Decision Support Systems 178 (2024) 114133

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Complex business ecosystem intelligence using AI-powered visual analytics

Rahul C. Basole [a], Hyunwoo Park [b,*], C. David Seuss [c]

[a] Accenture Data & AI, Atlanta, GA, USA
[b] Graduate School of Data Science, Seoul National University, Seoul, South Korea
[c] Northern Light Group, Boston, MA, USA

ARTICLE INFO

Keywords:
Business ecosystem
Artificial intelligence
Text mining
Complex networks
Interactive visualization

ABSTRACT

Business ecosystems are complex, dynamic systems characterized by a multitude of entities, including companies, ventures, and technologies, as well as activities and trends. Understanding the state of business ecosystems is an increasingly critical strategic imperative for many decision makers, but it is a resource-intensive activity as relevant information sources are dispersed, often highly unstructured, and not integrated or curated to deliver actionable insights. In this research, we present the design and implementation of an interactive visual analytic system that integrates artificial intelligence and graph visualization techniques to augment decision makers' understanding of the complex public narrative associated with business ecosystems entities. Our system is driven by a real-time content engine of 100,000+ global data sources including press releases, news articles, industry reports, analyst blogs in multiple languages organized across several domain-specific repositories. Following a user-specified query, the engine extracts both domain-agnostic and domain-specific entities and concepts for each document in the result set. We then model and visualize the resulting data as a dynamic, multipartite network and implement graph pruning algorithms and interactive data controls to enable users to interactively explore and discover the underlying business ecosystem from multiple perspectives. We illustrate and discuss the value of our system using representative use cases. Our study makes multiple contributions to visual decision support theory and practice, including mining unstructured data, constructing and interacting with knowledge graphs, and designing visual analytic tools for ecosystem intelligence. We conclude the study with implications and future research opportunities.

1. Introduction

Understanding the structure and dynamics of business ecosystems has become an important strategic imperative to decision makers across many domains [1–6]. In order to survive in today's hypercompetitive environment, companies must rapidly identify competitive actions, emerging trends, and innovation, technology, and investment opportunities [7,8]. These tasks often require a systematic lens across industries, markets, and countries. While many structured ecosystem data sources exist to help with these analyses individually, they are generally not integrated and thus provide only a partial perspective of the rich interactions that are occurring and evolving on a continuous basis [9]. Moreover, there is much greater amount of unstructured textual data – press releases, news articles, industry reports, analyst blogs, etc. – that can reveal potentially hidden and/or complementary insights of the current public narrative that decision makers need to make more informed decisions [9–11].

However, the burden of synthesizing the analysis of a business ecosystem from multiple textual data sources and perspectives often falls back on the investigator. Moreover, the amount of effort required for undertaking this synthesis is not trivial yet. There are many challenges associated with such an approach [12]. Simply having access to a wealth of unstructured data does not necessarily translate into the right level of insights. Consider a Google search for "strategic collaboration in artificial intelligence". This query can result in millions of documents, albeit ranked by relevance. Such a result set is often rendered as a list of hyperlinks paginated by ten to hundred items per page at most. Manually scouring through these documents would be incredibly resource intensive and, while maybe some insights to those aforementioned ecosystem questions could be derived, it would be difficult to assess whether something truly important was contained in some documents or what other insights not necessarily in focus could be revealed.

Leveraging artificial intelligence (AI) techniques to sift through large corpora of documents can overcome these challenges. Using natural language processing (NLP) and advanced text mining, important terms and concepts across the entire corpus can be extracted. The

* Corresponding author.
E-mail addresses: rahul.basole@accenture.com (R.C. Basole), hyunwoopark@snu.ac.kr (H. Park), david@northernlight.com (C.D. Seuss).

https://doi.org/10.1016/j.dss.2023.114133
Received 4 May 2023; Received in revised form 26 November 2023; Accepted 28 November 2023
Available online 30 November 2023

challenge with only using text mining results, however, is that it would still be difficult to consume how different extracted entities relate to each other. Without a systemic perspective encompassing the rich interactions among the extracted entities, decision makers may miss peripheral entities, concepts, clusters, or temporal patterns.

