Communication Analysis through Visual Analytics: Current Practices, Challenges, and New Frontiers

Maximilian T. Fischer, Frederik L. Dennig, Daniel Seebacher, Daniel A. Keim, and Mennatallah El-Assady

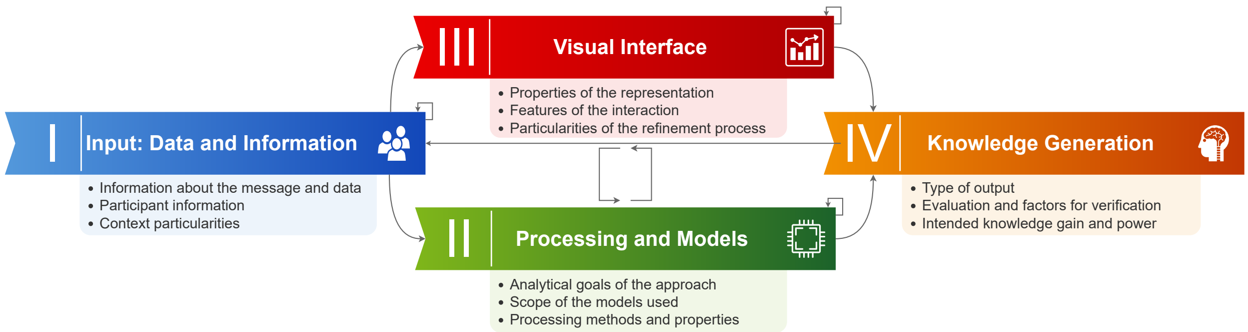

Fig. 1. **Characterization of the four main dimensions of our conceptual framework** for communication analysis systems, in the form of a concrete application of the *visual analytics process model* by Keim et al. [48] to the communication analysis domain.

Abstract— The automated analysis of digital human communication data often focuses on specific aspects such as content or network structure in isolation. Thereby, it often suffers from a limited perspective and makes cross-methodological analyses common in many domains, like investigative journalism, difficult. Communication research in psychology and the digital humanities instead stresses the importance of a holistic analysis approach to overcome these limiting factors. In this work, we conduct an extensive survey on the properties of over forty current semi-automated communication analysis systems and investigate how they cover concepts described in theoretical communication research. From these investigations, we derive a design space and contribute a conceptual framework based on communication research, technical considerations, and the surveyed approaches. The framework describes the systems' properties, capabilities, and composition through a wide range of criteria organized in the analysis dimensions (1) Data, (2) Processing and Models, (3) Visual Interface, and (4) Knowledge Generation. These criteria enable a formalization of digital communication analysis through visual analytics, which, we argue, is uniquely suited for this task by tackling automation complexity while leveraging domain knowledge. With our framework, we identify shortcomings and research challenges, such as group communication dynamics, trust and privacy considerations, and holistic approaches, for which we discuss relevant design considerations. Simultaneously, our framework supports the evaluation of systems and promotes the mutual exchange between researchers through a common language and taxonomy, laying the foundations for future research on communication analysis through visual analytics.

Index Terms—Communication analysis, visual analytics, conceptual framework, design space, state of the art.

1 Introduction

Human communication has been fundamentally transformed, especially in the last two decades, becoming increasingly digital, with cost-effective, location-independent, and instant access changing communication behavior. With this transformation to digital communication [80], new research opportunities have emerged in a wide variety of different domains, ranging from engineering to social sciences to business: For example, it has been studied how visualization can show the evolution of dynamic communication networks [86], how discourse analysis for digital communication can be enhanced [41], or how team communication performance in business settings can be evaluated [24]. For such analyses, digital analysis methods are often used to aid and support the (semi-)manual, domain-specific research methodologies.

In this paper, we focus on the field of interactive human communication analysis and specifically on **automated and interactive communication analysis *systems*** that target written human communication (in the following: communication analysis systems), most commonly e-mails, chats, or documents. For the purpose of this paper, we define these systems as semi-automated applications that also employ visual components to interactively analyze digital human communication. We do not consider approaches focusing primarily on a single methodology like sentiment analysis, but those that aim (to some degree) at a cross-methodological analysis among multiple parties, which becomes increasingly relevant in domains like investigative journalism [8].

As we highlighted in previous work [22], the research into communication analysis *systems* often lacks [8, 88] **cross-methodological** aspects: the vast majority of systems focus on either the content of communication *or* on the network aspect *in isolation* instead of taking into account the fundamental dynamics holistically. This is in contrast to seminal works on human communication research [62, 90], recent textbooks [63, 70], or current communication research in psychology or digital humanities [24, 64], where often – even when digitally supported [19, 91] – a holistic view is taken to consider explicit and implicit connotations together and in context. In contrast, the individual analysis of content, network, and metadata aspects alone can – for interrelated tasks – lead to an incomplete or biased view, while isolated approaches often introduce discontinuities, increasing manual work and hampering the cross-methodological detection of cross-matches.

Existing frameworks on digital communication analysis *systems* do not adequately cover these issue due to four reasons: First, the need

---

• *M. T. Fischer, F. L. Dennig, D. Seebacher, and D. A. Keim are with the University of Konstanz, M. El-Assady is with the AI Center, ETH Zürich. E-mail: {max.fischer, frederik.dennig, daniel.seebacher, keim}@uni-konstanz.de, melassady@ai.ethz.ch.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

for such a revised formalization has been recognized [87] in communication sciences. So far the opportunities, challenges, and pitfalls have primarily been described from an application domain-oriented perspective [19, 91] in the social sciences, while a systematic description is missing, only available for social-media-based approaches [19, 91]. Ethical considerations [21] so far play only a small role in the system design. Second, recent efforts have begun to map digital communication systems as a whole [23], with a focus on content, infrastructure, and policy aspects, but leaving out the technical considerations, like methods, interfaces, and interaction concepts. Third, the same is true for the classical communication analysis research [62, 70, 90], which lacks technical considerations and is primed for analog but not digital communication. For example, facial expressions and gestures are not retained. Fourth, digital communication has also transformed the way we communicate and the modalities we use [67], like shorter messages or emoji reactions, requiring an updated framework.

In this work, we want to bridge the **gap** between communication research and modern communication analysis system development. As evident in a few academic works [22, 35, 50, 92] and recent commercial systems [15, 66, 68], visualization and interactive user steering is a promising way [12, 32, 40, 48, 75, 91, 93] to begin to tackle the gap between different analysis modalities. Lack of a common description from both a technical perspective and psychological communication research has made the systematic exploration of the field difficult. This also prevented a broader analysis of how visual analytics principles are – and could be – employed in communication analysis, how such systems can be categorized, and what a relevant taxonomy would look like, hindering comparison between approaches. The main objective of this work is to explore these systems from a primarily capability-oriented perspective, in terms of communication research, technical state of the art, and human factors. While we consider and point out these human factors and ethical considerations as much as possible within this framing, we also refer to our accompanying paper [21] on ethical awareness and human factors in communication analysis. We refer to this work several times over the following sections for a more detailed background and a broader, in-depth discussion.

As part of this work, we survey state-of-the-art approaches and investigate concepts in communication research to derive a **design space** on communication research, making the following contributions:

- The creation of a **conceptual framework** (see Figure 3) of communication analysis systems, based on communication research, technical considerations, and a systematic review.
- A state-of-the-art **survey** and comparison of existing approaches, assessing their maturity and coverage (see Table 2 and the **interactive browser** at https://communication-analysis.dbvis.de)
- A discussion on the open challenges and implications for future research **opportunities** on communication analysis systems.

With this contribution, we identify research challenges and aid the comparison of approaches while creating a taxonomy for future research on communication analysis through visual analytics.

2 Background

Communication analysis can use a variety of different techniques to analyze communication behavior. The complexity and ambiguity of the exchanges and modalities [47] make automation difficult. As such and due to the privacy and ethical considerations, communication analysis is well suited [21] for **visual analytics** [48], which describes the concept of combining computational data analysis with interactive human sensemaking for knowledge generation. This has been formalized [75] as a knowledge generation process with feedback loops for the steps' exploration, verification, and knowledge generation.

The origins of **communication research** can be traced backed to ancient times, with the study of rhetoric and oratory as well as persuasion in Ancient Greece and Rome. However, as we described in previous work [22], the formal study of communication as a process did not really start until the early 20th-century through researchers like Simmel, Cooley, Lippmann or Moreno, to be later extended by Bevelas [5] and Leavitt [5] for groups as well as Shannon [82] and Savage and Deutsch [78] for computer-aided modeling. From the 1960s onward, the seminal works of McLuhan [62], Watzlawick et al. [90], and von Thun [79] established the field, which was later extended by Roger [73], now encompassing a wide range of different techniques [63, 70], from natural language processing over social network analysis to metadata exploration. They all share a strong focus on the **human factors** and **communication context**, in the form of channels (Watzlawick et al.) or the medium (McLuhan), forming an essential aspect of the analysis: Be it phrasing or omissions in the face of power relations, narrow- or broadcasting to target audiences, subtle implicit messages between the lines, expectations of confidentiality (and correlating frankness), or effort in crafting the messages.

With the **advent of digital** processing and computational power came the shift from laboursome manual analysis [31] first to digitally supported [58] and later to highly automated analysis. However, it is noticeable [22], while not completely surprising, that the completeness of the analysis developed in the opposite direction of automation level, with solutions being ever more specialized and focusing on just specific aspects the more they are automated. For example, modern systems allow us to analyze communication behavior and social ties using centrality measures [60] or describe complete artificial networks in social sciences [6]. Specialized toolkits have been developed to analyze such network structures, like Pajek [4] or Gephi [3]. All these approaches primarily focus on the network aspects, omitting most of the meta-data and especially the content. Other approaches instead focus only on it, like keyword-based searches [94] to filter communication for certain content or aiming to improve the understanding of communications meaning' through sentiment analysis [69] or topic modeling [72].

From a **visual perspective**, several approaches primarily follow a node-link-diagram-based approach, like Gephi [3], and many commercial solutions like IBM's i2 Analyst's Notebook [45], Pajek [4], Palantir Gotham [68], DataWalk [15], and Nuix Discover and Nuix Investigate [66]. Another class of approaches uses matrix-based approaches to analyze the communication relations, for example, MatrixExplorer [38] or NodeTrix [39]. Another set of approaches uses timeline designs like CloudLines [51], while others like Fu et al. [29] modify graph presentations through multiple planes. For a detailed discussion, we refer to our previous work [22], where we discuss and develop the visual analytics research landscape for holistic systems.

