* lists source nodes at row *infectedBy*, they can also be highlighted by clicking on them, thereby the user can trace back activation pathways (e.g, see fig. 3.8) (**T1**). This layer can be explored via three different node-link layouts, each serving a different purpose, as described in the following (**T1**) (see Figure 4.5 (a):

ForceAtlas2: By default, when landing on the *Detail Layer*, the ForceAtlas2 [JVHB14] coordinates from the previous layer are reused (see fig. 3.7). The reasoning behind this decision is that we want to keep the mental map of the user [BB99], as we are already switching from one mode of visualization (density-based) to another (node-link), we avoid simultaneously rearranging the nodes' positions. Node labels are shown and can be resized using a font slider.

d3.forceSimulation: By selecting d3js's force layout [BOH11], the nodes are rearranged,



Figure 3.7: Detail Layer of a co-authorship network using a **ForceAtlas2** layout, with the node profiles of Munzner T. (c) and Zhang X. (b) expanded and highlighted. A minimized view of the Local Layer (a) is placed at the top left.

thus leading to a more compact and optimal representation, where less zooming and panning are required (see fig. 3.8). However, we omit node labels at this stage to avoid clutter, as nodes are now much closer.

Sugiyama-style: If, and only if at least one ID simulation is loaded, we allow the usage of a more specialized layout (see fig. 4.1(c)). Here, for each time step, we only consider the following nodes: (i) all activated nodes, except those that currently are not, or were not activated by any node within this layer, (ii) inactive nodes that were active in previous time steps and have activated at least one neighboring node within this layer that is still active. This generates a Directed Acyclic Graph (DAG), which we can leverage by employing a hierarchical graph drawing layout such as Sugiyama [STT81]. Hierarchical graph drawing layouts perform better at providing more effective visualization of cascades through networks [AP16].



Figure 3.8: *Detail Layer* of a co-authorship network using a **d3-force** layout, with two simulations active. Authors along a path between the seed and the target author are highlighted.

All three layers encode the ID process, primarily via the respective colors of the simulations (see fig. 3.8). A node is (partially) painted in the colors of a simulation if, at the current time step, this node is active within the respective simulation. Similarly, we paint edges in the corresponding colors if their target node is active. An edge, that carries an activation, encodes the direction of the diffusion via an arrowhead. This arrowhead is also painted according to the activating simulation (**T1**, **T2**). Additionally, edges can be visually distinguished into inter- and intra-community edges, using dashed and solid

lines, respectively (T1) (see fig. 3.8).

At the top-left of the *Detail Layer*, we retain a minimized view of the *Local Layer* across all selectable layouts. This ensures that contextual spatial information remains accessible, allowing the user to always see which part of the community they are currently investigating. This is an additional measure to help maintain the user's mental map when diving into the final layer (**T1**).

Auxiliary Components. Besides the primary view, which hosts the three-layered visualization, we offer three auxiliary components. The panel on the left (see fig. 3.3-A) includes tools related to loading the network and the precalculated simulation files, as well as layer-specific drawing parameters used to modify and customize the visualization. We list all loaded simulations via their specified simulation names together with their assigned color in the *Simulation Control Panel* (see fig. 3.3-A). An eye-themed button can be pressed to show or hide a simulation across all other components. We offer a *Node Search* \mathcal{C} *Highlight* panel to not only search for specific nodes in the network but also highlight specific nodes. All highlighted nodes are compiled in a list where they can be removed too (see fig. 3.3-A) (T1).

The Global Metrics panel offers basic stats and information about the loaded "medium" network and ID simulations (see fig. 3.3-C). Some of the network stats are on-demand since the calculation for large networks can be computationally intensive and time-consuming. For each loaded simulation, its respective simulation card can be viewed. It contains the name, associated color, total coverage across the entirety of the network, and a list of its seed nodes (**T4**). Seed nodes can be directly highlighted here. By pressing the circular button with the arrows, we can access the alternative Chart View. Here we plot global diffusion trends for each simulation (**T4**) (see fig. 4.10a). Depending on the model employed, we can view the evolution of nodes inside the categories, Susceptible (dotted), Infected/Activated (solid), and Removed (dashed). As seen in Figure 4.10b, the absolute and relative number of nodes in each category can be viewed in tooltips by hovering over the desired time step. This hovering is coordinated across all charts and opens tooltips for each. We thus enable an additional way to easily compare diffusion trends across simulations (**T2**). We allow setting the time step by clicking on the corresponding points in the chart (**T3**).

The panel at the bottom-middle of the system serves as an extension of the *Global Metrics* panel (see fig. 3.3-B). It shows the same type of information, specifically for the selected communities. If multiple communities are selected, then it aggregates stats and charts **(T4)**.

Accessible from every layer is also a *Time Control Panel*, enabling the navigation between the time steps of the simulations **(T3)** (see fig. 3.3-E). *Thovea* requires all loaded simulations to be of the same length. The user can navigate in steps of 1 or 10, or use the *play* button to automatically advance through time.

1

2

 $\frac{3}{4}$

 $\mathbf{5}$

6

7

8

9 10

 $11 \\ 12$

13

14

15 16

17

18

19

20

21

22

 $23 \\ 24$

25 26

3.4 System Architecture

To provide all the functionalities required to fulfill T1-4, we require a comprehensive user interface. We decided on a single-page web application using React [Rea]. The interface depends mainly on the JavaScript libraries D3 [BOH11] for the different visualization modes and graphology [Pli25] for the representation and management of a *Graph* object. Graphology also provides implementations for the ForceAtlas2 Layout [JVHB14] and Louvain method [BGLL08]. For the Sugiyama layout [STT81], we use an implementation found within the *d3-dag* [Bri23] library. Wallinger et al. [WAA⁺23] have implemented the Edge Path Bundling using the programming language Python. We therefore connect the client-side via a REST API to a Flask [Pal24] server and include a module to calculate the edge path bundles on the server upon request from the client. Thus, *Thovea* implements a client-server architecture, and its implementation can be found within the GitHub repository [mer25].

To load a graph into *Thovea*, it has to be a .json file in the following format:

```
{
    "options": {
         "type": <"undirected"||"directed">
    },
    "nodes":
              ſ
         {
             "key": "0",
             "attributes": {
                  "label": "Some Label"
         },
         {
             "key": "1",
             "attributes": {
                  "label": "Some other Label"
         },
            . . .
    ],
    "edges":
         {
             "key": "0",
             "source":
                         "0",
             "target": "1"
         },
            . . .
```

36

The graph has to exclusively contain either directed or indirected edges. Furthermore, multi-graphs are not supported. The simulation files are also .json files that have to be formatted as follows

```
1 {
2 "type": < "SIR" || "IC" || "LT" >,
3 "name": "name of simulation",
4 "iterations": <model.iteration_bunch(n)>,
5 "trends":<model.build_trends(model.iteration_bunch(n))>
6 [0]["trends"]
7 }
```

We refer to NDLib [RMR18] functions in the format file. The term *model* inside the description refers to the chosen NDLib model, e.g., model = ep.SIRModel(G), where G is an instance of NetworkX's [HSS08] Graph() or DiGraph() object A Python script to create custom simulation files and the example datasets used in this thesis can be found within the GitHub repository [mer25].

The color palette used by this system is chosen to be color-blind friendly [TP23].



CHAPTER 4

Evaluation

We conduct a comprehensive assessment of our system to ensure a thorough evaluation from both qualitative and quantitative perspectives. First, we elaborate on the datasets used in Section 4.1. Second, we present two case studies in Section 4.2 to illustrate the system's real-world applicability and effectiveness in practical scenarios. Furthermore, we apply the ICE-T [WAM⁺19] evaluation methodology in Section 4.3 to provide a structured and standardized quantitative analysis.

4.1 Datasets

For the evaluations, we deliberately select four diverse datasets, each accompanied by its own collection of ID simulation files. Two of those are shown later in our case studies. These datasets are chosen to represent a range of real-world network structures and diffusion dynamics. By selecting these datasets, we aim to demonstrate the robustness, generalizability, and adaptability of our approach.

The first dataset is an *InfoVis co-authorship network*, constructed from publications at the IEEE InfoVis conference from 1995 to 2015 [SCH⁺16]. Each node represents an author and is labeled with the author's name. An undirected edge is placed between authors who have co-authored a paper published in one of the aforementioned conferences. The final network contains 698 nodes and 1,806 edges. This network captures the diffusion of academic collaboration within our community through the span of 20 years.

The second dataset includes a network of Facebook pages associated to companies and available on the Network Repository [RA15]. Each node corresponds to a company's Facebook page, identified by its name and a unique Facebook-specific page ID. An undirected edge is placed between two nodes if the two Facebook pages have both *liked* each other. This dataset provides insights into social media connectivity and mutual relationships between corporate entities. It entails 4,704 nodes and 21,921 edges.



Figure 4.1: Snapshot of *Thovea*'s Three-Layer pipeline for Case Study 1: InfoVis Coauthorship Network. (a) *Overview Layer* of two different ID simulations (blue and green), where three communities (2, 3, 8) have been selected; (b) *Local Layer* with a density-based visualization of the selected communities just for the green simulation, and two nodes highlighted (in magenta); (c) The selected circular area from (b) is viewed through a Sugiyama layout, enabling the investigation of detailed activation pathways between the selected nodes.

Thirdly, we include a dataset representing the European Power Grid Network [Mar22] via an undirected graph. Each node of the grid includes information about the country it's located in and the energy price at that location. The network contains 6,659 nodes and 8,309 undirected edges. In this case, ID simulations model cascades of power plant failures, reflecting the vulnerability and propagation dynamics within a critical infrastructure system.

Our fourth dataset is the German QAnon Telegram Dataset [Tho21], which maps the spread of information within conspiracy-related Telegram channels. Each node in this network represents a Telegram channel associated with conspiracy theories. If a message has been forwarded from one channel to another, a directed edge is placed between them. The network consists of 3,525 nodes and 8,471 edges. An ID simulation on this network can simulate how information, including misinformation, propagates across online communities.

4.2 Case Studies

In this section, we present two case studies and showcase how the design decisions and visual encodings can be used to fulfill the tasks described in Section 3.2. For that, we have chosen the InfoVis co-authorship network [SCH⁺16] and the Facebook network of company pages [RA15].