Ecosystem intelligence is intrinsically a human-centric activity [13]. Whether practitioners want to understand the competitive landscape and identify investment/innovation opportunities or scholars studying firms, technologies, industries, or markets, insight generation and decision support requires human sensemaking and judgment. Thus, whatever text mining analysis on large corpora should be presented to decision makers in a digestible visual format and a system designed for ecosystem intelligence should ideally also allow interactive exploration with quick feedback.

In this paper, we present the design, development, and implementation of an interactive visual analytics system that leverages both text mining and network visualization of massive, real-time data sources, to provide analysts, decision makers, and ecosystem scholars an ability to systematically explore and discover the public narrative around user-specified search terms through an intuitive user interface. In the context of business ecosystem intelligence, a narrative can be referred to as a description that conveys the key elements and dynamics of the ecosystem. Narratives encompass interconnected stories, serving as a vital tool for humans to comprehend and convey events spanning time [14]. Furthermore, they influence the actions of stakeholders and offer insights into the ecosystem's achievements, as narratives specific to the ecosystem are influenced by both the environment and the actors within it [15,16]. In this context, a narrative can help stakeholders to better understand how different components of an ecosystem are interconnected and how they influence each other [14,17]. It can also provide insights into market trends, emerging technologies, competitive forces, regulatory factors, and other factors that shape the ecosystem. Our novel interface provides novel ecosystem insight and sensemaking capabilities, thus contributing to the business intelligence and analytics domain [18]. Prior decision support systems (DSS) research has also long stressed the importance of visualization as an essential tool for data-driven decision making (e.g., [19–21]).

Moreover, it has been shown that many managerial decision making tasks often require large, complex, and interconnected data. To make this digestible and actionable, it is critical to transform this data into representations that can help humans understand it better. Given the particular strengths of human vision, this representation is frequently visual. When the underlying data is unknown, the decision-making process begins with exploratory sensemaking analysis [22]. Our study thus builds on and contributes to the existing visualization-centric business intelligence systems literature. We extend prior work by focusing on how to derive insights from unstructured data for business ecosystem intelligence purposes. On visual representation, we make the case for a multi-partite network representation approach to highlight the interrelatedness of business ecosystem entities. From an interface perspective, we show that dynamic queries, interactive controls, and details on demand can enhance the intelligence sensemaking process. Our prototype system is powered by an operating industry-class data aggregation engine. We illustrate the application and value of our system using real-world ecosystem intelligence use cases. Collectively, these aspects advance our theoretical and practical understanding of ecosystem intelligence, an emerging area of DSS research.

The rest of the paper is structured as follows. Section 2 provides a review of literature that provides the theoretical foundation of our work. Informed by the prior work, Section 3 establishes design requirements. Section 4 introduces the architecture and main features of our system and explains how the design requirements are satisfied by our design of the system. Section 5 demonstrates the value of the system by showcasing an illustrative use case. Section 7 concludes the paper with summary and future research opportunities.

2. Related work

Our work draws on three streams of research: text mining, business ecosystems, and visual ecosystem intelligence. In the following section, we briefly review each stream and describe how it influenced the design and development of our visual decision support tool.

2.1. Text mining

The explosive growth of many diverse types of unstructured textual data – including news articles, press releases, analyst reports, annual reports, corporate filings, blogs, social media posts, presentations, publications, etc. – is creating an unprecedented potential to develop a much deeper understanding of the rapidly evolving business environment [9,12,23]. While structured data, often found in curated, proprietary, and subscription-based repositories, has allowed researchers and analysts to explore retrospective, historical, "what happened" types of questions, the fluid, diverse, and more expressive nature of unstructured textual data allows a much deeper and more holistic exploration of what happened, why things may have happened, and what could potentially happen. In addition to being able to answer such important questions, the public availability and general pervasiveness makes the mining of textual data even more attractive for scholars and practitioners.