3 Methodology

In the following, we aim towards tackling the central question of a **common description** of communication analysis: *How can the different approaches in communication analysis systems be described within a common, conceptual framework to allow their mutual comparison?*

Framework Basis — We propose to base such a framework on three areas of consideration: (1) The **existing research landscape** of interactive communication analysis systems provides a foundation for the classification of approaches based on measures such as analytical goals, visualization and interaction methods, or the power of the knowledge generation process. (2) **Communication research** offers decades of research on the particularities of (often non-technical) communication analysis. For this work, we consider concepts from seminal and more recent summary works [12, 62–64, 70, 79, 82, 90], described in Section 2. Additionally, we study relevant theoretical (non-system) works in computer science, like a survey on text visualization [52], group discourse and role analysis [26, 43, 44, 56], as well as works dealing with semi-manual approaches and user studies, including the human factors (e.g., [8, 13, 21, 25, 30, 46, 61, 74]). However, many of these works miss the transfer from a theoretically analyzed concept to an actual system implementation. (3) **Technical considerations** of the approaches, taking into account design properties such as analyzable data types, data representation, and flow, or limitations like scalability from a technical standpoint. We discuss the findings from considerations (2) and (3) later in Section 4, while (1) requires a broad review:

Existing Research Landscape – Seed Papers — To analyze the state of the art and contribute one angle of classification criteria, we start with a keyword-based seed literature survey. As we aim to inform

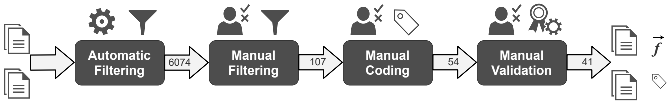

Fig. 2. **The paper collection and coding process.** It consists of four main steps: (1) Automated filtering, (2) manual filtering, (3) manual coding, and (4) manual validation.

about the most common ideas in *visual analysis* applications, we restrict the search to the following high-quality journals and conferences:

- IEEE Trans. of Vis. and Comp. Graph. (**TVCG** and **IEEE VIS**)
- Computer Graphics Forum (**EuroVis** and EuroVA)
- Proc. of the **CHI** Conference on Human Factors in Comp. Sys.

We discard technically obsolete approaches (older than 2006, i.e., 15 years), and intentionally neglect niche techniques [29, 96] that have not made in common visualization research canon or published in other journals like Digital Investigation [35]). The high number of initial approaches, and the high discard rate for CHI is due to the abundant use of the phrase *communication* when referring to user actions.

Selection Methodology — For the actual paper selection methodology, we follow a four-step approach (see also Figure 2 and Table 1). First, we conducted a keyword-based seed search for the words *communication* and *analysis* on the titles, abstract, index terms, and contents of publications in each of the venues described above. Secondly, we went through all these papers' titles and abstracts manually, discarding those which clearly are not concerned with communication analysis systems, reducing the selection significantly. In this step, we included approaches suggested by the domain experts. Third, we manually looked at the remaining papers and decided if they indeed describe a communication analysis system. In the final step, we validated the results by checking borderline cases, consequently removing seven papers. Or final collection includes 41 approaches.

Domain Expert Consultation — To broaden the perspective, we consulted with eight domain experts about **approaches used in practice**. The experts belong to the field of law enforcement, working for various European law enforcement agencies, and each has extensive experience with digital investigations, including communication analysis, working in the field from ten to over 30 years. They contributed a collection of six actual systems [3, 4, 15, 45, 66, 68] applied in the field (including commercial). Those approaches were included in the Selection Methodology from Step 2 onward and were not simultaneously discovered during the seed paper selection. However, some approaches [3, 45] can be considered to be universally known in the community. Further, the **domain experts** contributed **insights on their needs** and **perceived challenges**: They consider it unlikely an autonomous system can completely replace an experienced-saturated investigator with years of domain-specific knowledge [21] except in the narrowest or specialized of tasks. As soon as incomplete information is involved (virtually ever) and decisions under uncertainty have to be taken, the analysts often follow their hunches, exploring different options, but having difficulty in articulating their reasoning [22]. They explore related and connected information, which they consider important for contextual information [22]. As such, they are used to - and strongly prefer - visually-interactive tools for investigations, as it supports their understanding through rapid-feedback mechanisms [22], increasing their trust [21]. Nevertheless, many experts are open to new developments and consider systems their companions, supporting them without patronizing or limiting them [21], relieving them of labor-some manual work. However, they have to be developed with analysts in mind [21], otherwise potentially overwhelming them or missing key functionality like the inclusion of analog domain knowledge [20]. Black box AI models are received critically except for hints, as the domain experts are often no AI experts, lacking opacity and making it difficult to prove provenance and a chain-of-reasoning that fulfills moral or legal obligations [21]. Developing systems fulfilling these requirements while leveraging reliable XAI methods are the key challenges.

Table 1. Publications per venue and paper count for collection step.

Venue	#Coll.	#Filtered	#Coded	#Final
IEEE TVCG	790	35	27	23
Computer Graphics Forum	495	17	11	8
CHI Proceedings	4789	49	10	4
Commercial Systems	-	6	6	6
Total	6074	107	54	**41**

4 Conceptual Framework

In the following section, we aim to construct a framework that encompasses discerning aspects of communication analysis systems. As with any taxonomy, the framework is *one* possible version of a taxonomy, developed in several iterations. We justify our considerations overall and for each property, referencing relevant work when applicable. As a note of caution, we stress that many properties we present in the following are themselves multifaceted, and our choices and considerations can benefit from a critical discussion within the community. For the complete conceptual framework, see Figure 3. For the full classification of the surveyed approaches from above, consult Table 2.

Main Considerations — When designing a conceptual framework, the **structuring methodology** is of central importance. One standard methodology is to use a task-based grouping [12, 20]. However, sometimes very different methods are employed for the same task: for example, for discovering key persons in a communication network, SNA-based [32] approaches using centrality measures and node-link visualizations are equally applicable as geometric deep learning models [20] interactively visualized using a matrix-based method. However, both methods have very different side effects, visualization, and interaction techniques and make very different assumptions and requirements on the data. Instead, we follow the second large methodology and use a property, representation, and methods-based taxonomy [52].

As our **primary goal** is to design a conceptual framework of communication analysis through *visual analytics*, we motivate the main areas by Keim et al.'s established process model [48], but develop each area specifically for communication analysis using considerations from communication research and our survey. We slightly modify Keim et al.'s terminology, proposing **four main dimensions**, characterized further in Figure 1 and the following sections: (I) Input: Data and Information (4.1) encompasses the (inferred) content and context with respect to communication research, (II) Processing and Models (4.2) discusses the analytical goals and scopes of the systems, (III) Visual Interface (4.3) presents visual and interaction techniques employed, and (IV) Knowledge Generation (4.4) discusses the information flow.

4.1 Input: Data and Information

This category focuses on information, context, and environment of the communication, in particular theoretical aspects and data properties, with the structuring partly based on classical communication research [12, 62–64, 70, 79, 82, 90]. Therefore, we are discussing the content and meaning, context, and relationship aspects of communication extensively. Building on established frameworks [79, 82, 90], we propose to focus on three closely-interrelated areas: the information as *message*, the *communication participants*, and the *environment* (or context).

4.1.1 Message

The *message* [79] (also central channel [82] or content [90]) refers to the entailed information. From a system's perspective the distinction by *data type* is obvious, while the information can be considered from its actually transported content (*coding* [90]) and its orthogonal interpretation (*expression* levels [79]):

Data Type — The data type refers to the content type from a technical point of view. When looking at data classification in information visualization [11], we can identify several data types which are relevant for communication analysis: *text*, *audio*, *image*, *video*, and meta-data,

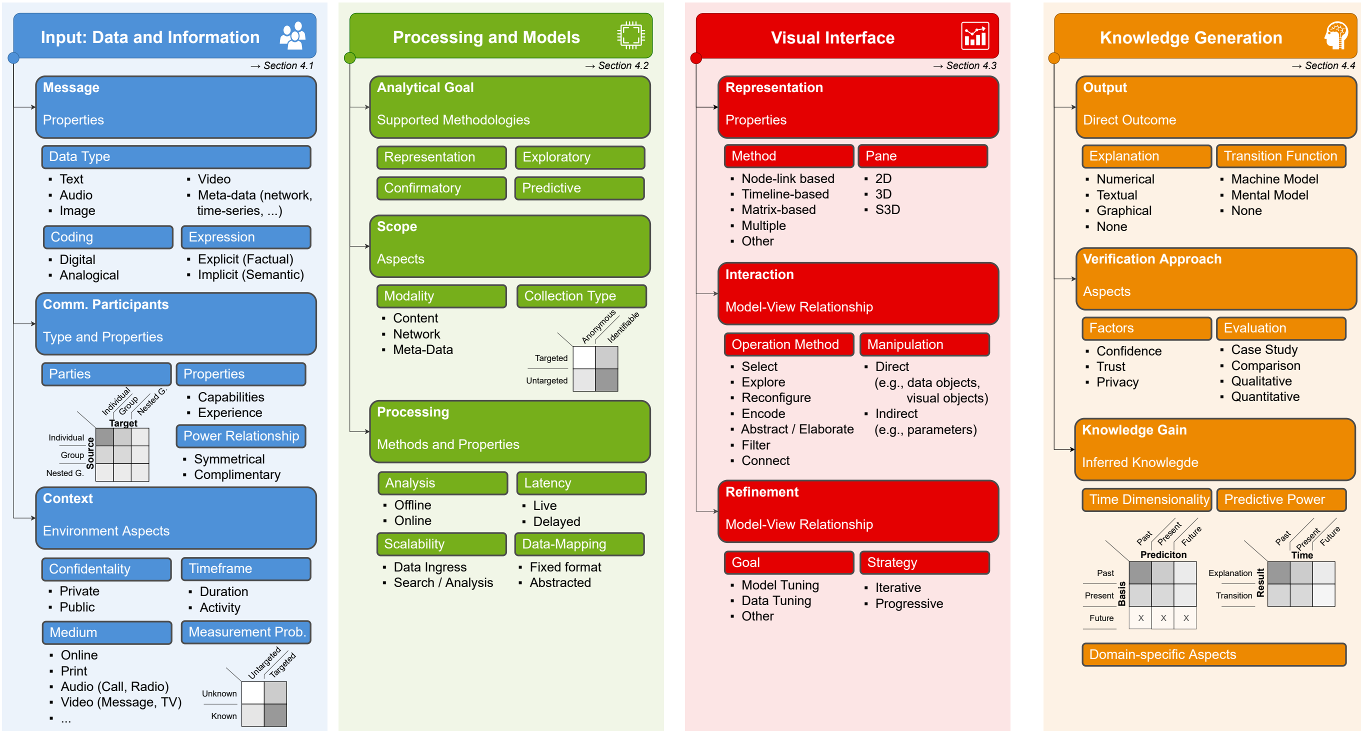

Fig. 3. **Conceptual framework of communication analysis systems.** It consists of the four main dimensions *Input: Data and Information*, *Processing and Models*, *Visual Interface*, and *Knowledge Generation*, described in more detail in Section 3 and also in Figure 1. Each category contains several properties and sub-properties, which allow for systematical analysis of such systems. Note that the graphic highlights some of the most relevant aspects for individual properties, which can be used for a simplified comparison. However, the properties themselves are multifaceted and, for a detailed analysis, should be discussed more nuanced and in more detail than indicated by these examples.

related to *network* as well as *time-series*. Based on the usage in current approaches (see Section 3), the two most relevant ones are text data (e.g., extracting topics from text [14]) and relation network (structure) data (e.g., social graphs between communication participants [3]). However, communication can also happen via audio (e.g., telephone or VoIP) or via video chats, comprising audio and moving images, i.e., video data. While our framework focuses primarily on this written (i.e., text) communication, we include these types for completeness. Therefore, it would be possible to include the detection of facial expressions using deep learning [65] to analyze the analogical code and set it into context (see below). Meta-data in the form of time-series data (e.g., [51], extracting event order and relevance) is often relevant for the communication *context*, for example regarding regularity and duration.