Case Study 1: InfoVis Co-Authorship Network. For this case study, we have prepared two simulation files. Both use the IC model to simulate an ID process over this network, with one seed author each. Since the model accepts an activation probability

40



Figure 4.2: Snapshot of Overview Layer during Case Study 1 with Metrics View opened: Panel for graph selection and upload (a). Panel for simulation selection or upload (b). Community 2 (contains the target author) is selected (c). Simulation card of blue and green simulation. The green simulation card's seed node list is expanded (d). Communitywise simulation cards of green and blue simulation (e). Simulation Control Panel with both simulations visible(f)

for each edge, we assign each edge a weight that is proportional to the number of contributions two authors have co-authored. We can interpret these edge weights as a measure of the influence that one author can have over the other. After all, more frequent collaboration can manifest as the spread of a particular common topic or idea across an author's academic sphere. Even though the co-authorship network itself is undirected, as the ID process unfolds, activated edges will become directed, thereby encoding which of the two authors has influenced whom. With this in mind, we will now investigate which of the two seed authors can assert their influence further into the co-authorship network and where differences lie.

We start by loading the InfoVis dataset (see fig. 4.2(a)) and both simulation files into the system (see fig. 4.2(b)). The seed nodes for these simulations are *Miksch S*. (blue) and *Munzner T*. (green), two well-connected authors and high-profile figures in visualization research. In the context of this case study, we look for communities where there is a noticeable difference in reach between the two simulations (**T2**) at a global level (**T4**). We can identify that Community 2 expresses such characteristics (see fig. 4.2(c)). Only author *Munzner T*. can assert influence over that community. This can also be confirmed by looking at the community-wise simulation cards in Figure 4.2(e), where *Munzner T*. reaches 16.67% convergence and *Miksch S*. 0%. Switching to the chart view reveals the temporal evolution of that same behaviour in Figure 4.3(b). Additionally, the global chart view reveals that the green simulation reaches convergence faster than the blue



Figure 4.3: Snapshot of Overview Layer during Case Study 1 with Chart View opened: Global chart (a). Community-wise chart view while Community 2 is selected (b).

one. It also reaches a higher global coverage of 47.71% rather than the 41.83% of the blue one (see fig. 4.2).

We refer to the author who serves as an entry point into Community 2 as the target author. To understand how the seed author can assert influence over our target author, we can trace back the diffusion process. Starting from Community 2, we identify each author who has played a role in that process (**T1**). In preparation for that, we first hide the blue simulation using the *Simulation Control Panel* (e.g. see fig. 4.2(f)), as for now, we only focus on the green one. Then, using the appropriate simulation card, we highlight the seed node (author *Munzner T.*) using the seed node list (e.g. fig. 4.2(d)). This marks the community that contains the seed author with a distinctive pink border (see fig. 4.2, Community 3).

Now that we can visually distinguish the seed community, let's focus our attention on Community 2: We hover over Community 2 and realize that it connects only to Community 8 (see fig. 4.1(a)). Therefore, we can conclude that an author from that community must have direct influence over Community 2. Hovering over Community 8 reveals all their inter-community *Edge Bundles* and their respective *Diffusion Flux Circles*. We immediately notice that there is only one such circle that is traveling towards Community 8. The source of this circle is Community 3, which happens to be the community that contains the seed author (see fig. 4.1(a)). Thus, we have successfully identified all the relevant communities. We select all three communities (3, 8, 2) and progress onto the *Local Layer*.

As seen in Figure 4.1(b), we can view the topological characteristics of our selected communities (T4). By default, the inter-community edges are shown and indicate the



Figure 4.4: Snapshot of Local Layer during Case Study 1: seed author Miksch S. (a) and target author Zhang X. (b) are highlighted via the pink halo. Both simulations are toggled on (c), and the current time step is set to the last one. We can see that Community 2 only contains a green simulation contour, and no blue contour (d). We can see which areas are superimposed by only one simulation.

connection points between the three communities. We pinpoint the target author by first setting the active time step to 0 and traversing through the time domain in increments of 1 (T3). At time step 5, we see the first simulation contour appear inside Community 2 (see fig. 4.1(b)). We click on it to briefly visit the *Detail Layer*. We do this to open the node profile window of the target author and highlight it, such that we can track the target author across all layers. The result of this can be seen in Figure 4.5(c). We go back to the *Local Layer* and can now clearly identify the seed and target author (see fig. 4.1(b)). To showcase the superimposed comparison of our two simulations' contours, we show the blue simulation again (see fig. 4.4(c)). In Figure 4.4, we navigate to the last step and investigate the activated areas within. In Community 3, which contains the seed author of the green simulation, we can see that both the green and blue simulations activate broad areas within the community (a). The same can be said for Community 8 (b). However, when it comes to Community 2, we notice an absence of blue simulation contours (d) (T2). To continue with our goal for this case study, we hide the blue simulation again and investigate the exact cascade between the seed and the target author. We open the selection circle tool and adjust the location and radius such that it includes not only our highlighted author, but also all authors that might have played a role in reaching the target author (see fig. 4.1(b)) (**T1**).

The Detail Layer initially renders the ForceAtlas2 layout [JVHB14] as seen in Figure 4.6.



Figure 4.5: Snapshot of Community 2 in detail Layer during Case Study 1: ForceAtlas2 [JVHB14] layout selected. Clicking on the "Layout" selection field opens a dropdown menu that contains the 2 alternative layouts: d3-force [BOH11] and Sugiyama [STT81] (a). A font slider scales the node labels (b). Zhang X. is highlighted. Within the green simulation, it influences 2 other authors in this Community: Gou L. and Hu Y. (c)

We immediately switch to the Sugiyama layout [STT81] as seen in Figure 4.1(c), as it supports a more efficient investigation of distinct cascades [AP16]. As seen in Figure 4.6, starting from the target author (*Zhang X.*) (d), we open the corresponding *node profile* (b) and highlight the author that has influenced *Zhang X.* (e) by clicking on the *infectedBy* field. We repeat this for every new source author that gets highlighted until we reach the seed author (c), thus revealing the four authors over whom *Munzner T.* has influenced *Zhang X.* (T1). At this point, we show the blue simulation again and see that the seed author of the green simulation is also activated in the blue one by author *Hoffman E.* (a) (T2). We refer to Figure 3.8 and assess the role of our newly identified key authors concerning the blue simulation, which could not make its way to Community 2. It manages to influence an author that has a direct connection to Community 2, however, this connection fails to activate (T2) (see fig. 3.8).

Case Study 2: Facebook Network of Company Pages. For the second case study,



Figure 4.6: Snapshot of Detail Layer during Case Study 1: All influenced authors from community 2,3, and 8 are visible. Dashed edges indicate an inter-community edge. Node profile of Miksch S. (a) and Zhang X. (b). All authors along the path from Miksch S. (c) and Zhang X. (d) are highlighted. Author (e) activates author (d)

we showcase the system's capabilities to handle larger networks and choose the Facebook network. The simulations for this dataset all use the SIR model [LWGZ17] with an infection rate β of 0.01 and a recovery rate γ of 0.005. We can interpret this as a piece of information or news spreading across the network of Facebook pages. Since an edge between two pages means encoding a mutual like, we can assume that this piece of information can spread from one page to the other. In accordance with the SIR model, a page is considered *Infected* if it discusses or interacts with this piece of information or news, and *Removed* if it no longer engages with it.

The seed selection technique is as follows: All three simulations use a seed set of 10% of the total node count. As seen in Figure 4.7(a), each employs a different selection metric. One uses randomly selected seed nodes (blue), the other two take the nodes



Figure 4.7: Snapshot of Overview Layer during Case Study 2:

with the highest closeness centrality (magenta) and degree (green), respectively. Our goal is to investigate the differences in reach among these seed selection strategies and identify communities that exhibit such discrepancies. This relates to the problem of IM. Accordingly, we start our investigation by assessing the different seed selection techniques in terms of their reach.

The simulation cards (see fig. 4.7(a)) reveal (**T4**) that the random selection strategy, with 92.98% reaches the most pages globally. It is followed closely by the strategy using closeness centrality (92.43%). Choosing the top 10% nodes with the highest degree gives us the smallest reach, with 91.41%. It is certainly interesting that the naive random seed selection method results in the highest coverage (**T2**).

To find a possible explanation, we can interact with the *Overview Layer*. Firstly, it seems that the network is tightly connected across the board, with larger, more central communities in the middle and smaller ones at the peripheries (**T4**) (see fig. 4.7(b)). The simulation bars inside the *Community Tiles* give us an idea of how engaging a piece of information is inside each community (see fig. 4.7(c)). The reach of all seed strategies across the larger communities tends to be more balanced than in the smaller ones (**T2,T4**). To further investigate our initial observation, we select all smaller communities tends to communities tends to be more balanced than in the smaller ones (**T1**) (see fig. 4.7(b)). The community-wise stats view (see fig. 4.7(d)) aggregates the convergence of each community and displays them for each simulation (**T4**). The gap



Figure 4.8: Snapshot of Overview Layer during Case Study 2 at time step 0: All the larger communities are selected (a) with community-wise simulation cards open (b).

between the random technique and the other metric-based techniques is evident compared to the global situation: While the random seed set selection reaches 91.19% coverage, the others do not even reach 89% (T2).

The seed selection method that only considers the most connected nodes seems to perform worse in reaching the smaller communities in the peripheries (see fig. 4.7(b), e.g., Communities 15,21,26,28). This is no surprise, since the best-connected nodes tend to be inside the larger communities. If we switch to the community-wise chart view (see fig. 4.10b), we can see that the closeness centrality technique barely has any seed nodes located in the selected set of small communities (**T1**). While the highest degree strategy employs 71 seed nodes in the selected region, the random selection almost doubles it with 136 nodes (**T2**). Though less in quantity, the metric-based strategies, due to their tighter connections, remain somewhat competitive in this selected subset of communities (**T3**).

Now, let us look at the larger communities. We deselect all currently selected and proceed to select only the larger communities as seen in Figure 4.8(a) and navigate to time step 0. We can see that among those larger communities, the seed-selection strategy going for the top 10% of nodes with the highest closeness centrality has marginally the highest coverage (see fig. 4.8(b)). Furthermore, in Figure 4.9(b) we can see that among those larger communities, the metric-based seed selection techniques (degree and closeness centrality) have more activated nodes at time step 0, thus more of these better-connected nodes are inside the larger communities. This is in line with our findings when we had only the smaller communities selected.



Figure 4.9: Snapshot of Overview Layer during Case Study 2 at time step 0: All the larger communities are selected (a) with the community-wise chart view open (b).

4.3 ICE-T Expert Interview

Many evaluation methods rely on low-level questions that fail to capture the broader significance of a visualization. To address this limitation, we adopt a value-driven assessment for evaluating *Thovea*, following the approach by Stasko [Sta14]. He proposes that the value of a system can be described via four different capabilities:

- Time. Captures whether the visualization can facilitate faster and more efficient searching of a particular information or casual scanning of the dataset.
- Insight. Refers to a visualization's capability to enable the extraction of intentional and incidental insights
- Essence. Captures the ability of a visualization to provide the essence of the dataset, in terms of overview and surrounding context.
- Confidence. Refers to the confidence the user has in the quality of the data and visualization. In particular, it describes the ability of the visualization to convey missing or false data and accurately represent the data, respectively.