Text mining leverages natural language processing (NLP) to transform unstructured textual data into normalized, structured data suitable for analysis or to drive machine learning (ML) algorithms. There are several key purposes for text mining. Broadly, they can be organized in terms of information extraction, topic tracking, document summarization, categorization, and sentiment analysis [24]. *Information extraction*, often the first step in mining unstructured textual data, includes the identification of key words, phrases, and entities and their relationship within the text. Information extraction can be driven by the frequency or importance of words, or guided by the identification of particular entities, such as people, places, organizations, and dates. *Topic tracking* examines how a topic evolves and can provide alerts if and when a given topic emerges. For instance, if a company or competitor is mentioned in the news, an analyst may be notified of the occurrence and pointed to the article of interest. *Summarization* is the task of reducing the length and detail of a document while retaining its main points and overall meaning. While important entities can be easily identified by contemporary algorithms, there are still some limitations in how machines capture semantics and provide meaning in text. *Categorization* involves the identification of the main themes of a document, often through a bag-of-words approach, which enables the detection of topic presence and document ranking. *Sentiment analysis* is the computational detection of subjective information such as opinions, attitudes, and feelings expressed in text. We often see this applied to the analysis of customer reviews, analyst reports, and more recently corporate disclosures.

The use of text mining in business, finance, accounting, marketing, and information systems research is not new and is growing steadily. It has been applied to a diverse set of contexts, domains, and textual corpora, including credibility of online reviews [25], behavior in e-commerce marketplaces [26], retention of enterprise blog users [27], design of ranking systems [28], social media use [29], corporate risks [30], IT innovation adoption [31], detecting cybercrime [32], linguistics styles in user communities [33], and analyst discussions to earning calls [34]. A comprehensive review would be beyond the purpose of this study; interested readers are thus referred to several review studies. [35] provide a survey of textual analysis in accounting and finance. [36] review the use of topic modeling in management research. [37,38] provide a review of text analytics in marketing.

The use of text analytics for understanding industries and markets has also been growing in recent years. In one of the earliest studies we could identify, [39] extract interfirm relations from news and press releases. Hoberg and colleagues [40,41] use text analytics applied to product descriptions to determine industry structure and segments. [42] demonstrate how to convert user-generated content (blogs, chats, forums) to market structures and competitive landscape insights.

The predominant focus of existing studies is to use text mining to extract entities of interest, determine document sentiment, or identify relevant topics that could be used for subsequent input into empirical models. Very rarely however do studies visualize the resulting text analytic insights directly, thus not necessarily allowing the researcher to interactively probe the data. In this study, we argue that the use of interactive visual analytics for understanding textual data is of particular importance to many researchers and scholars, in particular for ecosystem intelligence.

2.2. Business ecosystems

The term *ecosystem* is used pervasively in both industry and academia. Drawn from the natural sciences, it is a metaphor commonly used to describe the complex, dynamic, hyperconnected nature of many social, economic, and technical systems that exist today [43]. Ecosystems are typically characterized by a large, heterogeneous set of often geospatially distributed entities that are interconnected through various types of relationships [44]. The structure and behavior of an ecosystem cannot be identified by inspection of the entities alone, but rather by the interaction of the entities. Ecosystems are highly dynamic, with entities entering and leaving; relationships formed, renewed, and deleted; and entity and relationship attributes changing constantly [45]. It is this scale, complexity, and emergent dynamism that makes systematic understanding of ecosystems particularly challenging [3,46–48].

Given its applicability to describe a wide range of contexts, the study of ecosystems has been a topic of significant interest across many academic disciplines, including management, strategy, information systems, innovation, and entrepreneurship [49–52]. Broadly considered, the focus of these studies ranges from different types of ecosystems (business, innovation, entrepreneurial, technology, service, software, digital), at different levels of analysis (interorganizational, product/service, artifact, people), and with varying methodological approaches (theoretical, conceptual, empirical, analytical) [2,53–56].

There is a growing empirical evidence that ecosystems allow firms to innovate, increase revenue growth, access new markets, and access new customers [57]. Moreover, given that a firm is no longer an independent strategic actor [4], decision makers must continuously understand, analyze, and scan their ecosystem, in order to identify changes and competitive actions and craft appropriate strategies. Traditional strategy frameworks and analysis methods are of little help and new approaches are needed. What complicates matters is that a truly holistic understanding of business ecosystems will require a multi-faceted lens, including an understanding of the complex interdependencies between incumbent and emerging organizations, technologies, people, trends, and activities. Existing data sources are likely not capturing all these entities in a single place; similarly, existing methods likely fall short in being able to model this complex system.