Coding — The transported meaning is a core aspect of the communication, which splits into spelled out (*coding*) and inferred (*expression*) meaning. Based on Watzlawick et al. [90], the spelled out communication content [62] can be regarded as coding, either in *digital* or *analogical* (sic!) form. The digital code roughly refers to the actual meaning of the transmitted information in a symbolic system (e.g., the writing "the sky is blue"), while the analogical code refers to how something is communicated, including cues (e.g., biosignals, like winking, or emoticons). Analogical analysis is rare (e.g., message sentiment [16, 42]), partly due to the information loss in digital transmissions.

Expression — Similar, but orthogonal to it is the *expression*, which describes the intended or inferred information extracted from the content. It can be *explicit* factual information (the fact that the sky is blue) or information *implicitly* contained and must be inferred, for example, from the semantics (or character [62] of the message). For example, the sky is blue - "let's go hiking now". Code and expression together allow to classify systems by their capability to leverage both digital (e.g., actual content information) and also analogical codes (e.g., inferred as sentiment analysis) while judging support for explicit (e.g., keyword-based search like [45]) and also implicit (e.g., named entity recognition like [17]) content. Most approaches consider the explicit level, and several, especially text-based ones, also the implicit level.

4.1.2 Communication Participants

Of central importance are the participants in a communication. The *scale* of the communication is determined by its audience. In correlation with the *Context*, this determines different modes like narrow-casting (few, restricted participants), broad-casting (large audience), or targeting (specific participants), in turn influencing (or being influenced by the communication *medium*). In communication research [62, 90] these aspects are usually considered as part of the context (see below).

Parties — The involved types of the parties can be a single other participant (with oneself as a special case) or different forms of groups (homogeneous groups or heterogeneous groups with subgroups) and differ between the sender and receiver sides. Therefore, we propose to structure the approaches based on their support to analyze the communication between a source and a target in a 3x3 matrix (individual, group, nested groups), e.g., individual to individual is encoded as (e.g., no group support whatsoever [10]). Counterintuitively, the matrix might not always be symmetric.

Properties — The properties of these participant(s) can be manifold. One possible classification can describe them along their *capabilities* and their *experience* (knowledge of context).

Power Relationship — The (power) relations between the parties have a strong influence, with differences in *push* and *pull*. A possible classification [90] distinguishes between *symmetrical* (equal grounds) or *complementary* (dependence) relations. However, clearly, the relations can be described in more detail. The relationship is rarely taken into account explicitly in existing research (e.g., [9] analyses the changing relations inside a group during information diffusion).

4.1.3 Context

The context of communication is essential [62, 82, 90], because it strongly affects the implicit interpretation among participants. We focus on the context of the external environment (*confidentiality*, *measurement*, and *medium*), and of the message (*timeframe*).

Confidentiality — The confidentiality of the communication channel can strongly influence the communication coding, for example through aversion or code-words (see also *factors* in Section 4.4 and our work [21] on human factors).

Measurement Problem — Closely related is the measurement problem, where the analysis interferes and influences the communication coding and expression simply due to its (possible) presence. The communication is affected by the participants' awareness of the implications, so they might adapt their behavior, use coded language, are less honest, implicit, or communicate not at all [21]. This also concerns trust and reliability, both for the parties as well as the analysts [13, 21]. We categorize this aspect into a quadrant, between (expected) knowledge about the analysis, which can be either known or unknown to be true. Note that the inverse, known to be false, is only achievable with a specific degree of certainty. This effect also depends on the targeting: If it is known that all e-mails in the world are collected and analyzed (untargeted), they might have less influence than knowing that their e-mails are monitored for sure (targeted), as this can influence the degree of scrutiny and sophistication of the analysis employed. For example, diplomatic cables are often very frank, mainly because they are not expected to be leaked (assume untargeted, identifiable), while public policy discussions are much more considerate (targeted, identifiable).

Timeframe — The timeframe when communication is occurring is highly relevant (e.g., for event correlation [85]). It can be described from the perspective of its duration and the activity during it [81].

Medium — The communication medium [62, 79] is partly covered (or mutually induced) by the participants and also the message type, coding, expression, and other contextual factors. Nevertheless, it deserves its own spot, in particular, due to media-typical characteristics and its relevance in research [62].

4.2 Processing and Models

After defining the data and information available, we study the particularities of processing and model creation from this information. We consider a technical perspective in visual data analysis, following the Golden Circle model by Simon Sinek [83] to answer the why? (*Analytical Goal*), the what? (*Model Scope*), and the how? (*Processing*).

4.2.1 Analytical Goal

We start with why [83], categorizing them by the aim of the analysis, which determines the analytical tasks to achieve it. We align our classification by the standard definition of analytical tasks in visual analytics [77]. This includes the category *representation* for a fixed analysis task (present existing data), *confirmatory* analysis for a directed search (to validate a hypothesis about the data), and *exploratory* analysis for an undirected search (find interesting anomalies in the data). Another goal in communication analysis involves *predictive* analysis (e.g., to analyze the future diffusion of information [92]), to draw conclusions from the data, which could also be regarded as overarching all methodologies and is related to the knowledge extracted (see Section 4.4).

4.2.2 Scope

The scope answers the what?, determining the generic capabilities *on* the information. Other scopes are defined by their data (see Section 4.1) and knowledge generation (see Section 4.4) support.

Modality — The analysis modality categorizes into the three [22] core aspects: *meta-data* (e.g., like time-series [51]), *network* (e.g., social graphs [3]), and *content* (e.g., conversation order [18]).

Collection Type — The collection type is the logical composite to the *Measurement Problem*, defining how the data was acquired and its corresponding analysis implications. We propose to categorize it into a quadrant between targeting methodology and anonymity level (see the relevant part in Figure 3). The former can be either targeted (specific communication from a restricted set of users) or untargeted (unfound bulk collection). The latter can either be high (anonymized or pseudo-anonymized) or low (identifiable). Different configurations might pose particular challenges to the analysis model regarding aspects such as scalability and inference capability through class imbalance or uncertainty [21]. For example, the targeted analysis of identifiable communication participants can focus on the actual exchange and leverage context and relationship information. The untargeted analysis of pseudo-anonymized communication instead often results in a search for the needle in the haystack and can rarely leverage background.

4.2.3 Processing

Due to the breadth of different methods, we focus on generic aspects, namely the *analysis* approach and *latency*, *scalability* as KPI, and *data-mapping* as power.

Analysis — The employed techniques and algorithms often differ significantly between *offline* analysis and *online* analysis. Loading a dataset once would be considered the former type, while batch (e.g., updating data with changes [45] and, in particular, streaming approaches can be classed as the latter. Most approaches only cover offline analysis.

Latency — The latency is orthogonal to the analysis. Research [70] indicates that latency in the communication can significantly affect it, as well as its analysis. The two primary options are (nearly) instantaneous communication, like in an active *live* L chat (e.g., live monitoring and analysis [68]) or *delayed* D communication, such as e-mail or as a document Differentiation into these two groups [70] is often enough for most differences in reaction and behavior, although the latency can play a role (e.g., answering under time pressure).

Scalability — The scalability of a KPI can be defined on two levels: First, on the *data-ingress* level, which defines the amount a system can import, analyze, and visualize initially. The second aspect is the scalability on the *search* and analysis side, for example, during exploratory analysis. For example, how many results can be shown simultaneously? We roughly categorize both aspects into few (less than ten, I), medium (order of hundredths to thousands, II), and huge (more than 10k, IIII).

Data-Mapping — Supporting data mapping increases the analytical power of the systems. Supporting a flexible import system that allows mapping properties in contrast to a fixed data format often aligns with support for merging different data sources. For example, many systems cannot load multiple datasets and combine fields like sender and username but only consider a single dataset (e.g., a set of e-mails) in a fixed format. Only a few support the flexible integration from multiple data sources (e.g., supporting the combination of data-sets and data field operations like renaming [45]).

4.3 Visual Interface

While there can be many design principles involved [12], we describe the visual interface abstractly [48], focusing on three interrelated concepts: *representation* for the visualization, the techniques employed in *interaction*, and the synthesis of both through *refinement*.

4.3.1 Representation

The central aspect of visualization systems is their representations.

Method — We follow the established nomenclature of visualization techniques [48]. However, we only chose those common in communication analysis: *node-link-based* (e.g., [4]), *timeline-based* (e.g., [33]), and *matrix-based* (e.g., [20]). *Other* (e.g., chord diagrams [16]) techniques are grouped, while we additionally highlight *multiple-paradigm* (e.g., timeline, graph, and text [28]) approaches.

Pane — The different visualization methods can be employed in different visualization panes. We consider the three major ones, namely **2D**, **3D**, and **S3D** (stereoscopic 3D like VR or AR). For example, a communication network can be visualized as a node-link diagram in either way, and each choice may influence the interaction concepts.

4.3.2 Interaction

Interaction methods are of central importance in visual analytics.

Operation Method — We classify the approaches based on their interaction method according to the classification developed by Yi et al. [93], namely *Select*, *Explore*, *Reconfigure*, *Encode*, *Abstract/Elaborate*, *Filter*, and *Connect*. Some are extremely common, while others like *encode* depend on the capabilities.

Table 2. **Communication analysis system classification** summarizing the different properties of the approaches. The classification criteria are formed on a subset based on the conceptual framework we developed in Section 4 and which is shown in Figure 3. The selected approaches and most categories are also available on a dedicated website at communication-analysis.dbvis.de (for details, see Section 5.2).

Generic Properties: Approach supports ● a specific property or not ○, or support is only very basic or limited ◐ (e.g., show as associated entry without analysis).
†: Indicates approaches which are commercial systems or widely used in the industry.

The following encodings have special symbols:
Group Communication: The approach supports different types of group communication analysis, as encoded in a 3x3 matrix. The rows (source) and columns (target) specify individual, group, and nested groups, respectively. For example, ⊞ encodes that the approach supports only the analysis of communication between individuals and, additionally, of individual to groups. Groups as source are not supported, and nested groups not at all.
Latency: The approach supports either the live L analysis of communication or a delayed D analysis a posteriori.
Scalability: The approach supports analysis of tens (I), hundredth to thousands (II), or millions (IIII) of cases for analysis.
Evaluation: Case study or example E , comparison with other techniques ↔ , and expert study 👤 .
Time Dimensionality: The approach supports different time dimensionalities, as encoded in a 2x3 matrix. The basis of the knowledge is encoded in the rows (either past or present), and the prediction columns specify if knowledge about the past, present, or future is inferred. For example, ⊞ encodes that the approach only uses past knowledge to learn about the past, e.g., to find a historic pattern. The encoding ⊞ instead indicates the approach uses past and present data to predict future developments.
Predictive Power: The approach's predictive power encodes in a 2x3 matrix the type of output (explanation or transition, as rows) in relation to the time dimensionalities past, present, and future, as columns. For example, ⊞ encodes an explanation about the past, while ⊞ can also give explanations about the future. Instead, the encoding ⊞ indicates that the approach supports both an explanation but also a transition function to explain past as well as classify present and predict future developments.