Building on this framework, Wall et al. [WAM⁺19] introduce the ICE-T methodology, which enables a structured quantitative assessment of these four components. Each component is divided into three guidelines, further divided into a few low-level heuristics,



Figure 4.10: Community-wise chart view of three simulations without hovering over (a), and with hovering over such that a tooltip appears (b).All the smaller communities are selected within case study 2.

formulated as rateable statements. Study participants evaluate the system by rating these on a scale from 1 (strongly disagree) to 7 (strongly agree). In total, each participant assesses the system by rating the 21 heuristics. To consider our system *valuable*, we set a target global cumulative average score of 5 or higher, ensuring that *Thovea* meets a high standard of effectiveness and usability. The ICE-T questionnaires and all evaluation-related materials are included in our supplementary material.

Participants. As required by the ICE-T methodology, we recruited through personal contact five participants who are experts in the field of Information Visualization and VA. One has a strong focus on dynamic network analysis. Another has contributions in the field of network and multivariate data visualization. Two of our participants' research focuses on temporal data. Furthermore, one of our participants has experience with information visualization in an industry context. The participants comprised an assistant professor, a post-doctoral researcher, a PhD student, a senior researcher, and the head of a visual computing research group. In terms of demographics, 3 of our participants were male and two were female. None of the participants were involved in the design process of *Thovea* and had no knowledge of the system prior to the interview. All were required to sign a consent form compliant with the European General Data Protection Regulation before the interview (GDPR) and they did not receive compensation for their participation. All of them were familiar with the concept of ID, with some dealing with ID in their daily line of work.

Protocol. The ICE-T study was set up as a set of individual expert interviews. Since most of our participants could not be present physically, we conducted an online interview

with each participant individually via Microsoft Teams. The 1-hour expert interviews were structured as follows:

- Introduction (5'): At the start of each session, an introductory presentation highlighted the purpose behind the system. Here, we also indicated that the participants could at any point stop the experiment if they deemed so.
- **Demo** (10'): A live demo of *Thovea* showcased all its features and familiarized the participant with its usage.
- Free Interaction (30'): The participants were asked to explore the system on their own and vocalize their actions and thoughts following the think-aloud method. Each participant was offered a set of tasks for their example dataset of choice (among the four described above). However, it was made clear that solving those tasks was not mandatory.
- Feedback (15'): We asked each participant for some post-evaluation feedback to supplement their quantitative scoring. The questions were as follows: (i) The system was designed with four tasks in mind (Overview, Explore, Navigate, Compare). Do you agree with our design decisions regarding those tasks? And which ones do you not agree with?; (ii) What did you like?; (iii) What features did you miss when interacting with the system?

After the interview, each participant was provided with the ICE-T Questionnaire. They were instructed to fill it out at their convenience. Since the system was hosted as a web application, the participants could revisit the system while filling out the questionnaire. The evaluation was considered complete once we received the ICE-T questionnaire.

Results. The average score for each component across all participants is presented in Table 4.1. We follow a standard score aggregation procedure outlined by Wall et al. $[WAM^{+}19]$, where we compute the mean scores at each hierarchical level. This approach ensures a balanced and structured evaluation of the system's effectiveness. Those performs best in the Insight and Essence components, both receiving an average score of 6.2. Achieving a high score in Insight indicates that our approach not only provides an enhanced understanding of the data but also offers alternative perspectives to investigate the data and stumble upon new insights [WAM⁺19], as one participant stated: "For the exploration task, this design is very nice, because it also gives an overview with the clustering and the communities. Also, you can go one level deeper and see every single node. The Exploration task is very nice across all the layers.". This is a crucial strength that enables us to successfully support T1 by facilitating deep analytical exploration. A high score in Essence aligns with our goal of providing an overview of the network and ID according to T4. One participant stated: "The (...) overview I think is pretty good, because it's very easy to detect isolated clusters and connected ones, and (...) different groups in the graph. So i think this clustering is working very nicely.". This confirms that our visualization effectively distills complex data into meaningful

Parameter	Average Score	Std. Dev. (σ)
Insight	6.2	0.47
Confidence	5.23	0.45
Essence	6.2	0.45
Time	5.66	0.7

Table 4.1: Results of the ICE-T Evaluation. Scale from 1 (worst) to 7 (best)

representations. Following these, we observe slightly lower but still strong scores in Time (5.6) and Confidence (5.23). The score in Time indicates that *Thovea* enables users to efficiently extract key information without too many delays. The Confidence component, however, registers the lowest score among the four, with a particular challenge identified in the heuristic: "The visualization helps understand data quality" [WAM⁺19], which received an average score of 3.4. However, this is not surprising since the requirements for our design did not include verification of data quality. Overall, across all components, we achieve a global average score of 5.82. Individually, each component receives a score greater than 5. This surpasses our predefined success criterion, confirming that *Thovea* effectively supports its intended tasks and delivers a high-value visualization to users.



CHAPTER 5

Discussion

In this section, we discuss our lessons learned, especially regarding the feedback from the ICE-T [WAM⁺19]. Furthermore, we identify the main limitations of the system and point to directions for future work. Finally, we revisit our research questions from Section 1.1 and assess whether *Thovea* manages to answer them.

Mental Map Preservation. One of the more notable challenges of having a threelayered visualization method is that it requires the user to switch between multiple different modes of visualization. This can lead to initial confusion as elements on the screen are rearranged, vanish, or appear when traversing the layers. One participant noted that when switching from the node-link based Overview Layer to the density-based Local Layer, they had difficulties understanding the meaning of this transition. Here, an animated transition could guide the user through the transition by, for example, zooming into the selected communities [BB99]. This could make it clearer that we are traveling into the selected communities. In the *Detail Layer*, one participant suggested that an animated transition could also be leveraged for mental map preservation when switching between the available layouts. They also extend this possibility to the navigation between time steps. Especially inside the Sugiyama layout, some form of highlighting or movement (e.g., [WB04, WB05, VAA24]) could be useful (T3) due to the dynamic nature of ID [BB99]. Beyond animated transitions, additional contextual cues (e.g., subtle fading effects for disappearing elements, breadcrumbs views, or ghosting of previous states) could further reinforce continuity and help users maintain their mental map. Furthermore, there was a desire for a more consistent approach in community placement across the first two layers. The spatial relationship of Community Tiles in the Overview Layer is not preserved when switching to the *Local Layer*, but rather is recalculated. Taking this spatial information into account when drawing the respective communities could reduce confusion. A limitation of the current implementation that relates to that issue is overlapping communities within the *Local Layer*. This ambiguity led to some participants misinterpreting this as nodes within the borders of multiple communities being also

assigned to all of them. Additionally, some users might prefer faster or slower animations. Allowing them to adjust transition speed could improve accessibility and user experience Overall, the participants had positive remarks on the layered top-down approach to exploration as it allows them to focus on the region they deem interesting (T4), especially once they have familiarized themselves with the interface.

Guidance. During the investigation of the EU Power Grid dataset, one participant had difficulties in locating the source of a secondary bump in activation within a simulation. Depending on the selected resolution, that dataset can result in dozens of communities. Thus, trying to identify a growing simulation bar inside a single *Community Tile* can be challenging. This relates to the *change blindness* [NHT01] problem, where smaller changes within a larger picture can remain unnoticed. To solve this issue, future work could consider incorporating guidance mechanisms [CGM⁺17] that automatically highlight potential areas of interest and nudge the user onto communities that might contain useful insights. Such a mechanism would also help jump-start the exploration of the data, as initially, some participants were unsure where to start **(T1)**. Future work could investigate how and by what metric communities that exhibit unusual or dominant behavior within a simulation could be highlighted. Two participants also voiced their desire for some more information about the communities. Furthermore, integrating *onboarding* could help the users in navigating complex application dashboards [DWH⁺22].

Role of Local Layer. During the evaluation, questions about the purpose and role of the *Local Layer* have come up. This was especially the case for participants who predominantly explored the smaller InfoVis dataset. The *Local Layer* was designed to provide a high-level understanding of the local topology of communities of interest and the ID process. For larger networks, revealing detailed information about individual nodes and low-level diffusion pathways in this layer would not be feasible due to the large amount of visual clutter. In some cases, participants were confronted with small density maps that were based on only a handful of nodes. To bridge the gap between these differences in magnitudes, we provide customizability via the optional visualization of bundled edges and individual nodes (**T1**).

Ambiguities with Geospatial Data. The EU Power Grid dataset contains geographical information in the form of country codes. However, *Thovea* as a domain-agnostic platform treats this information as regular metadata that is displayed in the *Detail Layer* but not incorporated into the layout calculation. Thus, the partition into communities does not consider this geographical relation. This can lead to confusion, as users might expect communities to be separated by such geographical attributes. Future work could resolve this confusion and leverage such geospatial attributes (if provided) to calculate a more intuitive partition (e.g. [STP⁺17]).

Visual Scalability. Overall, the three-layer approach, combined with the opportunity to customize the visualizations, manages to reduce visual clutter such that a meaningful exploration and analysis can take place. However, there are limitations, especially when it comes to the comparison of multiple simulations (T2). In the *Local Layer*, we have

chosen a superimposed approach for this comparison. The main problem is that rendering the contours of three, four, or in some cases even two simulations on top of each other led to a significant amount of visual clutter that is typical for superimposed approaches [GAW⁺11]. Participants worked around this problem by showing and hiding specific simulations. A similar problem can be examined in the Sugiyama layout. The participant had positive impressions of this layout, when it was run for a single simulation. Once additional simulations are considered, it becomes increasingly difficult to perform a holistic assessment of the cascade. The implementation calculates a Sugiyama layout for all the nodes that are activated across the simulations. This mixes up the pathways between the simulations and can, for example, lead to edges that point upwards. Given these limitations, an alternative approach that leverages some form of difference visualization or small multiples might be worth exploring [KCK17].

Computational Scalability. The initial reduction of computational cost by partitioning the graph into smaller sub-graphs enables a swift experience for datasets that contain thousands of nodes. To test out the limit of *Thovea*, we use a social network dataset extracted from the website Slashdot [LHK10]. It contains 77350 nodes and 516575 edges. We ran the application locally on a laptop with an AMD Ryzen 7 5800U CPU and 16GB of RAM. Applying the Louvain algorithm [BGLL08] and the subsequent precalculation of the individual community layout using the ForceAtlas2 algorithm [JVHB14] took 46 seconds. The subsequent loading of the precalculated simulation file took 26 seconds. Thus, the simulation was ready to be explored after 72 seconds. Interactivity was smooth for the *Overview Layer* and *Detail Layer*. Inside the *Local Layer*, with a community open that contained 20623 nodes, we had to turn off the nodes for a smooth experience. The calculation of the Edge Path Bundling [WAA⁺23] for the inter- and intra-community edges took too much time to be considered usable for such large communities. Here, one could increase the resolution of the Louvain algorithm to further divide such large communities into smaller ones.