2.3. Visual ecosystem intelligence

The business data visualization literature can be generally categorized into four main research areas: business intelligence, finance, customer-centric application, and business ecosystems. In this section, we focus on prior work in business ecosystems only. Readers interested in a comprehensive review of the other areas are referred to [58].

The emergence of interactive business ecosystem visualization is in large part driven by the growing need for understanding complex ecosystems, the proliferation and availability of relevant data, and significant advances in visual decision support technologies [59]. Visual business ecosystem intelligence can thus be defined as the application of interactive visual analytic approaches to representing, analyzing, and managing complex business ecosystems for purposes of sensemaking, insight generation, and decision support.

At its core, visual business intelligence solutions consist of graphical representations of different ecosystem phenomena of interest. Given the typical desire to understand structure and composition, business ecosystems are frequently depicted graphically through network (node-link), matrix, or space-filling (treemap, sunburst, etc.) representation techniques. The appropriate choice of visual representation often depends on the underlying data complexity, task, and user characteristics. Indeed, visual representations can have a different effectiveness driven by these contextual elements [47]. Faber et al. [60] identify different types of commonly used visualization in ecosystem analysis and suggest requirements and design principles.

Existing business ecosystem visualizations have focused on developing visualizations for a variety of domains and contexts, including the composition and performance of stock markets data [61,62], alliances and partnerships in technology ecosystems [45], and supply chain networks [63,64], API ecosystems [65,66], entrepreneurial ecosystems [1,67], FinTech [68], cloud computing [69], microservices [70], machine learning and artificial intelligence [71,72], and coopetition [73].

These visual representations are typically enriched by encoding relevant ecosystem entity attributes as graphical elements typically denoted through position, size, color, shape, and thickness. In doing so, users' attention can be guided to specific entities, clusters, or relationships of interest. Since the amount of information/data to be encoded in a visual representation should not be overwhelming, additional details are often provided through interactive techniques (for example through details on demand, tooltips, or hover overs). In contrast to static visual representations, interactive visualizations allow a deeper probing and exploring of ecosystems. Dynamic selection, filtering, searching, or positioning of entities and relationship as well as adding and removing them, for instance, provides decision makers the ability to not only discover and make sense of ecosystems but also pursue potential what-if questions that traditional visual analysis approaches may not afford.

Another aspect that is critical to visual business ecosystem intelligence is the consideration of the dynamic, temporal aspect of ecosystems. Ecosystems, by definition, are continuously evolving, with new entities and relationships entering and departing [2]. While existing studies have provided temporal snapshots, it is sometimes difficult to track and compare how the position of individual firms in relation to the overall structure may be changing. Effective visual business ecosystem intelligence solutions must therefore be able to capture, represent, and enable playback the dynamic aspects inherent in business ecosystems. If future models are generated, a projection of the emerging structure would be beneficial.

The field of visual business ecosystem intelligence is still in its infancy, but over the past few years several notable studies have emerged that have addressed these desirable capabilities. One of the first visual business ecosystem intelligence solutions was dotlink360 which provided multiple coordinated views to allow analysts to explore, discover, and understand interfirm networks for a focal firm, specific market segments or countries, or the entire business ecosystem [31]. Ecoxight extended this solution by creating a web-based platform that provided multiple coordinated views of multipartite, multiattribute, dynamic, and geospatial ecosystem data and allowed users to upload their own ecosystem data [59]. Building on this work, [74] suggest a visual language to formally model and visualize business ecosystems.

Other visualization studies have focused on alternative ecosystem characteristics, analyses, or issues. Epheno, for instance, focuses explicitly on visualizing the temporal and sequential characteristics of ecosystem relationships, enabling decision makers to gain both systemic (macro) and detailed (micro) insights into a firm's relationship activities and discover patterns of multidimensional relationship formation [75]. In addition, epheno provided a rich set of additive crosslinked filters, comparative views, and a dynamically updated Markov model visualizations to inform decision makers of past and likely future strategy moves. Bicentric Diagrams provided a means to compare the distinct and overlapping network aspects of two ecosystem entities, focusing on the identification of sets, relationships, and reach and enabling visual exploration of strong and weak ties [76]. Graphiti, while not explicitly a visual business intelligence tool, provides capabilities to interaction technique to model networks that allows users to demonstrate to the system a subset of nodes and links they wish to see in the resulting network, thus facilitating what-if analyses [77]. Graphicle combines a unit-based with network visualization approach along with a dynamic histogram based filters to enable rich simultaneous, contextual exploration of individual data units and their attributes and their underlying network structures [78].