			Data Type					Coding		Expression		Parties	Power Rel.	Measur. Problem	Methodologies				Modality			Analysis	Latency	Scalability		Data-Mapping
	Approach	Reference	Text	Audio	Image	Meta (Network)	Meta (Time-Series)	Digital	Analogical	Explicit	Implicit	Matrix	Support?	Support?	Representation	Confirmatory	Exploratory	Predictive	Content	Network	Meta-Data	Online	L./D.	Ingress	Analysis	Abstracted
Node-link	Pajek†	[4]	○	○	○	●	○	●	○	●	○	⊞	○	○	●	●	●	○	○	●	◐	○	D	IIII	II	○
	Gephi†	[3]	○	○	○	●	○	●	○	●	○	⊞	○	○	●	●	●	○	○	●	◐	●	D	IIII	II	◐
	SaNDVis	[71]	●	○	○	●	○	●	○	●	○	⊞	○	○	●	●	●	○	●	●	○	○	D	IIII	II	○
	Ghani et al.	[32]	●	○	○	●	○	●	○	●	○	⊞	○	○	●	●	●	○	◐	●	●	●	D	II	II	○
	Elzen et al.	[88]	◐	○	○	●	○	●	○	●	○	⊞	○	○	●	●	●	○	◐	●	●	●	D	II	II	○
	NEREx	[17]	●	○	○	○	○	●	○	●	●	⊞	○	○	●	●	●	○	●	○	●	○	D	IIII	II	○
	i2 Analyst's NB†	[45]	●	◐	◐	●	◐	●	○	●	○	⊞	○	○	●	●	●	●	●	●	●	●	D	IIII	I	●
	Palantir Gotham†	[68]	●	◐	◐	●	◐	●	○	●	○	⊞	○	○	●	●	●	●	●	●	●	●+	L D	IIII	I	●
	DataWalk†	[15]	●	◐	◐	●	◐	●	○	●	○	⊞	○	○	●	●	●	●	●	●	●	●	D	IIII	I	●
	Nuix D./I.†	[66]	●	◐	◐	●	◐	●	○	●	○	⊞	○	○	●	●	●	●	●	●	●	●	L D	IIII	II	●
Matrix	Reccurence Plots	[1]	●	○	○	○	○	●	◐	●	○	⊞	○	○	●	●	●	○	●	○	○	○	D	II	II	○
	GestaltMatrix	[7]	○	○	○	●	○	●	○	●	○	⊞	○	○	●	●	○	○	○	●	○	○	D	II	I	○
	MatrixWave	[96]	○	○	○	○	●	●	○	●	○	⊞	○	○	●	●	●	○	○	○	●	○	D	II	I	○
	SmallMultiPiles	[2]	○	○	○	●	○	●	○	●	○	⊞	○	○	●	●	●	○	○	●	○	○	D	II	II	○
	HyperMatrix	[20]	●	○	○	●	○	●	○	●	○	⊞	◐	○	●	●	●	●	●	●	◐	○	D	II	I	○
Timline	Themail	[89]	●	○	○	○	●	●	○	●	○	⊞	○	○	●	●	●	○	●	○	◐	○	D	IIII	II	○
	TextFlow	[14]	●	○	○	○	●	●	◐	●	○	⊞	○	○	●	●	●	○	●	○	●	○	D	II	II	○
	CloudLines	[51]	●	○	○	○	●	●	○	●	○	⊞	○	○	●	●	●	○	◐	○	●	○	D	IIII	I	○
	DecisionFlow	[33]	◐	○	○	○	●	●	○	●	○	⊞	○	○	●	●	●	○	◐	○	●	○	D	II	II	○
	OpinionFlow	[92]	●	○	○	○	◐	●	◐	●	●	⊞	○	○	●	●	●	●	●	○	◐	○	D	IIII	I	○
	Han et al.	[37]	○	○	○	○	●	●	◐	◐	●	⊞	○	○	●	●	●	●	◐	○	◐	○	D	IIII	II	○
	Liu et al.	[57]	●	○	○	○	●	●	○	●	○	⊞	○	○	●	●	●	○	●	○	●	○	D	II	I	○
	ThreadReconst.	[18]	●	○	○	○	●	●	○	●	●	⊞	○	○	●	●	●	○	●	○	●	○	D	IIII	II	○
	T-Cal	[27]	●	○	○	○	●	●	◐	●	○	⊞	○	○	●	●	●	○	●	○	●	○	D	IIII	II	○
	netflower	[85]	●	○	○	●	●	●	○	●	○	⊞	○	○	●	●	●	○	◐	○	●	○	D	II	I	○
	WeSeer	[55]	●	○	◐	◐	●	●	◐	●	○	⊞	○	○	●	●	●	●	●	◐	●	○	D	IIII	II	○
Multi-Paradigm	MatrixExplorer	[38]	●	○	○	●	○	●	○	●	○	⊞	○	○	●	●	●	○	○	●	●	○	D	II	I	○
	NodeTrix	[39]	●	○	○	●	○	●	○	●	○	⊞	○	○	●	●	●	○	○	●	●	○	D	II	I	○
	Hadlak et al.	[36]	○	○	○	●	●	●	○	●	○	⊞	○	○	●	●	●	○	○	●	●	○	D	II	I	○
	Overview	[8]	●	○	○	○	●	●	○	●	○	⊞	○	○	●	●	●	○	●	○	●	○	D	IIII	II	○
	ConVis	[42]	●	○	○	●	●	●	●	●	●	⊞	●	○	●	●	●	○	●	●	●	○	D	II	II	○
	TargetVue	[10]	●	○	◐	●	◐	●	◐	●	●	⊞	○	○	●	●	●	○	●	●	●	○	D	II	II	○
	MediaDiscourse	[59]	●	○	○	○	●	●	○	●	○	⊞	○	○	●	●	●	○	●	○	●	○	D	II	I	○
	VisOHC	[53]	●	○	○	○	●	●	◐	●	●	⊞	○	○	●	●	●	○	●	○	●	○	D	II	I	○
	iForum	[28]	●	○	○	●	◐	●	◐	●	●	⊞	○	○	●	●	●	○	●	●	●	○	D	II	II	○
	VASSL	[49]	●	○	◐	◐	●	●	○	●	○	⊞	○	○	●	●	●	○	●	○	●	●+	D	II	I	○
	CommAID	[22]	●	○	◐	●	●	●	○	●	●	⊞	◐	◐	●	●	●	○	●	●	●	○	D	II	I	◐
Other	Zhao et al.	[95]	●	○	○	○	◐	●	◐	●	○	⊞	○	○	●	●	●	○	●	○	●	○	D	II	I	○
	Whisper	[9]	●	○	◐	●	●	●	◐	●	○	⊞	●	○	●	●	●	○	●	●	●	○	D	IIII	I	○
	ConToVi	[16]	●	○	○	◐	○	●	○	●	●	⊞	●	○	●	●	●	○	●	○	○	○	D	IIII	II	○
	Beagle	[50]	●	○	○	●	◐	●	○	●	●	⊞	○	○	●	●	●	○	●	●	●	○	D	IIII	II	○