Diffusion Flux Circle. The choice of using animation was perceived positively by our participants, as it provided a quick glimpse into the flow of diffusion among neighboring communities **(T4)**. However, the main difficulty with this encoding lies in accurately assessing the speed of the circles to infer the activation rate of the respective *Edge Bundle*. Future work could further explore the concept of *unfolding edges* [BDT23] to design a more suitable representation for the application domain of ID.

Data Quality. In the ICE-T evaluation [WAM⁺19], we got our lowest score in *Confidence*. This is not surprising since *Thovea* is not designed to consider data quality. Thus, the results of the evaluation need to be assessed with this context in mind.

5.1 Revisiting Research Questions

In the final section of this chapter, we revisit our research questions that have been stated in Section 1.1. We will start with the sub-questions Q2, Q3, and Q4 and then assess whether the overarching main question Q1 has been given a meaningful answer. **Q2** Which visualization technique or combination of techniques is suitable for exploring dynamic diffusion processes over large networks?

Looking back at the state-of-the-art in Section 2.3, many approaches are designed to solve problems specific to a certain application domain. They introduce novel visualization techniques and model using, for example, map [CCW⁺19, CLCY20, CCL⁺17] and river metaphors [SWL⁺14, WLY⁺14] that are tailored to their specific application domain. Many of those are found within the domain Social Networks. When it comes to the field of epidemiology, the use of geographical maps was very common (see table 2.2) [AME11, BGG⁺11, YDH⁺17, MLR⁺11]. Overall, Table 2.2 reveals that there is a great variety in employed visualization methods in the *Diffusion Exploration* category, where the data itself encodes diffusion behaviour. When it comes to simulating an ID process, such variety can no longer be found. Apart from the geographical maps used in an epidemiological context, node-link visualizations are a popular choice. Some of the approaches do support larger datasets such as R-Map [CLCY20], DMap+ [CCW⁺19], and Opinionflow [WLY⁺14]. Even though they are successful in revealing insights hidden in complex data sets, they are not easily exportable to different application domains. The number of scalable VA platforms that provide a domain-agnostic approach is limited. Namely, DiVA $[SGM^+23]$ and VAIM $[ADL^+22]$ approach this issue. The main problem with DiVA, however, is that it solely relies on node-link visualization. Computationally speaking, DiVA is successful in providing a platform that can load and offer interactivity for larger networks; however, visual scalability is an area it struggles. Even though nodelink visualizations are popular (see table 2.2), they do cause clutter in larger instances. VAIM's [ADL⁺22] focus+context approach introduces Density and Diffusion Matrices that are more abstract and can be used to explore certain areas of interest through a local node-link visualization. These matrices serve as an overview of not only the topology of the network but also the diffusion process. Similarly, in our ICE-T evaluation, all participants appreciated the overview visualization that *Thovea* offers. It serves as a starting point for further exploration and can help form hypotheses about the ID process or network itself. The node-link visualization of the Overview layer and Detail Layer was easily interpretable and understandable during our study, especially since we used aggregation and filtering, respectively. Beyond new techniques for large graph drawings, aggregation and filtering are identified as a popular route [vLKS⁺11] for working with large networks. Therefore, it is not necessarily a specific visualization method that excels in conveying complex patterns, but rather the way the data is processed such not to overwhelm the cognitive load of the user $[YAD^+18]$.

Q3 Does a layered visualization approach that uses different levels of abstraction help break down the complexity of analyzing ID over large networks?

Looking back at our ICE-T evaluation, all participants appreciated the three-layer visualization approach of *Thovea* as it combines a high-level overview with an ondemand exploration of distinct pathways. The global average score of 5.82 on the ICE-T questionnaire also reinforces that *Thovea*'s layered approach is successful in breaking down complex datasets into smaller and simpler problems. Especially when considering

56
a score of 6.2 in both *Essence* and *Insight*. Not only does *Thovea* succeed in providing a meaningful overview of the network and the ID process, it also enables the user to reveal insight that otherwise would remain hidden in the complexity of the data [WAM⁺19], by its ability to filter down to a handful of interesting nodes. However, there is still room for improvement, as discussed above in this chapter.

Q4 How can two or more diffusion models be meaningfully compared in a single view such that local node-level and broader network-wide differences can be intuitively spotted and assessed?

Those a integrates the comparison of multiple ID processes into each layer's visualization technique. We have enabled the comparison of up to four concurrent simulations. While expressing the desire for some additional features in the *Overview Layer*, especially concerning guidance, the participants were positive about the visual encodings for the comparison feature. The simulation bar chart inside each *Community Tile* was intuitive and easy to read. However, the superimposed simulation contours were prone to visual clutter. Even though we support the comparison of up to four concurrent simulations, our participants primarily avoided displaying more than two simulation contours. The visual encodings for comparison in the *Detail Layer* were also well received by the participants. However, comparing two or more simulations on the Sugiyama layout was very difficult. In those cases, the participants predominantly made use of either the ForceAtlas2 [JVHB14] or the d3.forceSimulation [BOH11] layout.

Q1 How can VA be leveraged to explore and analyze diffusion processes in large networks to uncover useful, otherwise hidden insights?

With *Thovea*, we have shown that employing VA to solve problems in the field of ID can yield great benefits. The two case studies (see section 4.2) and the ICE-T study (see section 4.3) showcase and confirm their uses and advantages. By splitting the complexity of dynamic diffusion datasets (see section 3.1) into multiple layers with different abstraction levels, we enable a top-down exploration that enables a comprehensive understanding of the underlying diffusion process.



CHAPTER 6

Summary

Within this thesis, we introduce *Thovea*, a three-layer VA platform that enables the analysis, exploration, and comparison of ID processes over large networks. ID is a process that is related to a variety of different application domain. From pathogen transmission in epidemiology, spreading of fake news inside a social network, to the distribution of malware inside critical infrastructure. Leveraging VA to gain a deeper understanding on these processes can reveal otherwise hidden insight that can than be used to either maximize, minimize or trace back the diffusion process. To get a comprehensive overview of the existing contributions in this field we first review the literature in a state-of-the-art on the visualization of ID. We introduce a taxonomy that brings structure into this field and generalizes across all application domains by categorizing the limited existing contributions by their main goal: Diffusion Exploration, Spread Simulation, Influence Maximization, Influence Summarization, Model Comparison, and Anomaly Detection. We further highlight certain key features that are found within the respective contributions: Key-Player Detection, Comparison Tools, Model/Parameter Tweaking to understand which features are crucial for the different main goals. We identify the lack of domainagnostic platforms that generalize over different application domains. Furthermore, we acknowledge visual scalability as a main issue in dealing with larger ID processes, especially if we want to compare multiple ID simulations.

With these shortcomings identified, we design and implement *Thovea*. We want to provide a domain-agnostic platform for scalable ID. As we followed Munzner's [Mun09] nested model for visualization design and validation, we extracted a set of four abstract tasks that are commonly found within the contributions listed in our taxonomy in Table 2.2. Based on these tasks, we design and implement visual encodings. The core novelty of *Thovea* is found within the layered approach to ID. The data can be separated into the "medium" network, which serves as the underlying base network, and the "diffusion" dynamic graph, where each of its static components represents the ID process at a specific time step. The complexity is amplified when we consider large networks. By splitting this complex data into three distinct layers that each provide a unique perspective on the data, we reduce the cognitive load on the user and enable a stepwise top-down exploration of the underlying diffusion process. The first layer offers a high-level aggregated overview of the network, highlighting inter- and intra-community diffusion trends. The second layer provides a density-based estimation and visualization of selected communities, detailing intra- and inter-community diffusion progression. The last layer enables in-depth investigation and analysis of specific regions within communities, allowing users to explore individual diffusion paths down to the node level.

The evaluation is done twofold: First, we conduct two case studies that showcase the utility and features of our system. Furthermore, an ICE-T [WAM⁺19] study confirms the value of *Thovea* with a global average score of 5.82, where each individual component has been awarded a score greater than 5. We especially notice the high scores in the components *Essence* and *Insight* with a score of 6.2, underlining *Thovea*'s ability to provide a meaningful overview of the network and underlying ID process, and its ability to help users uncover hidden insights.

Regarding lessons learned we refer to the difficulty in comparing multiple diffusion processes in a single view. It is very easy to overload the cognitive load as the comparative burden is left with the user. Furthermore, static transitions between layers and time steps leave the user alone in deciphering changes, as stated by some of our participants.

For future work, we point in the direction of researching methodologies to further preserve the mental map of the user while travelling across layers and time steps. Introducing animated transitions could lead to improvements. Furthermore, we highlight the need to develop a methodology that specializes in the comparison of ID processes to further reduce clutter and improve visual scalability.

60

Overview of Generative AI Tools Used



List of Figures

$2.1 \\ 2.2$	Map metaphors used by D -map+[CCW ⁺ 19] (a) and R -Map[CLCY20] (b). River metaphors for diffusion exploration used by $EvoRiver[SWL^+14]$ (a) and $OpinionFlow[WLY^+14]$ (b).	11 12
2.3	Node-link diagram for diffusion exploration by <i>Whisper</i> [NYX ⁺ 12] (a) and <i>SocialWave</i> [STP ⁺ 17] (b)	13
2.4	Interface of BloomWind [YJZ ⁺ 25]	14
2.5	Diffusion exploration via matrix view [HNWC23] (a) and storylines [BPW ⁺ 21]	
	(b)	14
2.6	Spread simulation with the system of von Landesberger et al. [VLDBF15].	15
2.7	Interface of $DiVA$ [SGM ⁺ 23] in dual diffusion visualization mode	16
2.8	PanViz interface [MLR ⁺ 11] (a), Decision history tree [AME11] (b)	17
2.9	GLEaMviz [BGG ⁺ 11] map interface (a), and model builder (b)	18
2.10	VAIM [ADL ⁺ 22] interface (a), and <i>Eiffel</i> [HSS ⁺ 20] interface (b)	18
2.11	Model Comparison [VKPM15] (a), and $\#FluxFlow$ [ZCW ⁺ 14] (b)	19
3.1	Munzner's [Mun09] nested model for design and evaluation \dots Flowebart of <i>Theway's</i> layers: <i>Overview Layer</i> (1) <i>Local Layer</i> (2) and <i>Datail</i>	23
3.2	Layer (3). One Node is highlighted throughout the layers	27
3.3	Thovea interface: (A) Setup view containing drawing parameters, node search & highlight, simulation control panel, and loading tools; (B) Community-wise stats and metrics; (C) Global stats and metrics; (D) Primary visualization view—currently showing the <i>Overview Layer</i> for three simulations, (E) Panel for austomization of primary visualization including the Time Control Panel	20
3.4	The <i>Local Layer</i> with two communities selected (indicated within the dashed blocks) and nodes and edges toggled on. The dominant <i>Edge Bundle</i> between the communities is shown with the thick, connecting line. Two simulations	20
	(blue vs. green) are loaded, shown at time step 0	30
3.5	Local Layer representation of one community: (a) without and (b) with	
	additional context through nodes and edges.	31
3.6	Independent Edge Ambiguity as described by Wallinger et al. [WAA ⁺ 23]	32
3.7	Detail Layer of a co-authorship network using a ForceAtlas2 layout, with the	
	node profiles of Munzner T. (c) and Zhang X. (b) expanded and highlighted.	0.0
	A minimized view of the <i>Local Layer</i> (a) is placed at the top left. \ldots .	33