The discussion above underlines in particular the critical importance of data to ecosystem intelligence. All of the aforementioned systems require the curation and import of a somewhat complete dataset. Important data can come from highly structured as well as unstructured data sources [9,79]. It also requires an a priori focus on the entity types of interest. The resulting dataset is then visualized. Yet, the generation of these datasets can be extremely time consuming, and when generated can already be out-of-date, thus not providing the necessary timely insight needed for ecosystem intelligence. Furthermore, any subsequent analyses, potentially triggered by initial observations and findings, will have to follow a new data curation process. It is this manual, tedious, and time-consuming process that could stifle repeated business ecosystem intelligence activities. In our study, we aim to fill this gap by providing a more complete end-to-end workflow and leveraging real-time textual data sources.

3. Design requirements

We followed a design science approach to develop our visual analytic system. The design of our system was guided by both prior work developing visual analytic tools for business intelligence (e.g., ecoxight [59], epheno [75], VisualSCM [64], Graphicle [78]). The design of all these prior systems leveraged an iterative human-centered design approach in that they put prototypical users at the center of the development process. For our study, we elicited requirements from discussions with analysts, technology executives, investors, and ecosystem scholars. Specifically, we identified the following five overarching requirements:

- **Open form search query (R1).** Analysts and decision makers are familiar with Google like search interfaces to formulate queries. The system should allow users to input any combination of terms and keywords to begin their investigation. In ecoxight and epheno, for instance, users can initiate their exploration using a keyword and text entered in a search box.
- **Coordinated, interactive views of visualization and document results (R2).** Users should be able to view the search results in raw format (documents) as well as in visual form. The two views should be connected and coordinated to allow users to understand what is driving the visual and where individual documents reside in the big picture of the result corpus. In all aforementioned systems, we utilized a combination of main window and side panels to display relevant information on demand. This approach has proven to be effective to limit the cognitive load and providing contextual information when needed.
- **Rich filtering capabilities (R3).** Searches can be broad and lead to more results than needed. The system should provide rich filtering capabilities, thus enabling a top-down (overview first) then focus inquiry paradigm. In all aforementioned systems, we provided dynamic filters (including checkboxes, dropdown lists, etc.), to enable deep exploratory capabilities.
- **Intuitive, point-and-click UI (R4).** Users are likely going to be business users and analysts not familiar with command line techniques. An intuitive graphical point-and-click UI can help facilitate learnability and adoption. All previous systems use a graphical interface commonly found in web-based applications.
- **Export and Post Processing Controls (R5).** The results of the ecosystem intelligence will be embedded in corporate presentations and/or social media. Users thus require image exporting and post-processing (e.g., network layout optimization or alternative layouts) controls. In ecoxight, for instance, we enable the export of network visualizations into Gephi formats to facilitate extended analyses not afforded by the system.

Based on the aforementioned design requirements, we develop a conceptual system architecture for building a text-powered visual analytic application for ecosystem intelligence. Specifically, we propose that an effective unstructured data-driven ecosystem intelligence system should contain three main capabilities. The text mining engine should dynamically identify relevant documents and extract entities of interest based on user input. The visual analytics engine should convert the resulting dataset in a graphical way that facilitates the interpretation and sensemaking process and provide relevant descriptive analytics. Both engines should be tightly integrated with a human-in-the-loop consideration that not only augments and amplifies human intelligence, but also provides confidence and trust in the underlying data, system interactions, and ecosystem insights.

4. System design and architecture

4.1. Workflow

Fig. 1 shows a high-level workflow diagram of using the interactive text mining and visual analytic capabilities of our system. In the following sections we introduce each part of the workflow. A corresponding video describing the system and its capability is available upon request.