	Pane			Operation M.							Manipulation		Goal		Strategy			Explanation			Transition Func.		Factors			Evaluation	Time Dim.	Predictive Power
Approach	2D	3D	S3D	Select	Explore	Reconfigure	Encode	Abstract	Filter	Connect	Direct	Indirect	Model Tuning	Data Tuning	Active Learning	Iterative	Progressive	Numerical	Textual	Graphical	Machine Model	Mental Model	Confidence	Trust	Privacy	Types	Matrix	Matrix
Pajek†	●	●	○	●	●	●	◐	●	●	●	●	●	○	●	○	●	●	●	○	●	○	●	◐	◐	○	E	⊞	⊞
Gephi†	●	◐	○	●	●	●	◐	●	●	●	●	●	○	●	○	●	●	●	○	●	○	●	◐	◐	○	E	⊞	⊞
SaNDVis	●	○	○	●	●	●	○	●	●	●	●	●	○	●	○	●	●	○	○	●	○	●	○	○	○	E 👤	⊞	⊞
Ghani et al.	●	○	○	●	●	●	○	●	●	●	●	●	○	●	○	●	●	○	○	●	○	●	○	○	○	👤	⊞	⊞
Elzen et al.	●	○	○	●	●	●	○	●	●	●	●	●	○	●	○	●	●	●	○	●	○	●	○	○	○	E	⊞	⊞
NEREx	●	○	○	●	●	●	●	●	●	●	●	●	○	●	●	●	●	○	●	●	○	●	○	○	○	E 👤	⊞	⊞
i2 Analyst's NB†	●	○	○	●	●	●	●	●	●	●	●	●	○	●	●	●	●	●	●	●	●	●	◐	◐	◐	-	⊞	⊞
Palantir Gotham†	●	○	○	●	●	●	●	●	●	●	●	●	○	●	●	●	●	●	●	●	●	●	◐	◐	◐	-	⊞	⊞
DataWalk†	●	○	○	●	●	●	●	●	●	●	●	●	○	●	●	●	●	●	●	●	○	●	◐	◐	◐	-	⊞	⊞
Nuix D./I.†	●	○	○	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	◐	◐	◐	-	⊞	⊞
Reccurence Plots	●	○	○	●	●	●	○	○	●	●	○	●	○	○	○	○	●	○	●	●	○	●	○	○	○	E	⊞	⊞
GestaltMatrix	●	○	○	○	●	○	○	○	○	○	○	○	○	○	○	●	○	○	○	●	○	●	○	○	○	E 👤	⊞	⊞
MatrixWave	●	○	○	●	●	●	○	○	●	●	●	●	○	○	○	●	●	○	○	●	○	●	○	○	○	👤	⊞	⊞
SmallMultiPiles	●	○	○	●	●	●	○	●	●	●	●	●	○	●	○	●	●	○	○	●	○	●	○	○	○	👤	⊞	⊞
HyperMatrix	●	○	○	●	●	●	◐	●	●	●	●	●	●	●	●	●	●	●	○	●	●	●	●	●	○	E ↔ 👤	⊞	⊞
Themail	●	○	○	●	●	●	○	●	●	●	◐	●	○	●	○	●	●	○	●	●	○	●	○	◐	○	E 👤	⊞	⊞
TextFlow	●	○	○	●	●	●	●	●	●	●	○	●	○	●	○	●	●	○	●	●	○	●	○	○	○	E	⊞	⊞
CloudLines	●	○	○	●	●	○	○	●	●	●	○	●	○	●	○	●	○	○	○	●	○	●	○	○	○	E	⊞	⊞
DecisionFlow	●	○	○	●	●	●	○	●	●	●	●	●	○	●	○	●	●	●	○	●	○	●	○	○	○	👤	⊞	⊞
OpinionFlow	●	○	○	●	●	○	○	●	●	●	●	●	○	●	○	●	●	○	●	●	●	●	○	○	○	E 👤	⊞	⊞
Han et al.	●	○	○	●	●	●	○	●	●	●	○	●	○	●	○	●	●	○	○	●	○	●	○	○	○	E	⊞	⊞
Liu et al.	●	○	○	●	●	○	○	●	●	●	○	●	○	●	○	●	●	○	○	●	○	●	○	○	○	E	⊞	⊞
ThreadReconst.	●	○	○	●	●	●	●	●	●	●	●	●	●	●	●	●	●	○	●	●	●	●	○	●	○	E 👤	⊞	⊞
T-Cal	●	○	○	●	●	●	○	●	●	●	○	●	○	●	○	●	●	○	●	●	○	●	○	○	○	E 👤	⊞	⊞
netflower	●	○	○	●	●	●	○	●	●	●	●	●	○	●	○	●	●	●	●	●	○	●	○	○	○	👤	⊞	⊞
WeSeer	●	○	○	●	●	●	●	●	●	●	●	●	○	●	○	●	●	●	●	●	○	●	○	○	○	E 👤	⊞	⊞
MatrixExplorer	●	○	○	●	●	●	●	●	●	●	●	●	○	●	○	●	●	●	○	●	○	●	●	○	○	👤	⊞	⊞
NodeTrix	●	○	○	●	●	●	●	●	●	●	●	●	○	●	○	●	●	●	○	●	○	●	●	○	○	E	⊞	⊞
Hadlak et al.	●	○	○	●	●	●	●	●	●	●	◐	●	○	●	○	●	●	●	○	●	○	●	○	○	○	E	⊞	⊞
Overview	●	○	○	●	●	●	●	●	●	●	●	●	○	●	○	●	●	○	●	●	○	●	○	◐	○	E 👤	⊞	⊞
ConVis	●	○	○	●	●	●	●	●	●	●	●	●	○	●	○	●	●	●	●	●	○	●	○	○	○	👤	⊞	⊞
TargetVue	●	○	○	●	●	●	●	●	●	●	●	●	○	●	○	●	●	●	●	●	○	●	○	○	○	E 👤	⊞	⊞
MediaDiscourse	●	○	○	●	●	●	●	●	●	●	●	●	○	●	○	●	●	●	●	●	○	●	○	○	○	E 👤	⊞	⊞
VisOHC	●	○	○	●	●	●	●	●	●	●	●	●	●	●	○	●	●	●	●	●	○	●	●	○	○	👤	⊞	⊞
iForum	●	○	○	●	●	●	●	●	●	●	●	●	○	●	○	●	●	●	●	●	○	●	○	○	○	E 👤	⊞	⊞
VASSL	●	○	○	●	●	●	●	●	●	●	●	●	○	●	○	●	●	●	○	●	○	●	◐	○	○	E 👤	⊞	⊞
CommAID	●	○	○	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	○	●	●	●	●	●	○	E 👤	⊞	⊞
Zhao et al.	●	○	○	●	●	●	●	●	●	●	●	●	○	●	○	●	●	○	●	●	○	●	●	●	○	E 👤	⊞	⊞
Whisper	●	○	○	●	●	●	○	●	●	●	●	●	○	●	○	●	●	●	●	●	○	●	○	○	○	E 👤	⊞	⊞
ConToVi	●	○	○	●	●	●	○	●	●	●	●	●	○	●	○	●	●	○	●	●	○	●	○	○	○	E 👤	⊞	⊞
Beagle	●	○	○	●	●	○	○	●	●	●	●	●	○	●	○	●	●	●	●	●	○	●	○	○	○	E 👤	⊞	⊞

Manipulation — The manipulation [48] of the elements can be either *direct*, for example, when interacting with data or visual objects. Alternatively, it can be *indirect*, for example, when modifying parameters. Most approaches support both.

4.3.3 Refinement

In addition to the interaction concept, other discerning factors are the particularities of the refinement, for which we differentiate [74] between the *goal* and the *strategy* to achieve it.

Goal — Two primary goals can be differentiated [84]: is the goal to tune an underlying *model* (e.g., for predicting communication behavior [20]) or the *data* (e.g., to select a fitting representation [57])?

Strategy — The basic refinement strategy [74] might vary greatly with the approach: does it follow an *iterative* (e.g., improving the results through continuous interactions [16]) or *progressive* (e.g., incrementally discovering events [51]) strategy?

4.4 Knowledge Generation

Knowledge, generated and learned, is the ultimate analysis goal. We propose three subcategories: *output* to conceptualize the direct outcome, *knowledge gain* to cover the power of the outcome, and *verification approach* to consider implications and evaluations.

4.4.1 Output

Based on the classification of Spinner et al. [84], we propose two distinct categories for the learned knowledge type: An *explanation* can consist, for example, of numerical (e.g., graph algorithms [3]), textual (e.g., presented text [85]), or graphical representations (e.g., visual network representations [7]). It represents knowledge but in a factual representation that is not easily transferable and can be regarded as a (final) result of the existing data and is intended for humans. In contrast to this, another type of result can be a *transition function*, which is closer to an actual model, one example being the analyst's mental model. Another type is a machine model, for example, a trained, applicable classifier (e.g., diffusion model [92] or neural communication prediction model [20]) that encapsulates learned knowledge.

4.4.2 Knowledge Gain

As a final step in the learning process, the question arises which knowledge [34] is actually gained and how powerful the process is.

Time Dimensionality — The time dimensionality describes the relationship between data and knowledge generation. A 2x3 matrix shows the possible combinations of data basis and prediction type, each with the entries past, present, and future. For example (like [36]), a system can use past data and predict past data, for example for a search. Then, (like [55]) would be an article analysis and prediction system which has been trained on past data to analyze a text, either an existing one or one on the fly in the present and future. Another example for a future prediction is a model that forecasts communication activity based on past events. Note that by causality, the future is excluded.

Predictive Power — A second important consideration describes the predictive power of the knowledge generated, which is represented as a 2x3 matrix, where the result (explanation or transition function) and the time are combined. For example , can a system explain (i.e., show) past events (virtually all systems)? Or is it able to provide factual information for future events (e.g., information cascade prediction [55], internal model is inaccessible). A more powerful example is a controllable model which can explain and predict (e.g., opinion diffusion [92]).

Domain-specific Aspects — Last but not least, depending on different analysis tasks, more specific options might be of interest. As discussed above, these are out of scope here; however, we could imagine this as future work (see Section 5).

4.4.3 Verification Approach

The presentation, as well as automated analysis of knowledge, raises a plethora of ethical as well as technical questions.

Factors — Visual analytics is very well suited to address important factors like confidence, trust, and privacy and consider aspects like fairness, accountability, and privacy [13, 21]. For example, probability scores could be used to estimate the results' confidence stemming from automatic processes (e.g., visually indicating confidence scores [20]) and visualize it to the expert. Other examples include analysis log files, integrity protection, traceability, and verify-ability, as well as a provenance history. The lack of such certainty measures might exclude systems from sensitive areas. While essential, as shown by Correll et al. [13], many approaches are oblivious. For a more detailed discussion on the human factors in communication analysis, the ethical dilemmas, and design considerations for communication analysis systems, we refer to our companion work [21].

Evaluation — To evaluate approaches, the author's own evaluation method can be considered. Several options are possible: Either a convincing *example* E or a case study (e.g., describing a potential application [39]). A second option is a *comparison* ↔ with similar, existing approaches through feature comparison (e.g., comparing the main features with related work [20]). A third option would be a *qualitative* interview (e.g., interviewing eight domain experts [22]) or a *quantitative* user study .

5 Discussion and Future Work

This section discusses the main findings and lessons learned before reflecting on the difficulties in creating a conceptual framework for communication analysis. In particular, we discuss the potential implications and opportunities for future research while highlighting the shortcomings in the current formalization.

We have defined four main dimensions containing over fifty different properties, providing a conceptual framework for interactive communication analysis systems employing visual analytics principles.

We imagine several **applications for this framework**: Its primary goal is to provide a state-of-the-art overview of the current techniques employed, laying the foundations for a longer and more detailed survey in the field. Further, it aims at structuring the research field and providing a common language for the community while supporting comparison between approaches for practitioners and developers alike. Last but not least, it identifies gaps and research opportunities, which we discuss in the following section.

5.1 Survey Findings and Research Opportunities

Visual analytics is especially suited to support semi-automatic communication analysis [40, 47]. The complexity, multi-modality, and ambiguity of communication make it an ideal companion to interactively combine domain knowledge from experts and computing power. Concepts like interactive learning allow refining models, while uncertainty awareness enables automatic judging of results, fostering user trust and possibly identifying bias.

Findings — To apply the framework, we have taken the 41 selected approaches (see Section 3) and coded them according to our conceptual framework. Based on the results in Table 2, we can discover several interesting aspects. For once, for the **data type**, while the analysis of text data seems mostly universal across representation methods, this is not the case for the other data types. Somewhat unsurprisingly, when network data is included, the visualization is often node-link-based or multi-paradigm, while for time-series, it is either timeline or multi-paradigm. Given the scope of the survey, the lack of audio and image is not surprising. Given that all the approaches belong to the category of visual analytics, it is also unsurprising that virtually all support representative, confirmatory and exploratory analysis as their **analysis methodology**, their **operation methods** covered most options, and their **explanation** is at least always graphical.

More interesting, however, are the **differences** and the **research opportunities** we can conclude from their discrepancies, which we highlight in the following for each category (see Figure 1); those of particular relevance are highlighted with a star: ★.

★ I.1 Analysis of the Meaning and Analogical Code
Only a subset of approaches analyzes the implicit meaning of the communication (e.g., [17, 28, 42]). However, almost none analyze

the **analogical code** of the communication.
Implication: The analogical code can contain important cues which might support the analysis of the content and provide supportive information about the relationships inside the network, which makes it especially relevant to consider [90]. Leveraging it can lead to a richer and more-complete analysis, while it can support resolving contradictions and ambiguities [18].

I.2 Include Power Relations
Again, almost no approach considers the power relations between the participants, which can similarly influence the communication semantics, meaning, and modalities.
Implication: Power relations between participants [79, 90] might influence content aspects like choice of words, formality, use of irony, meaning, or meta-data aspects like dynamics [81], timestamps, or message count. Results can be used and considered in context with the content analysis.

I.3 Dynamic Analysis
While some might consider this a technical problem, the development of systems that support the dynamic analysis of communication data and batch/stream approaches also sets considerable hurdles to established analysis and visualization methods, which makes it an interesting academic research problem.
Implication: Exploring how new data and updated results can be integrated [9], how fluctuating analysis can be stabilized, and how changed predictions [57] are communicated offers more effective ways for visual communication.