34

Snapshot of Thovea's Three-Layer pipeline for Case Study 1: InfoVis Co-4.1authorship Network. (a) Overview Layer of two different ID simulations (blue and green), where three communities (2, 3, 8) have been selected; (b) Local Layer with a density-based visualization of the selected communities just for the green simulation, and two nodes highlighted (in magenta); (c) The selected circular area from (b) is viewed through a Sugiyama layout, enabling 40the investigation of detailed activation pathways between the selected nodes. Snapshot of Overview Layer during Case Study 1 with Metrics View opened: 4.2Panel for graph selection and upload (a). Panel for simulation selection or upload (b). Community 2 (contains the target author) is selected (c). Simulation card of blue and green simulation. The green simulation card's seed node list is expanded (d). Community-wise simulation cards of green and blue simulation (e). Simulation Control Panel with both simulations visible(f) 41Snapshot of Overview Layer during Case Study 1 with Chart View opened: 4.3Global chart (a). Community-wise chart view while Community 2 is selected 42Snapshot of Local Laver during Case Study 1: seed author Miksch S. (a) and 4.4target author Zhang X. (b) are highlighted via the pink halo. Both simulations are toggled on (c), and the current time step is set to the last one. We can see that Community 2 only contains a green simulation contour, and no blue contour (d). We can see which areas are superimposed by only one simulation. 43Snapshot of Community 2 in detail Layer during Case Study 1: ForceAtlas2 4.5[JVHB14] layout selected. Clicking on the "Layout" selection field opens a dropdown menu that contains the 2 alternative layouts: d3-force [BOH11] and Sugiyama [STT81] (a). A font slider scales the node labels (b). Zhang X. is highlighted. Within the green simulation, it influences 2 other authors in this Community: Gou L. and Hu Y. (c) 44 Snapshot of Detail Laver during Case Study 1: All influenced authors from 4.6community 2,3, and 8 are visible. Dashed edges indicate an inter-community edge. Node profile of Miksch S. (a) and Zhang X. (b). All authors along the path from Miksch S. (c) and Zhang X. (d) are highlighted. Author (e) 454.7464.8Snapshot of Overview Layer during Case Study 2 at time step 0: All the larger communities are selected (a) with community-wise simulation cards 47Snapshot of Overview Layer during Case Study 2 at time step 0: All the 4.9larger communities are selected (a) with the community-wise chart view open 48

4.10	Community-wise chart view of three simulations without hovering over (a), and with hovering over such that a tooltip appears (b). All the smaller communities	
	are selected within case study 2	49
1	Filled-out questionnaire from Participant 1	82
2	Filled-out questionnaire from Participant 2	83
3	Filled-out questionnaire from Participant 3	84
4	Filled-out questionnaire from Participant 4	85
5	Filled-out questionnaire from Participant 5	86
6	Average score per ICE-T [WAM ⁺ 19] component per participant (P1-5) with	
	Total scores and standard deviation	86



List of Tables

2.1	Selection of keywords split into two sets	7
2.2	Taxonomy of all 28 contributions with three main classifications: visualization	
	technique, goal, and utility features	10
4.1	Results of the ICE-T Evaluation. Scale from 1 (worst) to 7 (best) \ldots	51



Acronyms

DAG Directed Acyclic Graph. 33

- **IC** Independent Cascade. 1, 5, 19, 24, 40
- ID Information Diffusion. 1–3, 5, 6, 8, 13, 14, 16, 22–25, 27, 31–35, 39–41, 49, 50, 53–57, 59, 60, 64
- **IM** Influence Maximization. 1, 5, 6, 17, 18, 21, 24, 46

LoD Level-of-Detail. 25

LT Linear Threshold. 1, 5, 19, 24, 81

SIR Susceptible-Infected-Removed. 5, 7, 24, 45, 81, 82

SIRS Susceptible-Infected-Removed-Susceptible. 5, 7

SIS Susceptible-Infected-Susceptible. 5, 7

VA Visual Analytics. 1, 2, 6–9, 11, 14–20, 22, 49, 56, 57, 59



Bibliography

- [ADL⁺22] Alessio Arleo, Walter Didimo, Giuseppe Liotta, Silvia Miksch, and Fabrizio Montecchiani. Influence Maximization With Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics*, 28(10):3428–3440, October 2022.
- [AGB⁺21] Zahra Aghaee, Mohammad Mahdi Ghasemi, Hamid Ahmadi Beni, Asgarali Bouyer, and Afsaneh Fatemi. A survey on meta-heuristic algorithms for the influence maximization problem in the social networks. *Computing*, 103(11):2437–2477, November 2021.
- [AME11] Shehzad Afzal, Ross Maciejewski, and David S. Ebert. Visual analytics decision support environment for epidemic modeling and response evaluation. In 2011 IEEE Conference on Visual Analytics Science and Technology (VAST), pages 191–200, Providence, RI, USA, October 2011. IEEE.
- [and17] Yen-Chi Chen and. A tutorial on kernel density estimation and recent advances. *Biostatistics & Epidemiology*, 1(1):161–187, 2017.
- [AP16] Daniel Archambault and Helen C. Purchase. On the effective visualisation of dynamic attribute cascades. *Information Visualization*, 15(1):51–63, 2016.
- [BAEA11] Ceren Budak, Divyakant Agrawal, and Amr El Abbadi. Limiting the spread of misinformation in social networks. In *Proceedings of the 20th International Conference on World Wide Web*, WWW '11, page 665–674, New York, NY, USA, 2011. Association for Computing Machinery.
- [BB99] B.B. Bederson and A. Boltman. Does animation help users build mental maps of spatial information? In *Proceedings 1999 IEEE Symposium on Information Visualization (InfoVis'99)*, pages 28–35, 1999.
- [BDT23] M.-J. Bludau, M. Dörk, and C. Tominski. Unfolding edges: Adding context to edges in multivariate graph visualization. *Computer Graphics Forum*, 42(3):297–309, 2023.
- [BGG⁺11] Wouter Van Den Broeck, Corrado Gioannini, Bruno Gonçalves, Marco Quaggiotto, Vittoria Colizza, and Alessandro Vespignani. The GLEaMviz

computational tool, a publicly available software to explore realistic epidemic spreading scenarios at the global scale. *BMC Infectious Diseases*, 11(1):37, December 2011.

- [BGLL08] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, oct 2008.
- [BM08] J.A. Bondy and U.S.R Murty. *Graph Theory*. Springer Publishing Company, Incorporated, 1st edition, 2008.
- [BOH11] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D³ data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, 2011.
- [BPW⁺21] 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. In Search of Patient Zero: Visual Analytics of Pathogen Transmission Pathways in Hospitals. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):711–721, February 2021. arXiv:2008.09552 [cs].
- [Bra08] Fred Brauer. Compartmental Models in Epidemiology, pages 19–79. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.
- [Bri23] Erik Brinkman. d3-dag: Layout algorithms for directed acyclic graphs in d3.js, 2023. Version 1.1.0, accessed March 20, 2025.
- [CCL⁺17] Siming Chen, Shuai Chen, Lijing Lin, Xiaoru Yuan, Jie Liang, and Xiaolong Zhang. E-Map: A Visual Analytics Approach for Exploring Significant Event Evolutions in Social Media. In 2017 IEEE Conference on Visual Analytics Science and Technology (VAST), pages 36–47, Phoenix, AZ, October 2017. IEEE.
- [CCW⁺16] Siming Chen, Shuai Chen, Zhenhuang Wang, Jie Liang, Xiaoru Yuan, Nan Cao, and Yadong Wu. D-Map: Visual analysis of ego-centric information diffusion patterns in social media. In 2016 IEEE Conference on Visual Analytics Science and Technology (VAST), pages 41–50, Baltimore, MD, USA, October 2016. IEEE.
- [CCW⁺19] Siming Chen, Shuai Chen, Zhenhuang Wang, Jie Liang, Yadong Wu, and Xiaoru Yuan. D-Map+: Interactive Visual Analysis and Exploration of Ego-centric and Event-centric Information Diffusion Patterns in Social Media. ACM Transactions on Intelligent Systems and Technology, 10(1):1– 26, January 2019.

- [CGM⁺17] Davide Ceneda, Theresia Gschwandtner, Thorsten May, Silvia Miksch, Hans-Jörg Schulz, Marc Streit, and Christian Tominski. Characterizing guidance in visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):111–120, 2017.
- [CLCY20] Shuai Chen, Sihang Li, Siming Chen, and Xiaoru Yuan. R-Map: A Map Metaphor for Visualizing Information Reposting Process in Social Media. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):1204– 1214, January 2020.
- [CWW10] Wei Chen, Chi Wang, and Yajun Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. KDD '10, page 1029–1038, 2010.
- [DR01] Pedro Domingos and Matt Richardson. Mining the network value of customers. In *Proceedings of the Seventh ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining*, KDD '01, page 57–66, 2001.
- [DWH⁺22] Vaishali Dhanoa, Conny Walchshofer, Andreas Hinterreiter, Holger Stitz, Eduard Groeller, and Marc Streit. A process model for dashboard onboarding. *Computer Graphics Forum*, 41(3):501–513, 2022.
- [DXZ⁺14] Donghao Ren, Xin Zhang, Zhenhuang Wang, Jing Li, and Xiaoru Yuan. WeiboEvents: A Crowd Sourcing Weibo Visual Analytic System. In 2014 IEEE Pacific Visualization Symposium, pages 330–334, Yokohama, March 2014. IEEE.
- [GAW⁺11] Michael Gleicher, Danielle Albers, Rick Walker, Ilir Jusufi, Charles D. Hansen, and Jonathan C. Roberts. Visual comparison for information visualization. *Information Visualization*, 10(4):289–309, 2011.
- [GHFZ13] Adrien Guille, Hakim Hacid, Cecile Favre, and Djamel A Zighed. Information diffusion in online social networks. *SIGMOD Record*, 42(2), 2013.
- [HNWC23] Dongyun Han, Abdullah-Al-Raihan Nayeem, Jason Windett, and Isaac Cho. PDViz: A Visual Analytics Approach for State Policy Data. Computer Graphics Forum, 42(1):305–318, February 2023. Publisher: John Wiley & Sons, Ltd.
- [HSS08] Aric Hagberg, Pieter J Swart, and Daniel A Schult. Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Laboratory (LANL), Los Alamos, NM (United States), 2008.
- [HSS15] Steffen Hadlak, Heidrun Schumann, and Hans-Jörg Schulz. A Survey of Multi-faceted Graph Visualization. In *Eurographics Conference on Visualization (EuroVis) - STARs.* The Eurographics Association, 2015.