4.2. User interface

The user interface (UI) consists of several main sections, as shown in Fig. 2. The top header includes the data repository selection and search query interface (A). The left-hand contains a collapsible panel with filtering controls (B). The right-hand contains a collapsible panel for network and node/link characteristics as well as a listing of the documents (C). The main visualization is displayed in the center of the UI (D). At the bottom of the UI are control options for the visualization as well as exporting and personalization capabilities (E).

4.3. Search

Each session begins with the selection of a focus domain data repository (IT, Life Sciences, Healthcare, Financial Services, Energy, Agribusiness, Chemical, Food and Beverage) and the specification of a query by the end user. We used a search query approach (instead of predefined concepts/terms) as users are familiar with Google Search like interfaces and it provides the most flexible means to inquire about a topic of interest. Query terms can be entered in any form including individual words, keyword terms, terms in quotation marks, all with Boolean logic.

The query then triggers an action to search for relevant documents through the Northern Light (NL) engine that contain the search terms and according to the search options selected. The NL search engine continuously indexes the document corpora in the domain repository. The indexing process employs established text analytics techniques such as document tokenization, sentence segmentation, chunking, syntax parsing, and sentence chaining. One of the technically challenging aspects addressed by the NL search engine is named entity recognition (NER)

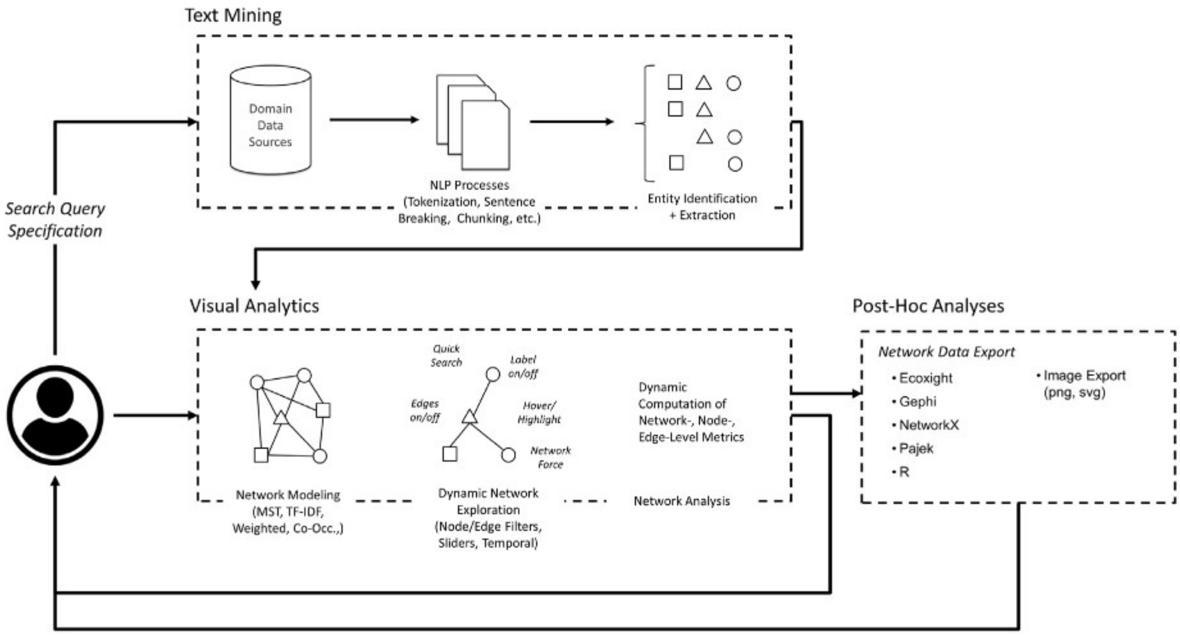

Fig. 1. Conceptual workflow diagram.