I.4 Research the Measurement Problem
The mitigation of the measurement influence is rarely explored.
Implication: Being aware of the *measurement problem* and explore mitigations [76] can strengthen user trust, while avoiding missed or erroneous results (e.g., due to codewords) [90].

II.1 Multi-Environment Inclusion
Many approaches lack support for data mapping and multiple data sources (like, e.g., [45]), requiring preprocessed data.
Implication: Automating the merging of heterogeneous data sources [48] with few or no user input reduces the amount of manual preprocessing or knowledge transfer required, make leveraging multiple data sources simultaneously less complicated [28], while exploring optimal interface strategies.

★ II.2 Analyze Group Communication
Only a few approaches support the analysis of group communication (e.g., [16, 24]), and almost none support nested groups.
Implication: New and more detailed knowledge can be drawn how groups operate [24] and information diffuses [54] within, in particular because much communication actually happens inside or between groups, which can involve specific particularities [5].

III.1 Visually Interactive Model Analysis
Virtually all approaches use the 2D pane for visualizations and many automations focus on filtering instead of model tuning.
Implication: Leveraging visual data analysis techniques [40, 48, 93] and explore unused approaches like VR for improving the analysis process [12, 32], focusing on the model [84] instead of only data selection may allow for the higher level conclusions, supporting the knowledge generation [75].

★ IV.1 Model / Transfer Function / Knowledge Gain
Few approaches contain an actual, powerful machine models [20, 66] to analyze communication.
Implication: Using such models can potentially support the analysis [12, 34], through measures as active learning [10, 65], intelligent filtering [22], or confidence-based predictions [74]. Transfer Functions allow for a more universal machine learning, applying knowledge to new problems, increasing the predictive power. This reduces manual work while increasing analytical capabilities.

★ IV.2 Confidence, Trust, and Privacy
These factors are insufficiently considered the majority of approaches, leading to a black-box analysis. Instead, one could include confidence estimates (e.g., [20]), logs, provenance (e.g., [22]), data minimization, or other concepts.
Implication: Several applications have strong requirements for confidence and trust [74], provenance [22, 84], and privacy [21]. Exploring how these can be fulfilled [21] without limiting the analysis can replace manual analysis by automated system.

IV.3 Guidelines and Quantitative User Studies
While several approaches include case studies and (qualitative) expert interviews, almost none make actual comparisons with related approaches or conduct quantitative user studies.
Implication: Case studies and qualitative expert interviews are not always comparable or conducted to the same standards [8]. While we do not doubt the systems work well as advertised, for reproducible comparisons between approaches, quantitative studies are required and evaluations along design guidelines [12], providing a more objective overview.

★ O.1 Holistic Approaches
Only few approaches work towards a holistic analysis by considering multiple analysis aspects in context, covering all modalities.
Implication: A holistic perspective [88] can increase the analytical capabilities considerably [8, 22], supporting cross-matches beyond analysis boundaries [21], while reducing mental load and manual work required.

★ O.2 Context / Analysis Reference Window
Similar to the holistic analysis, a specific focus on the communication context in reference to each other should be explored further for both inter- and intra-modality analysis.
Implication: Few approaches consider other modalities or external factors to explain particularities. For example, a break in a communication sequence might appear as a gap, but when combined with location information (e.g., same building) might indicate that the participants might have met for lunch and continued their conversation offline. In summary, the correct interpretation of communication is extremely context-dependent [10, 12], with different applicability of analysis methods. Analyzing references and clues can improve the determination of the highly variable context [64] for choosing appropriate analysis methodologies.

Implications — All the previously described opportunities offer potential improvements for a more complete analysis of communication. Fusing together multiple methods can lead to a richer and more complete analysis, potentially resolving contradictions and ambiguities. Some opportunities might primarily support existing analysis steps (e.g., I.2) as a precursor, while others provide new areas in itself (e.g., II.2, O.1). While the relevance of aspects might differ in any given analysis, our framework identifies and describes areas that users can consider and potentially leverage, depending on their analytical needs.

5.2 Limitations and Future Work

While our conceptual framework offers exciting research opportunities for more powerful communication analysis systems and the application of visual analytics concepts, the area itself is under active research with several limitations and unsolved challenges.

One problem with the described taxonomy is the basis it is designed upon (**completeness**). We are confident we included the essential results from classical communication research, technical considerations, and many aspects of the surveyed approaches. However, one issue is the completeness of the surveyed approaches. A significant problem in this research area is that relevant approaches are rarely labeled as belonging to communication analysis and are not easily definable by a set of keywords. We initially thought about compiling a list of domain-specific keywords to select papers by, for example, social network analysis, sentiment analysis, e-mail analysis, etc. However, we found it highly likely that such a selection would be highly biased by our knowledge of relevant topics, which is why we decided to go for a more extensive drag-net search to include previously unknown areas. However, while we took great care when defining the seed search and then selecting the approaches, it is inadvertently likely we missed individual ones, not least by restricting the target journals and conferences. Also, it could happen that a few approaches fell through our automatic or manual search pattern (see Section 3), while still being applicable to communication analysis. Some missing approaches may contain novel aspects which might widen the framework and introduce new elements or pose problems for existing ones. However, due to the restrictions discussed in Section 3, we do not claim overall completeness. The survey forms one of three pillars for our goal of constructing the framework, and together with the other two, we are confident the majority of cases can be described within our framework. Nevertheless, to address the issue of missing approaches, we created an accompanying **survey website** available at https://communication-analysis.dbvis.de, which lists the approaches we considered and also allows readers to submit methods missing methods.

Another possible limitation concerns the **orthogonality** of the framework itself. Due to the complexity and heterogeneity of the area, it contains some overlaps. As there is a need to balance the trade-off's accuracy, usability, and relevance, we think it is challenging to create a wholly consistent yet easily usable taxonomy. The choices we made for selecting the categories are often based on the literature and justified when required. However, given sparse taxonomy and non-standardized vocabulary, some groupings and namings could arguably have been chosen differently with the same validity. To advance research in this area, however, we decided to propose our framework as a first possible draft and one step towards a universally accepted framework. We, therefore, invite the research community to give feedback to stimulate the scientific discussion and extend upon the framework in further iterations, which can be enhanced further by more input from diverse research communities. As part of this process, the individual, multi-faceted aspects can be formalized in more detail.

Another aspect is the extension of the framework to **non-human communication**. Several aspects of the framework could similarity be applied to communication in general, for example, machine to machine. Indeed, nothing in the framework is specifically tailored to a human communicator, as all the properties could also apply to other types of communication. However, human communication is often more nuanced than machine communication, making parts of the framework less relevant, while other features (e.g., structuring, exchange content scope) might be missing so far.

6 Conclusion

In the last decades, communication analysis has experienced a shift away from manual analysis to computer-aided or even highly automated approaches. However, to the same extent as automation levels increased, the analysis itself has often become more specialized, moving away from an overarching exploration. This trend is in contradiction to traditional communication research, which stresses the importance of a holistic approach to capture the full meaning and context of communication. As a result, many modern digital communication analysis systems are highly adapted to a narrow range of tasks, often either in the area of content *or* in network analysis. While this might be perfectly sufficient and suitable for their intended use, such an isolated analysis can sometimes lead to a less effective exploration and lead to incomplete or biased results. Using separate approaches requires more manual work, often complicates analysis tasks, can introduce domain discontinuities, and increase the struggle domain experts face when trying to integrate their domain knowledge. Further, an isolated analysis may not be sufficient to capture the full available information and can make the automatic as well as the manual detection of cross-matches more difficult. The development of more holistic and advanced approaches for automated communication analysis systems is hindered by the lack of a clear framework and the absence of a common language that combines both technical aspects as well as results from traditional communication research.

We address this challenge by developing and formalizing a design space for digital communication analysis systems based on the existing tool landscape and communication research while making a case for how visual analytics principles can be employed for a more holistic approach. By systematically discussing and structuring the different analysis areas and aspects of the design space, we arrive at a conceptual framework to provide an overview and assess the maturity of communication analysis systems. As part of an initial survey, we have also categorized a large set of existing approaches using our framework.

By bridging the gap in the formalization of digital communication analysis systems by describing a design space for communication analysis, we aim to provide researchers with a common language, provide guidelines for building and to assess the maturity of such approaches, as well as point out gaps in the literature which offer exciting research opportunities. The results of this work are widely applicable in a variety of domains that are concerned with communication analysis like civil security, the digital humanities, or business intelligence, both from a theoretical point of view as well as for the development of more powerful communication analysis systems.

Acknowledgments

The authors acknowledge the financial support by the Federal Ministry of Education and Research of Germany (BMBF) in the framework of PEGASUS under the program "Forschung für die zivile Sicherheit 2018 - 2023" and its announcement "Zivile Sicherheit - Schutz vor organisierter Kriminalität II".

References

[1] D. Angus, A. Smith, and J. Wiles. Conceptual Recurrence Plots: Revealing Patterns in Human Discourse. *IEEE Transactions on Visualization and Computer Graphics*, 18(6):988–997, 2012. doi: 10.1109/TVCG.2011.100

[2] B. Bach, N. Henry-Riche, T. Dwyer, T. Madhyastha, J.-D. Fekete, and T. Grabowski. Small MultiPiles: Piling Time to Explore Temporal Patterns in Dynamic Networks. *Computer Graphics Forum*, 34(3):31–40, 2015. doi: 10.1111/cgf.12615

[3] M. Bastian, S. Heymann, and M. Jacomy. Gephi: An Open Source Software for Exploring and Manipulating Networks. In *Proc. of the Int. AAAI Conference on Weblogs and Social Media (ICWSM)*, 2009.

[4] V. Batagelj and A. Mrvar. Pajek - Program for Large Network Analysis. *Connections*, 21(2):47–57, 1998.

[5] A. Bavelas. Communication Patterns in Task-oriented Groups. *The Journal of the Acoustical Society of America*, 22(6):725–730, 1950.

[6] S. P. Borgatti, A. Mehra, D. J. Brass, and G. Labianca. Network Analysis in the Social Sciences. *Science*, 323(5916):892–895, 2009. doi: 10.1126/science.1165821

[7] U. Brandes and B. Nick. Asymmetric Relations in Longitudinal Social Networks. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2283–2290, 2011. doi: 10.1109/TVCG.2011.169

[8] M. Brehmer, S. Ingram, J. Stray, and T. Munzner. Overview: The Design, Adoption, and Analysis of a Visual Document Mining Tool for Investigative Journalists. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2271–2280, 2014. doi: 10.1109/TVCG.2014.2346431

[9] N. Cao, Y.-R. Lin, X. Sun, D. Lazer, S. Liu, and H. Qu. Whisper: Tracing the Spatiotemporal Process of Information Diffusion in Real Time. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2649–2658, 2012. doi: 10.1109/TVCG.2012.291

[10] N. Cao, C. Shi, S. Lin, J. Lu, Y.-R. Lin, and C.-Y. Lin. TargetVue: Visual Analysis of Anomalous User Behaviors in Online Communication Systems. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):280–289, 2016. doi: 10.1109/TVCG.2015.2467196

[11] S. K. Card, J. D. Mackinlay, and B. Shneiderman. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann Series in Interactive Technologies. Morgan Kaufmann, San Francisco, CA, 1999.