- [HSS⁺20] Yucheng Huang, Lei Shi, Yue Su, Yifan Hu, Hanghang Tong, Chaoli Wang, Tong Yang, Deyun Wang, and Shuo Liang. Eiffel: Evolutionary Flow Map for Influence Graph Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 26(10):2944–2960, October 2020.
- [JVHB14] Mathieu Jacomy, Tommaso Venturini, Sebastien Heymann, and Mathieu Bastian. Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. *PLOS ONE*, 9(6):1–12, 06 2014.
- [KCK17] Kyungyoon Kim, John V. Carlis, and Daniel F. Keefe. Comparison techniques utilized in spatial 3d and 4d data visualizations: A survey and future directions. Computers & Graphics, 67:138–147, 2017.
- [KKT03] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '03, page 137–146, 2003.
- [LFWT18a] Yuchen Li, Ju Fan, Yanhao Wang, and Kian-Lee Tan. Influence maximization on social graphs: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 30(10):1852–1872, 2018.
- [LFWT18b] Yuchen Li, Ju Fan, Yanhao Wang, and Kian-Lee Tan. Influence Maximization on Social Graphs: A Survey. *IEEE Transactions on Knowledge and Data Engineering*, 30(10):1852–1872, October 2018.
- [LHK10] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. Signed networks in social media. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10, page 1361–1370. Association for Computing Machinery, 2010.
- [LW14] Cheng Long and Raymond Chi-Wing Wong. Visual-VM: A Social Network Visualization Tool for Viral Marketing. In 2014 IEEE International Conference on Data Mining Workshop, pages 1223–1226, Shenzhen, China, December 2014. IEEE.
- [LWGZ17] Mei Li, Xiang Wang, Kai Gao, and Shanshan Zhang. A Survey on Information Diffusion in Online Social Networks: Models and Methods. *Information*, 8(4):118, September 2017.
- [LWY⁺20] Quan Li, Ziming Wu, Lingling Yi, Kristanto Sean N., Huamin Qu, and Xiaojuan Ma. WeSeer: Visual Analysis for Better Information Cascade Prediction of WeChat Articles. *IEEE Transactions on Visualization and Computer Graphics*, 26(2):1399–1412, February 2020.

- [Mar22] Antoine Marot. European power grid network with nodal prices. Zenodo, August 2022.
- [MBB⁺11] Adam Marcus, Michael S. Bernstein, Osama Badar, David R. Karger, Samuel Madden, and Robert C. Miller. Twitinfo: aggregating and visualizing microblogs for event exploration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 227–236, Vancouver BC Canada, May 2011. ACM.
- [MBO⁺22] Shynar Mussiraliyeva, Gulshat Baispay, Ruslan Ospanov, Zhanar Medetbek, and Kazybek Shalabayev. Graphical Visualization of the Connections of Involved Users and Identifying Influential Spreaders in a Social Network. In 2022 9th International Conference on Electrical and Electronics Engineering (ICEEE), pages 311–315, Alanya, Turkey, March 2022. IEEE.
- [mer25] mert96. Thovea github repository. https://github.com/mert96/ THOVEA-Information-Diffusion-Visual-Analytics, 2025. Accessed: Mar. 2025.
- [MLR⁺11] Ross Maciejewski, Philip Livengood, Stephen Rudolph, Timothy F. Collins, David S. Ebert, Robert T. Brigantic, Courtney D. Corley, George A. Muller, and Stephen W. Sanders. A pandemic influenza modeling and visualization tool. Journal of Visual Languages & Computing, 22(4):268–278, August 2011.
- [Mun09] Tamara Munzner. A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):921–928, 2009.
- [NBW14] Rudolf Netzel, Michel Burch, and Daniel Weiskopf. Comparative eye tracking study on node-link visualizations of trajectories. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2221–2230, 2014.
- [New06] M. E. J. Newman. Modularity and community structure in networks. Proceedings of the National Academy of Sciences, 103(23):8577–8582, 2006.
- [NHT01] L. Nowell, E. Hetzler, and T. Tanasse. Change blindness in information visualization: a case study. In *IEEE Symposium on Information Visualization*, 2001. INFOVIS 2001., pages 15–22, 2001.
- [Noa07] Andreas Noack. Energy models for graph clustering. Journal of graph algorithms and applications, 11(2):453–480, 2007.
- [NYX⁺12] Nan Cao, Yu-Ru Lin, Xiaohua Sun, D. Lazer, Shixia Liu, and Huamin Qu. Whisper: Tracing the Spatiotemporal Process of Information Diffusion in Real Time. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2649–2658, December 2012.

- [OK17] Yosuke Onoue and Koji Koyamada. Optimal tree reordering for group-in-abox graph layouts. In *Proc. SIGGRAPH Asia*, SA '17, 2017.
- [Pal24] Pallets. Flask [computer software]. PyPI, 2024. Version 3.0.3, Accessed: Feb. 2024.
- [PL20] Bernhard Preim and Kai Lawonn. A survey of visual analytics for public health. *Computer Graphics Forum*, 39(1):543–580, 2020.
- [Pli25] Guillaume Plique. Graphology, a robust and multipurpose graph object for javascript. Zenodo, February 2025.
- [RA15] Ryan A. Rossi and Nesreen K. Ahmed. The network data repository with interactive graph analytics and visualization. In *AAAI*, 2015.
- [Rea] React. React: A javascript library for building user interfaces. https: //reactjs.org/. Accessed: Feb. 2024.
- [RMR18] Giulio Rossetti, Letizia Milli, and Salvatore Rinzivillo. Ndlib: a python library to model and analyze diffusion processes over complex networks. In *Companion Proceedings of the The Web Conference 2018*, pages 183–186, 2018.
- [SCH⁺16] John Stasko, Jaegul Choo, Yi Han, Mengdie Hu, Hannah Pileggi, Ramik Sadana, and Charles D. Stolper. Citevis citation datafile. http://www.cc.gatech.edu/gvu/ii/citevis/ infoviscitation-data.txt, 2016. Accessed: May 2025.
- [SGM⁺23] Dhruv Sehnan, Vasu Goel, Sarah Masud, Chhavi Jain, Vikram Goyal, and Tanmoy Chakraborty. Diva: A scalable, interactive and customizable visual analytics platform for information diffusion on large networks. ACM Trans. Knowl. Discov. Data, 17(4), February 2023.
- [Shn03] Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *The Craft of Information Visualization*, pages 364–371. Morgan Kaufmann, San Francisco, 2003.
- [SRMV16] Konstantinos Skianis, Maria-Evgenia G. Rossi, Fragkiskos D. Malliaros, and Michalis Vazirgiannis. SpreadViz: Analytics and Visualization of Spreading Processes in Social Networks. In 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), pages 1324–1327, Barcelona, Spain, December 2016. IEEE.
- [ST20] Emilio Sulis and Marcella Tambuscio. Simulation of misinformation spreading processes in social networks: an application with NetLogo. In 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), pages 614–618, sydney, Australia, October 2020. IEEE.

- [Sta14] John Stasko. Value-driven evaluation of visualizations. In Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization, BELIV '14, page 46–53, 2014.
- [STP⁺17] Guodao Sun, Tan Tang, Tai-Quan Peng, Ronghua Liang, and Yingcai Wu. Socialwave: Visual analysis of spatio-temporal diffusion of information on social media. ACM Trans. Intell. Syst. Technol., 9(2), oct 2017.
- [STT81] Kozo Sugiyama, Shojiro Tagawa, and Mitsuhiko Toda. Methods for visual understanding of hierarchical system structures. *IEEE Transactions on* Systems, Man, and Cybernetics, 11(2):109–125, 1981.
- [STTL15] Lei Shi, Hanghang Tong, Jie Tang, and Chuang Lin. VEGAS: Visual influEnce GrAph Summarization on Citation Networks. *IEEE Transactions* on Knowledge and Data Engineering, 27(12):3417–3431, December 2015.
- [SWL⁺14] Guodao Sun, Yingcai Wu, Shixia Liu, Tai-Quan Peng, Jonathan J. H. Zhu, and Ronghua Liang. EvoRiver: Visual Analysis of Topic Coopetition on Social Media. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1753–1762, December 2014.
- [Tho21] W.F. Thomas. German QAnon Telegram Dataset. figshare, November 2021.
- [TP23] Martijn Tennekes and Marco J. H. Puts. cols4all: a color palette analysis tool. In *EuroVis (Short Papers)*. Wiley, 2023.
- [UA24] Mert Usul and Alessio Arleo. Peeking at Visualization Research on Information Diffusion. In *EuroVis 2024 - Posters*. The Eurographics Association, 2024.
- [VAA24] B. Vago, D. Archambault, and A. Arleo. Dyntrix: A hybrid representation for dynamic graphs. *Computer Graphics Forum*, 43(3):e15076, 2024.
- [VKPM15] Jason Vallet, Hélène Kirchner, Bruno Pinaud, and Guy Melançon. A Visual Analytics Approach to Compare Propagation Models in Social Networks. *Electronic Proceedings in Theoretical Computer Science*, 181:65–79, April 2015.
- [VLDBF15] Tatiana Von Landesberger, Simon Diel, Sebastian Bremm, and Dieter W Fellner. Visual analysis of contagion in networks. *Information Visualization*, 14(2):93–110, April 2015.
- [vLKS⁺11] T. von Landesberger, A. Kuijper, T. Schreck, J. Kohlhammer, J.J. van Wijk, J.-D. Fekete, and D.W. Fellner. Visual analysis of large graphs: State-of-the-art and future research challenges. *Computer Graphics Forum*, 30(6):1719–1749, apr 2011.

- [VWH⁺13] Fernanda Viégas, Martin Wattenberg, Jack Hebert, Geoffrey Borggaard, Alison Cichowlas, Jonathan Feinberg, Jon Orwant, and Christopher R. Wren. Google+ripples: a native visualization of information flow. pages 1389–1398, 05 2013.
- [WAA⁺23] Markus Wallinger, Daniel Archambault, David Auber, Martin Nöllenburg, and Jaakko Peltonen. Faster edge-path bundling through graph spanners. *Computer Graphics Forum*, 42(6):e14789, 2023.
- [WAM⁺19] Emily Wall, Meeshu Agnihotri, Laura Matzen, Kristin Divis, Michael Haass, Alex Endert, and John Stasko. A heuristic approach to value-driven evaluation of visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):491–500, 2019.
- [WB04] Colin Ware and Robert Bobrow. Motion to support rapid interactive queries on node–link diagrams. *ACM Trans. Appl. Percept.*, 1(1):3–18, July 2004.
- [WB05] Colin Ware and Robert Bobrow. Supporting visual queries on medium-sized node–link diagrams. *Information Visualization*, 4(1):49–58, 2005.
- [WCG⁺16] Yingcai Wu, Nan Cao, David Gotz, Yap-Peng Tan, and Daniel A. Keim. A Survey on Visual Analytics of Social Media Data. *IEEE Transactions on Multimedia*, 18(11):2135–2148, November 2016.
- [WLY⁺14] Yingcai Wu, Shixia Liu, Kai Yan, Mengchen Liu, and Fangzhao Wu. OpinionFlow: Visual Analysis of Opinion Diffusion on Social Media. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1763–1772, December 2014.
- [YAD⁺18] Vahan Yoghourdjian, Daniel Archambault, Stephan Diehl, Tim Dwyer, Karsten Klein, Helen C. Purchase, and Hsiang-Yun Wu. Exploring the limits of complexity: A survey of empirical studies on graph visualisation. Visual Informatics, 2(4):264–282, 2018.
- [YDH⁺17] Andrea Yanez, Jim Duggan, Conor Hayes, Musfira Jilani, and Maire Connolly. PandemCap: Decision support tool for epidemic management. In 2017 IEEE Workshop on Visual Analytics in Healthcare (VAHC), pages 24–30, Phoenix, AZ, October 2017. IEEE.
- [YJZ⁺25] Jianing Yin, Hanze Jia, Buwei Zhou, Tan Tang, Lu Ying, Shuainan Ye, Tai-Quan Peng, and Yingcai Wu. Blowing seeds across gardens: Visualizing implicit propagation of cross-platform social media posts. *IEEE Transactions* on Visualization and Computer Graphics, 31(1):185–195, 2025.
- [YWL⁺14] Xiaoru Yuan, Zhenhuang Wang, Zipeng Liu, Cong Guo, Hongwei Ai, and Donghao Ren. Visualization of social media flows with interactively identified key players. In 2014 IEEE Conference on Visual Analytics Science and Technology (VAST), pages 291–292, Paris, France, October 2014. IEEE.

- [ZBDS12] Michael Zinsmaier, Ulrik Brandes, Oliver Deussen, and Hendrik Strobelt. Interactive level-of-detail rendering of large graphs. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2486–2495, 2012.
- [ZCW⁺14] Jian Zhao, Nan Cao, Zhen Wen, Yale Song, Yu-Ru Lin, and Christopher Collins. #FluxFlow: Visual Analysis of Anomalous Information Spreading on Social Media. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1773–1782, December 2014.



Appendix

ICE-T Tasks

The following contains the set of tasks prepared for the ICE-T [WAM⁺19] evaluation. Each example dataset contains two tasks

InfoVis

- (i) In the Simulation "IC Miksch", there seems to be a delay in the spread of influence within the seed community. Trace the path of influence to find out why. Use the Highlighting feature to mark seemingly relevant nodes and explore the area of interest using all 3 Layers
- (ii) Highlight an author of your choice with the "Node Search & Highlight" Menu, investigate their role in the loaded simulation(s)

Qanon

- (i) Load SIR_AH and SIR_WeTheMedia. These two simulate the spread of fake news according to the SIR model through the telegram channel network, starting from 1 seed node respectively. Looking at the global and community trend charts, which simulation reaches a broader audience. Are there differences in the evolution of the diffusion of Fake News.
- (ii) Which channels are key influencers? Investigate communities in the 2nd and 3rd layer

EU Power Grid

• (i) Load the Dataset at resolution 1. Load the "LT_seed_FR_n50" simulation. It uses a LT Model to simulate a Power Grid Failure starting from 50 random Power Plants in France. After the first surge of failures there is a secondary bump in cluster 15, securing which power plant could have prevented this?

• (ii) Use the Sugiyama Layout to identify Power Plants that lead to a sizeable chain reaction of additional failures.

Facebook Pages

- (i) Load all 3 SIR simulations. We interpret this as information spreading through Facebook pages. Removed pages no longer engage with the new information. Which seed selection method reaches the most pages.
- (ii) At Louvain resolution 1, in which communities do we have a noticeably difference in coverage among the 3 simulations. Take a closer look at those communities and investigate from which neighboring communities the information is coming from.

ICE-T Questionnaires

terminology	data case- refers to individual nodes of the network ortribute, refers to meta riata Jahois metrics and diffusion	information of individual nodes							
	relationship in the data-refers to attributes among the data	s, such as clusters and diffusion paths							
			н	w would vo	u rate vour ag	reement with	h the followin	z statement	2
						Neither			
			Disagree	Disagree	Disagree	Agree nor Disagree	Agree	Agree	Agree
		The visualization exposes individual data cases and their attributes	0	0	0	0	0	0	Х
	The visualization facilitates answering questions about the data	The visualization facilitates perceiving relationships in the data like patterns & distributions of the variables	0	0	0	0	0	х	0
		The visualization promotes exploring relationships between individual data cases as well as different groupings of data cases	0	0	0	0	0	х	0
	The visualization provides a new or better	The visualization helps generate data-driven questions	0	0	0	0	0	х	0
insight	understanding of the data	The visualization helps identify unusual or unexpected, yet valid, data characteristics or values	0	0	0	0	х	0	0
		The visualization provides useful interactive capabilities to help investigate the data in multiple ways	0	0	0	0	0	0	х
	The visualization provides opportunities for serendipitous discoveries	The visualization shows multiple perspectives about the data	0	0	0	0	0	0	x
		The visualization uses an effective representation of the data that shows related and partially related data cases	0	0	0	0	0	0	X
	The visualization affords rapid parallel comprehension for efficient browsing	The visualization provides a meaningful spatial organization of the data	0	0	0	0	0	х	0
		The visualization shows key characteristics of the data at a glance	0	0	0	х	0	0	0
Time	The visualization provides mechanisms for quickly seeking specific information	The interface supports using different attributes of the data to reorganize the visualization's appearance	0	0	0	х	0	0	0
		The visualization supports smooth transitions between different levels of detail in viewing the data	0	0	0	0	х	0	0
		The visualization avoids complex commands and textual queries by providing direct interaction with the data representation	0	0	0	0	0	0	х
	The visualization provides a big picture	The visualization provides a comprehensive and accessible overview of the data	0	0	0	0	0	0	х
	perspective of the data	The visualization presents the data by providing a meaningful visual schema	0	0	0	0	0	x	0
Essence	The visualization provides an understanding of	The visualization facilitates generalizations and extrapolations of patterns and conclusions	0	0	0	0	0	0	х
	the data beyond individual data cases	The visualization helps understand how variables relate in order to accomplish different analytic tasks	0	0	0	0	0	х	0
	The visualization helps avoid making incorrect	The visualization uses meaningful and accurate visual encodings to represent the data	0	0	0	0	0	x	0
	inferences	The visualization avoids using misleading representations	0	0	0	0	0	0	×
Confidence	The visualization facilitates learning more broadly about the domain of the data	The visualization promotes understanding data domain characteristics beyond the individual data cases and attributes	0	0	0	0	0	×	0
	The visualization helps understand data quality	If there were data issues like unexpected, duplicate, missing, or invalid data, the visualization would highlight those issues		0	0	Y	0	0	0

Figure 1: Filled-out questionnaire from Participant 1

ICE-T Results

Value of Visualization

	relationship in the data- refers to attributes among the dat	a, such as clusters and diffusion paths							
			н	ow would yo	u rate your ag	reement wit	h the following	statement	3
			Strongly Disagree	Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Agree	Strongly Agree
		The visualization exposes individual data cases and their attributes	0	0	0	0	0	0	×
	The visualization facilitates answering questions about the data	The visualization facilitates perceiving relationships in the data like patterns & distributions of the variables	0	0	0	0	0	0	哭
		The visualization promotes exploring relationships between individual data cases as well as different groupings of data cases	0	0	0	0	0	0	M
Includes	The visualization provides a new or better	The visualization helps generate data-driven questions	0	0	0	0	0	0	×
magne	understanding of the data	The visualization helps identify unusual or unexpected, yet valid, data characteristics or values	0	0	0	0	0	0	哭
		The visualization provides useful interactive capabilities to help investigate the data in multiple ways	0	0	0	0	0	0	ы
	The visualization provides opportunities for serendipitous discoveries	The visualization shows multiple perspectives about the data	0	0	0	0	20	0	0
		The visualization uses an effective representation of the data that shows related and partially related data cases	0	0	0	0	×	0	0
	The visualization affords rapid parallel comprehension for efficient browsing	The visualization provides a meaningful spatial organization of the data	0	0	0	0	0	0	×
		The visualization shows key characteristics of the data at a glance	0	0	0	0	0	0	16
Time	The visualization provides mechanisms for quickly seeking specific information	The interface supports using different attributes of the data to reorganize the visualization's appearance	0	0	0	0	0	0	90
		The visualization supports smooth transitions between different levels of detail in viewing the data	0	0	0	0	×	0	0
		The visualization avoids complex commands and textual queries by providing direct interaction with the data representation	0	0	0	0	0	0	×
	The visualization provides a big picture	The visualization provides a comprehensive and accessible overview of the data	0	0	0	0	0	0	哭
	perspective of the data	The visualization presents the data by providing a meaningful visual schema	0	0	0	0	0	0	哭
Essence	The visualization provides an understanding of	The visualization facilitates generalizations and extrapolations of patterns and conclusions	0	0	0	0	0	0	×
	the data beyond individual data cases	The visualization helps understand how variables relate in order to accomplish different analytic tasks	0	0	0	0	0	×	0
	The visualization helps avoid making incorrect	The visualization uses meaningful and accurate visual encodings to represent the data	0	0	0	0	0	0	×
	Inferences	The visualization avoids using misleading representations	0	0	0	0	×	0	0
Confidence	The visualization facilitates learning more broadly about the domain of the data	The visualization promotes understanding data domain characteristics beyond the individual data cases and attributes	0	0	0	0	0	м	0
	The visualization helps understand data quality	If there were data issues like unexpected, duplicate, missing, or invalid data, the visualization would highlight those issues	0	×	0	0	0	0	0