[12] M. Conlen, S. Stalla, C. Jin, M. Hendrie, H. Mushkin, S. Lombeyda, and S. Davidoff. Towards Design Principles for Visual Analytics in Operations Contexts. In R. Mandryk, M. Hancock, M. Perry, and A. Cox, eds., *Proc. of the 36th CHI Conference on Human Factors in Computing Systems*, pp. 1–7. ACM, 2018. doi: 10.1145/3173574.3173712

[13] M. Correll. Ethical Dimensions of Visualization Research. In S. Brewster, G. Fitzpatrick, A. Cox, and V. Kostakos, eds., *Proceedings of the 2019 CHI*

Conference on Human Factors in Computing Systems, pp. 1–13. ACM, New York, NY, USA, 2019. doi: 10.1145/3290605.3300418
[14] W. Cui, S. Liu, L. Tan, C. Shi, Y. Song, Z. J. Gao, X. Tong, and H. Qu. TextFlow: Towards Better Understanding of Evolving Topics in Text. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2412–2421, 2011. doi: 10.1109/TVCG.2011.239
[15] DataWalk Inc. DataWalk, 2020.
[16] M. El-Assady, V. Gold, C. Acevedo, C. Collins, and D. A. Keim. ConToVi: Multi-Party Conversation Exploration using Topic-Space Views. *Computer Graphics Forum*, 35:431–440, 2016. doi: 10.1111/cgf.12919
[17] M. El-Assady, R. Sevastjanova, B. Gipp, D. A. Keim, and C. Collins. NEREx: Named-Entity Relationship Exploration in Multi-Party Conversations. *Computer Graphics Forum*, 36:213–225, 2017. doi: 10.1111/cgf.13181
[18] M. El-Assady, R. Sevastjanova, D. A. Keim, and C. Collins. ThreadReconstructor: Modeling Reply-Chains to Untangle Conversational Text through Visual Analytics. *Computer Graphics Forum*, 37(3):351–365, 2018. doi: 10.1111/cgf.13425
[19] W. Fan and M. D. Gordon. The power of social media analytics. *Communications of the ACM*, 57(6):74–81, 2014. doi: 10.1145/2602574
[20] M. T. Fischer, D. Arya, D. Streeb, D. Seebacher, D. A. Keim, and M. Worring. Visual Analytics for Temporal Hypergraph Model Exploration. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):550–560, 2020. doi: 10.1109/TVCG.2020.3030408
[21] M. T. Fischer, S. D. Hirsbrunner, W. Jentner, M. Miller, D. A. Keim, and P. Helm. Promoting Ethical Awareness in Communication Analysis: Investigating Potentials and Limits of Visual Analytics for Intelligence Applications, 2022. doi: 10.48550/arXiv.2203.09859
[22] M. T. Fischer, D. Seebacher, R. Sevastjanova, D. A. Keim, and M. El-Assady. CommAID: Visual Analytics for Communication Analysis through Interactive Dynamics Modeling. *Computer Graphics Forum*, 40(3):25–36, 2021. doi: 10.1111/cgf.14286
[23] S. Flensburg and S. S. Lai. Mapping Digital Communication Systems: Infrastructures, Markets, and Policies as Regulatory Forces. *Media, Culture & Society*, 42(5):692–710, 2020. doi: 10.1177/0163443719876533
[24] P. W. Foltz and M. J. Martin. Automated Communication Analysis of Teams. *Team Effectiveness in Complex Organizations*, 2008.
[25] A. Forghani, C. Neustaedter, and T. Schiphorst. Investigating the communication patterns of distance-separated grandparents and grandchildren. In W. E. Mackay, S. Brewster, and S. Bødker, eds., *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, p. 67. ACM, 2013. doi: 10.1145/2468356.2468370
[26] S. Fu, Y. Wang, Y. Yang, Q. Bi, F. Guo, and H. Qu. VisForum: A Visual Analysis System for Exploring User Groups in Online Forums. *ACM Transactions on Interactive Intelligent Systems*, 8(1):1–21, 2018. doi: 10.1145/3162075
[27] S. Fu, J. Zhao, H. F. Cheng, H. Zhu, and J. Marlow. T-Cal: Understanding Team Conversation Data with Calendar-based Visualization. In R. Mandryk, M. Hancock, M. Perry, and A. Cox, eds., *Proc. of the 36th CHI Conference on Human Factors in Computing Systems*, pp. 1–13. ACM, 2018. doi: 10.1145/3173574.3174074
[28] S. Fu, J. Zhao, W. Cui, and H. Qu. Visual Analysis of MOOC Forums with iForum. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):201–210, 2017. doi: 10.1109/TVCG.2016.2598444
[29] X. Fu, S.-H. Hong, N. S. Nikolov, X. Shen, Y. Wu, and K. Xu. Visualization and Analysis of Email Networks. In *Proceedings of the 6th International Asia-Pacific Symposium on Visualization*, PacificVis, pp. 1–8. IEEE, 2007. doi: 10.1109/APVIS.2007.329302
[30] G. Gao and S. R. Fussell. A Kaleidoscope of Languages. In G. Mark, S. Fussell, C. Lampe, m. schraefel, J. P. Hourcade, C. Appert, and D. Wigdor, eds., *Proceedings of the 35th CHI Conference on Human Factors in Computing Systems*, pp. 760–772. ACM, New York, NY, USA, 05022017. doi: 10.1145/3025453.3025839
[31] G. Gerbner. Analysis of Communication Content. 1969.
[32] S. Ghani, B. C. Kwon, S. Lee, J. S. Yi, and N. Elmqvist. Visual Analytics for Multimodal Social Network Analysis: A Design Study with Social Scientists. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2032–2041, 2013. doi: 10.1109/TVCG.2013.223
[33] D. Gotz and H. Stavropoulos. DecisionFlow: Visual Analytics for High-Dimensional Temporal Event Sequence Data. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1783–1792, 2014. doi: 10.1109/TVCG.2014.2346682
[34] T. M. Green, W. Ribarsky, and B. Fisher. Visual analytics for complex concepts using a human cognition model. In *2008 IEEE Symposium on Visual Analytics Science and Technology*, pp. 91–98. IEEE, 2008. doi: 10.1109/VAST.2008.4677361
[35] R. Hadjidj, M. Debbabi, H. Lounis, F. Iqbal, A. Szporer, and D. Benredjem. Towards an Integrated E-mail Forensic Analysis Framework. *Digital Investigation*, 5(3-4):124–137, 2009. doi: 10.1016/j.diin.2009.01.004
[36] S. Hadlak, H. Schumann, C. H. Cap, and T. Wollenberg. Supporting the Visual Analysis of Dynamic Networks by Clustering Associated Temporal Attributes. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2267–2276, 2013. doi: 10.1109/TVCG.2013.198
[37] Y. Han, A. Rozga, N. Dimitrova, G. D. Abowd, and J. Stasko. Visual Analysis of Proximal Temporal Relationships of Social and Communicative Behaviors. *Computer Graphics Forum*, 34(3):51–60, 2015. doi: 10.1111/cgf.12617
[38] N. Henry and J.-D. Fekete. MatrixExplorer: A Dual-Representation System to Explore Social Networks. *IEEE Trans. on Visualization and Computer Graphics*, 12(5):677–684, 2006. doi: 10.1109/TVCG.2006.160
[39] N. Henry, J.-D. Fekete, and M. J. McGuffin. NodeTrix: A Hybrid Visualization of Social Networks. *IEEE Trans. on Visualization and Computer Graphics*, 13(6):1302–1309, 2007. doi: 10.1109/TVCG.2007.70582
[40] S. C. Herring. New Frontiers in Interactive Multimodal Communication. *The Routledge Handbook of Language and Digital Communication*, pp. 398–402, 2015.
[41] S. C. Herring. The Coevolution of Computer-Mediated Communication and Computer-Mediated Discourse Analysis. In P. Bou-Franch and P. Garcés-Conejos Blitvich, eds., *Analyzing Digital Discourse*, pp. 25–67. Springer, 2019. doi: 10.1007/978-3-319-92663-6-2
[42] E. Hoque and G. Carenini. ConVis: A Visual Text Analytic System for Exploring Blog Conversations. *Computer Graphics Forum*, 33(3):221–230, 2014. doi: 10.1111/cgf.12378
[43] E. Hoque and G. Carenini. ConVisIT: Interactive Topic Modeling for Exploring Asynchronous Online Conversations. In O. Brdiczka, P. Chau, G. Carenini, S. Pan, and P. O. Kristensson, eds., *Proceedings of the 20th ACM International Conference on Intelligent User Interfaces*, IUI, pp. 169–180. ACM, New York, USA, 2015. doi: 10.1145/2678025.2701370
[44] E. Hoque and G. Carenini. MultiConVis: Visual Text Analytics System for Exploring a Collection of Online Conversations. In J. Nichols, J. Mahmud, J. O'Donovan, C. Conati, and M. Zancanaro, eds., *IUI'16*, pp. 96–107. ACM, New York, USA, 2016. doi: 10.1145/2856767.2856782
[45] IBM. i2 Analyst's Notebook, 2020.
[46] D. Jagdish, A. Attarwala, and U. Fischer. Behavior assessment and visualization tool. In E. Mynatt, D. Schoner, G. Fitzpatrick, S. Hudson, K. Edwards, and T. Rodden, eds., *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, pp. 3871–3876. ACM, New York, NY, USA, 04092010. doi: 10.1145/1753846.1754071
[47] C. Jensen, S. D. Farnham, S. M. Drucker, and P. Kollock. The Effect of Communication Modality on Cooperation in Online Environments. In T. Turner and G. Szwillus, eds., *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '00*, pp. 470–477. ACM Press, New York, New York, USA, 2000. doi: 10.1145/332040.332478
[48] D. A. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon. Visual Analytics: Definition, Process, and Challenges. In A. Kerren, ed., *Information Visualization*, pp. 154–175. Springer, 2008. doi: 10.1007/978-3-540-70956-5-7
[49] M. Khayat, M. Karimzadeh, J. Zhao, and D. S. Ebert. VASSL: A Visual Analytics Toolkit for Social Spambot Labeling. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):874–883, 2020. doi: 10.1109/TVCG.2019.2934266
[50] J. Koven, C. Felix, H. Siadati, M. Jakobsson, and E. Bertini. Lessons Learned Developing a Visual Analytics Solution for Investigative Analysis of Scamming Activities. *IEEE Trans. on Visualization and Computer Graphics*, 25(1):225–234, 2019. doi: 10.1109/TVCG.2018.2865023
[51] M. Krstajić, E. Bertini, and D. A. Keim. CloudLines: Compact display of Event Episodes in Multiple Time-series. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2432–2439, 2011. doi: 10.1109/TVCG.2011.179
[52] K. Kucher and A. Kerren. Text Visualization Techniques: Taxonomy, Visual Survey, and Community Insights. In S. Liu, G. Scheuermann, S. Takahashi, and I. P. V. Symposium, eds., *2015 IEEE Pacific Visualization Symposium (PacificVis)*, pp. 117–121. IEEE, Piscataway, NJ, 2015. doi: 10.1109/PACIFICVIS.2015.7156366
[53] B. C. Kwon, S.-H. Kim, S. Lee, J. Choo, J. Huh, and J. S. Yi. VisOHC: Designing Visual Analytics for Online Health Communities. *IEEE Trans-*

actions on Visualization and Computer Graphics, 22(1):71–80, 2016. doi: 10.1109/TVCG.2015.2467555

[54] H. J. Leavitt. Some Effects of Certain Communication Patterns on Group Performance. *The J. of Abnormal and Social Psychology*, 46(1):38, 1951.

[55] Q. Li, Z. Wu, L. Yi, K. Seann, H. Qu, and X. Ma. WeSeer: Visual Analysis for Better Information Cascade Prediction of WeChat Articles. *IEEE Transactions on Visualization and Computer Graphics*, 26(2):1399–1412, 2020. doi: 10.1109/TVCG.2018.2867776

[56] F. Liu, C. Park, Y. J. Tham, T.-Y. Tsai, L. Dabbish, G. Kaufman, and A. Monroy-Hernández. Significant Otter: Understanding the Role of Biosignals in Communication. In Y. Kitamura, A. Quigley, K. Isbister, T. Igarashi, P. Bjørn, and S. Drucker, eds., *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–15. ACM, New York, NY, USA, 05062021. doi: 10.1145/3411764.3445200

[57] S. Liu, J. Yin, X. Wang, W. Cui, K. Cao, and J. Pei. Online Visual Analytics of Text Streams. *IEEE Transactions on Visualization and Computer Graphics*, 22(11):2451–2466, 2016. doi: 10.1109/TVCG.2015.2509990

[58] M. Lombard, J. Snyder-Duch, and C. C. Bracken. Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability. *Human communication research*, 28(4):587–604, 2002.

[59] Y. Lu, M. Steptoe, S. Burke, H. Wang, J.-Y. Tsai, H. Davulcu, D. Montgomery, S. R. Corman, and R. Maciejewski. Exploring Evolving Media Discourse Through Event Cueing. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):220–229, 2016. doi: 10.1109/TVCG.2015.2467991

[60] Q. Luo and D. Zhong. Using Social Network Analysis to Explain Communication Characteristics of Travel-related Electronic Word-of-Mouth on Social Networking Sites. *Tourism Management*, 46:274–282, 2015. doi: 10.1016/j.tourman.2014.07.007

[61] A. Mastrianni, L. Kulp, E. Mapelli, and A. Sarcevic. Understanding Digital Checklist Use Through Team Communication. *Extended Abstracts on Human factors in computing systems. CHI Conference*, 2020, 2020. doi: 10.1145/3334480.3382817

[62] M. McLuhan. *Understanding Media: The Extensions of Man*. Routledge, London, 1964.

[63] M. McLuhan. The Medium Is The Message. In C. D. Mortensen, ed., *Communication Theory*, pp. 390–402. Routledge, 2017. doi: 10.4324/9781315080918-31

[64] G. S. Mesch. Social Context and Communication Channels Choice Among Adolescents. *Computers in Human Behavior*, 25(1):244–251, 2009. doi: 10.1016/j.chb.2008.09.007

[65] H.-W. Ng, V. D. Nguyen, V. Vonikakis, and S. Winkler. Deep Learning for Emotion Recognition on Small Datasets using Transfer Learning. In Z. Zhang, P. Cohen, D. Bohus, R. Horaud, and H. Meng, eds., *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pp. 443–449. ACM, New York, 2015. doi: 10.1145/2818346.2830593

[66] Nuix Pty Ltd. Nuix Discover and Nuix Investigate, 2020.

[67] I. Onyeator and N. Okpara. Human Communication in a Digital Age: Perspectives on Interpersonal communication in the family. *New Media and Mass Communication*, 78:35–45, 2019.

[68] Palantir Technologies, Inc. Gotham, 2020.

[69] B. Pang and L. Lee. Opinion Mining and Sentiment Analysis. *Foundations and Trends in Inf. Retr.*, 2(1-2):1–135, 2008. doi: 10.1561/1500000011

[70] J. C. Pearson. *Human Communication*. McGraw-Hill, 4th ed., 2011.

[71] A. Perer, I. Guy, E. Uziel, I. Ronen, and M. Jacovi. The Longitudinal Use of SaNDVis: Visual social Network Analytics in the Enterprise. *IEEE Transactions on Visualization and Computer Graphics*, 19(7):1095–1108, 2013. doi: 10.1109/TVCG.2012.322

[72] R. Řehůřek and P. Sojka. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of LREC 2010 workshop: New Challenges for NLP Frameworks*, pp. 46–50. Valletta, Malta, 2010.

[73] E. M. Rogers and D. L. Kincaid. *Communication Networks: Toward a New Paradigm for Research*. Macmillan USA, 1980.

[74] D. Sacha, H. Senaratne, B. C. Kwon, G. Ellis, and D. A. Keim. The Role of Uncertainty, Awareness, and Trust in Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):240–249, 2016. doi: 10.1109/TVCG.2015.2467591

[75] D. Sacha, A. Stoffel, F. Stoffel, B. C. Kwon, G. P. Ellis, and D. A. Keim. Knowledge Generation Model for Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1604–1613, 2014. doi: 10.1109/TVCG.2014.2346481

[76] K. Sameera and P. Vishwakarma. Cybercrime: To Detect Suspected User's Chat Using Text Mining. In S. C. Satapathy and A. Joshi, eds., *Information and Communication Technology for Intelligent Systems*, vol. 106 of *Smart Innovation, Systems and Technologies*, pp. 381–390. Springer Singapore, Singapore, 2019. doi: 10.1007/978-981-13-1742-2-37

[77] A. Sarikaya, M. Gleicher, and D. A. Szafir. Design Factors for Summary Visualization in Visual Analytics. *Computer Graphics Forum*, 37(3):145–156, 2018. doi: 10.1111/cgf.13408

[78] I. R. Savage and K. W. Deutsch. A Statistical Model of the Gross Analysis of Transaction Flows. *Econometrica*, 28(3):551–572, 1960.

[79] F. Schulz von Thun. *Miteinander reden: Störungen und Klärungen - Psychologie der zwischenmenschlichen Kommunikation*. Rowohlt, 1981.

[80] C. A. Scolari. Mapping Conversations About New Media: The Theoretical Field of Digital Communication. *New Media & Society*, 11(6):943–964, 2009. doi: 10.1177/1461444809336513

[81] D. Seebacher, M. T. Fischer, R. Sevastjanova, D. A. Keim, and M. El-Assady. Visual Analytics of Conversational Dynamics. In T. von Landesberger and C. Turkay, eds., *EuroVis Workshop on Visual Analytics (EuroVA)*, EuroVA. The Eurographics Association, Porto, Portugal, 2019. doi: 10.2312/eurova.20191130

[82] C. E. Shannon and W. Weaver. *Mathematical Theory of Communication*. Univ Illinois Press, 1949.

[83] S. Sinek. *Start With Why: How Great Leaders Inspire Everyone to Take Action*. Penguin, 2009.

[84] T. Spinner, U. Schlegel, H. Schafer, and M. El-Assady. explAIner: A Visual Analytics Framework for Interactive and Explainable Machine Learning. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):1064–1074, 2020. doi: 10.1109/TVCG.2019.2934629

[85] C. Stoiber, A. Rind, F. Grassinger, R. Gutounig, E. Goldgruber, M. Sedlmair, Š. Emrich, and W. Aigner. Netflower: Dynamic Network Visualization for Data Journalists. *Computer Graphics Forum*, 38(3):699–711, 2019. doi: 10.1111/cgf.13721

[86] M. Trier. Research Note —Towards Dynamic Visualization for Understanding Evolution of Digital Communication Networks. *Information Systems Research*, 19(3):335–350, 2008. doi: 10.1287/isre.1080.0191

[87] W. van Atteveldt and T.-Q. Peng. When Communication Meets Computation: Opportunities, Challenges, and Pitfalls in Computational Communication Science. *Communication Methods and Measures*, 12(2-3):81–92, 2018. doi: 10.1080/19312458.2018.1458084

[88] S. van den Elzen and J. J. van Wijk. Multivariate Network Exploration and Presentation: From Detail to Overview via Selections and Aggregations. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2310–2319, 2014. doi: 10.1109/TVCG.2014.2346441

[89] F. B. Viégas, S. Golder, and J. Donath. Visualizing email content. In R. Grinter, T. Rodden, P. Aoki, E. Cutrell, R. Jeffries, and G. Olson, eds., *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 979–988. ACM, 2006. doi: 10.1145/1124772.1124919

[90] P. Watzlawick, J. H. Beavin, and D. D. J. Jackson. *Menschliche Kommunikation: Formen, Störungen, Paradoxien*. Hans Huber, 4 ed., 1974.

[91] Y. Wu, N. Cao, D. Gotz, Y.-P. Tan, and D. A. Keim. A Survey on Visual Analytics of Social Media Data. *IEEE Transactions on Multimedia*, 18(11):2135–2148, 2016. doi: 10.1109/TMM.2016.2614220

[92] Y. Wu, S. Liu, K. Yan, M. Liu, and F. Wu. OpinionFlow: Visual Analysis of Opinion Diffusion on Social Media. *IEEE Transactions on Visualization and Computer Graphics*, 2014. doi: 10.1109/TVCG.2014.2346920

[93] J. S. Yi, Y. A. Kang, J. Stasko, and J. Jacko. Toward a Deeper Understanding of the Role of Interaction in Information Visualization. *Transactions on Visualization and Computer Graphics*, 13(6):1224–1231, 2007. doi: 10.1109/TVCG.2007.70515

[94] B. Yoon and Y. Park. A Text-Mining-Based Patent Network: Analytical Tool for High-Technology Trend. *Journal of High Technology Management Research*, 15(1):37–50, 2004. doi: 10.1016/j.hitech.2003.09.003

[95] J. Zhao, F. Chevalier, C. Collins, and R. Balakrishnan. Facilitating Discourse Analysis with Interactive Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2639–2648, 2012. doi: 10.1109/TVCG.2012.226

[96] J. Zhao, Z. Liu, M. Dontcheva, A. Hertzmann, and A. Wilson. MatrixWave. In B. Begole, J. Kim, K. Inkpen, and W. Woo, eds., *Proceedings of the 33rd CHI Conference on Human Factors in Computing Systems*, pp. 259–268. ACM, 2015. doi: 10.1145/2702123.2702419