Figure 2: Filled-out questionnaire from Participant 2

Value of Visualization

How would you rate your generates with the following states Intervisualization functions a new or better The visualization functions a new or better The visualization provides a new or better The visualization hydroxide dista class a well a The visualization provides a new or better The visualization provides a new or better The visualization provides and dista direct expandibility to help integrate the data in multiple O	strongly Agree O V V V
Interface Name	Strongly Agree
The visualization exposes included data cases and their attributes The visualization exposes included data cases and their attributes The visualization functionary and the attributes The visualization functionary and the attributes The visualization functionary and their attributes The visualization functionary	· · ** · **
The visualization facilitates answering quantities in the data is answering quantities are visible in the data is the pattern & distributions of the data is a visible in the data is a set of the data in the pattern is distributions of the data is a visible in the data in the visible intervisible in the data is a visible in the data in the visible intervisible in the data is a visible in the data in the visible intervisible intervisible in the visible intervisible intervisible intervisible intervisible intervisible in the visible intervisible intervisible capabilities to help intervisible in the data in the visible intervisible intervisible capabilities to help intervisible inte	。 メメ 。 メメ
Insight The visualization provides a previous registry stationalitys between individual data cases as well as difference propunging of that cases as well as The visualization provides a new or better individual data cases as well as The visualization provides a new or better individual data cases as well as The visualization helps benefit y unusual or unexpected, yet valid, data characteristics or values The visualization helps benefit y unusual or unexpected, yet valid, data characteristics or values The visualization provides useful interactive capabilities to help investigate the data in mutriple ways are usualized on the provide set of the data of the data in mutriple ways are usualized on the provide set of the data in mutriple ways are usualized on the provide set of the data in mutriple ways are usualized on the data in the data in mutriple ways are usualized on the data in the data in mutriple ways are usualized on the data in the data	XX · XX
Insight The visualization provides a new or better understanding of the data The visualization helps generate data diven questions The visualization helps identify unusual or unequested, yet valid, data characteristics or values The visualization provides useful interactive capabilities to help investigate the data in mutigie ways	× ° ××
mingen understanding of the data The visualization helps: identity musual or unexpected, yet valid, data characteristics or values 0 0 0 0 V 0 0 The visualization provides useful interactive capabilities to help investigate the data in mutigite 0 0 0 0 0 0 0 0 0 0	° ××
The visualization provides useful interactive capabilities to help investigate the data in multiple willy/s	XX
	×
The visualization provides opportunities for The visualization shows multiple perspectives about the data OOOOOOOO	
The visualization uses an effective representation of the data that shows related and partially related data cases O O O O O O O O O	×
The visualization affords rapid parallel The visualization provides a meaningful spatial organization of the data OOOOO 🏹	0
comprehension for efficient browsing The visualization shows key characteristics of the data at a glance O O O O O O	0
Time The interface supports using different attributes of the data to reorganize the visualization's OOOOX	0
The visualization provides mechanisms for quickly seeking specific information data The visualization supports smooth transitions between different levels of detail in viewing the data O O O	0
The visualization avoids complex commands and textual queries by providing direct interaction with the data representation O O O O O O O O O	×
The visualization provides a big picture The visualization provides a comprehensive and accessible overview of the data 0 0 0 0 0 0 0	0
perspective of the data The visualization presents the data by providing a meaningful visual schema 0 0 0 0 0 0 0	0
Essence The visualization provides an understanding of OOOOO	0
the data beyond individual data cases The visualization helps understand how variables relate in order to accomplish different analytic tasks O O O O O O O O O O	×
The visualization helps avoid making incorrect	0
Inferences The visualization avoids using mideading representations OOOOOOO	×
Confidence The visualization facilitates learning more troubulation promotes understanding data domain characteristics beyond the individual broadly about the domain of the data tractases and attributes 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0
The visualization helps understand data quality if there were data issues like unexpected, duplicate, missing, or invalid data, the visualization would highlight those issues	0

Figure 3: Filled-out questionnaire from Participant 3

Value of Visualization

terminology	g. disc care when is indicated and and its antown antibush-whet is such add, back, here, areas and additions of indicated and as relationships in the data-refers to attribute arrange the data, such as clusters and diffusion pands.										
			How would you rate your agreement with the following statements								
			Strongly Disagree	Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Agree	Strongly Agree		
		The visualization exposes individual data cases and their attributes	0	0	0	0	0	×	0		
	The visualization facilitates answering questions about the data	The visualization facilitates perceiving relationships in the data like patterns & distributions of the variables $% \left(\frac{1}{2}\right) = 0$	0	0	0	0	0	\checkmark	0		
		The visualization promotes exploring relationships between individual data cases as well as different groupings of data cases	0	0	0	0	0	0	×		
	The visualization provides a new or better	The visualization helps generate data-driven questions	0	0	0	0	0	0	*		
Insight	understanding of the data	The visualization helps identify unusual or unexpected, yet valid, data characteristics or values	0	0	0	X	0	0	0		
		The visualization provides useful interactive capabilities to help investigate the data in multiple ways	0	0	0	0	0	0	×		
	The visualization provides opportunities for serendipitous discoveries	The visualization shows multiple perspectives about the data	0	0	0	0	0	X	0		
		The visualization uses an effective representation of the data that shows related and partially related data cases	0	0	0	0	0	*	0		
	The visualization affords rapid parallel comprehension for efficient browsing	The visualization provides a meaningful spatial organization of the data	0	0	0	0	0	\checkmark	0		
		The visualization shows key characteristics of the data at a glance	0	0	0	0	*	0	0		
Time	The visualization provides mechanisms for quickly seeking specific information	The interface supports using different attributes of the data to reorganize the visualization's appearance	0	0	×	0	0	0	0		
		The visualization supports smooth transitions between different levels of detail in viewing the data	0	0	0	>	0	0	0		
		The visualization avoids complex commands and textual queries by providing direct interaction with the data representation	0	0	0	0	0	0	×		
	The visualization provides a big picture	The visualization provides a comprehensive and accessible overview of the data	0	0	0	0	0	×	0		
	perspective of the data	The visualization presents the data by providing a meaningful visual schema	0	0	0	0	0	×	0		
Essence	The visualization provides an understanding of	The visualization facilitates generalizations and extrapolations of patterns and conclusions	0	0	0	0	0	×	0		
	the data beyond individual data cases	The visualization helps understand how variables relate in order to accomplish different analytic tasks	0	0	0	0	Y	0	0		
	The visualization helps avoid making incorrect	The visualization uses meaningful and accurate visual encodings to represent the data	0	0	0	0	0	0	X		
	Interences	The visualization avoids using misleading representations	0	0	0	0	0	×	0		
Lonridence	The visualization facilitates learning more broadly about the domain of the data	The visualization promotes understanding data domain characteristics beyond the individual data cases and attributes	0	0	0	0	0	0	×		
	The visualization helps understand data quality	If there were data issues like unexpected, duplicate, missing, or invalid data, the visualization would highlight those issues	0	0	0	$\mathbf{\mathbf{x}}$	0	0	0		

Figure 4: Filled-out questionnaire from Participant 4

Value of Visualization

			Ho	w would yo	u rate your ag	reement wit	h the following	statement	s?
			Strongly Disagree	Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Agree	Strongly Agree
		The visualization exposes individual data cases and their attributes	0	0	0	0	0	•	0
	The visualization facilitates answering questions about the data	The visualization facilitates perceiving relationships in the data like patterns & distributions of the variables	0	0	0	0	•	0	0
		The visualization promotes exploring relationships between individual data cases as well as different groupings of data cases	0	0	0	0	0	•	0
-1-0-1	The visualization provides a new or better	The visualization helps generate data-driven questions	0	0	0	•	0	0	0
Insight	understanding of the data	The visualization helps identify unusual or unexpected, yet valid, data characteristics or values	0	0	0	0	•	0	0
		The visualization provides useful interactive capabilities to help investigate the data in multiple ways	0	0	0	0	0	0	•
	The visualization provides opportunities for serendipitous discoveries	The visualization shows multiple perspectives about the data	0	0	0	0	0	•	0
		The visualization uses an effective representation of the data that shows related and partially related data cases $% \left($		0	0	0	0	•	0
Time	The visualization affords rapid parallel comprehension for efficient browsing	The visualization provides a meaningful spatial organization of the data	0	0	0	•	0	0	0
		The visualization shows key characteristics of the data at a glance	0	0	0	0	0	•	0
	The visualization provides mechanisms for quickly seeking specific information	The interface supports using different attributes of the data to reorganize the visualization's appearance		0	0	0	0	0	
		The visualization supports smooth transitions between different levels of detail in viewing the data		0	0	0		0	0
		The visualization avoids complex commands and textual queries by providing direct interaction with the data representation	0	0	0	0	0	0	•
	The visualization provides a big picture	The visualization provides a comprehensive and accessible overview of the data	0	0	0	0	0	0	•
	perspective of the data	The visualization presents the data by providing a meaningful visual schema	0	0	0	0	•	0	0
isence	The visualization provides an understanding of	The visualization facilitates generalizations and extrapolations of patterns and conclusions	0	0	0	0	•	0	0
	the data beyond individual data cases	The visualization helps understand how variables relate in order to accomplish different analytic tasks	0	0	0	0	0	•	0
	The visualization helps avoid making incorrect	The visualization uses meaningful and accurate visual encodings to represent the data	0	0	0	0	0	•	0
	Inferences	The visualization avoids using misleading representations		0	0	0	0		0
.nfidence	The visualization facilitates learning more broadly about the domain of the data	The visualization promotes understanding data domain characteristics beyond the individual data cases and attributes		0	0	0	•	0	0
	The visualization helps understand data quality	If there were data issues like unexpected, duplicate, missing, or invalid data, the visualization would highlight those issues		0	0		0	0	0

Figure 5: Filled-out questionnaire from Participant 5

Ps	l	Т	E	С	TOTAL
P1	6,28	5,17	6,50	5,50	
P2	6,78	6,83	6,75	4,67	
P3	6,40	5,50	6,25	5,17	
P4	6,05	5,08	5,7 <mark>5</mark>	5,83	
P5	5,50	5,70	5,75	5,00	
Avg	6,20	5,66	6,20	5,23	5,82
S.D.	0,473096	0,702232	0,447214	0,448029	0,469248

Figure 6: Average score per ICE-T [WAM $^+19$] component per participant (P1-5) with Total scores and standard deviation