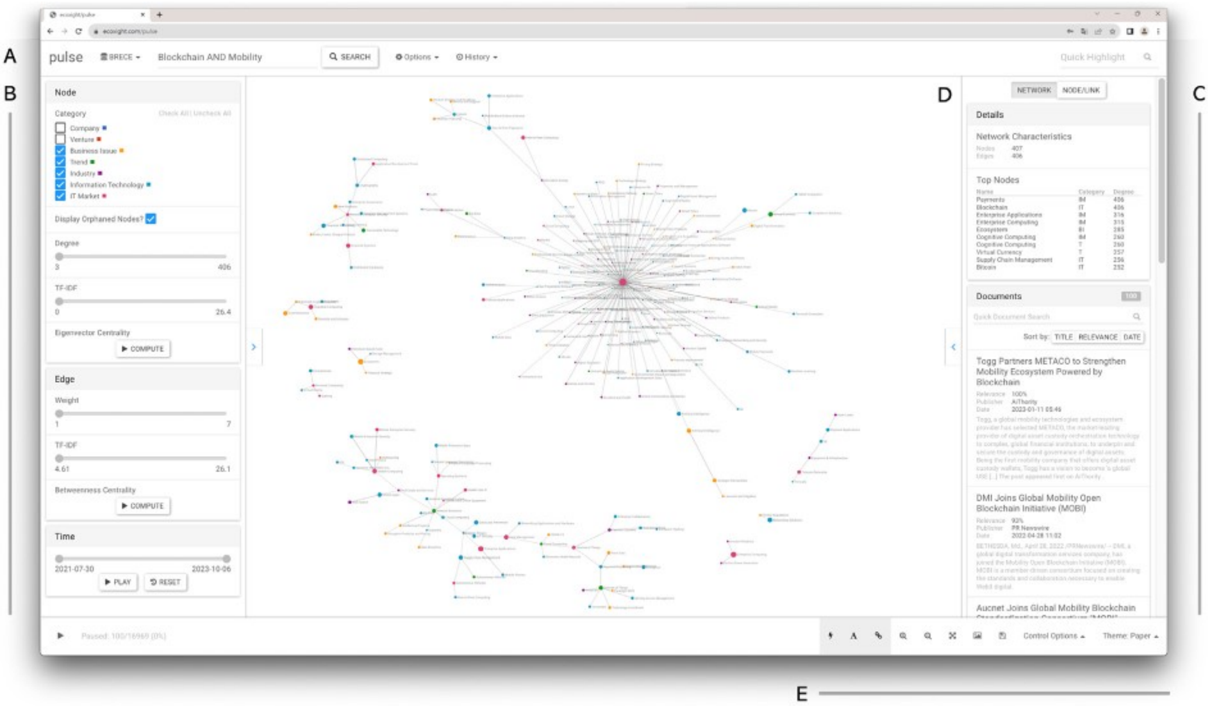

Fig. 2. Visual depiction of main user interface.

within the vast document corpora. The NL search engine generates annotated search results, with each article contained in a search result tagged by various categories of keywords, including named entities such as company names actually used, making this system and its search results highly relevant to end users.

We are cognizant that the precision of the search terms can significantly influence the identification of relevant documents. An overspecification can lead to too few documents, a general query can lead to too much noise. The user must balance these two extremes when specifying the query. In the future, semantic search capabilities could help (similar terms, did you mean, etc.) to overcome this issue. A History dropdown keeps track of all search queries during a session and can be selected for reuse.

4.4. Search options

Users have the ability to specify various search options to constrain their queries. Search Order determines whether documents should be extracted in terms of relevance score (numerical value assigned to a document that indicates how relevant that document is for a given query) or in terms of document date (most recent first). Search Field determines whether the query terms apply to the entire document or

Table 1
Data source domain repositories and entities types extracted.

Domain	Entity type																
	C	V	BI	T	I	IT	ITM	TA	CT	H	HW	FS	EN	EC	AC	IC	N
IT news	x	x	x	x	x	x	x										
Life sciences	x	x	x	x	x			x	x								
Healthcare	x	x	x	x	x					x	x						
Financial services	x	x	x	x	x							x					
Energy	x	x	x	x	x								x				
Agribusiness	x	x	x	x	x									x			
Chemical	x	x	x	x	x										x	x	
Food and beverage	x	x	x	x	x												x

only the title of the document. We note that more granular search criteria, such as sentence or paragraph level may provide more fine-grained control. Search Period determines what timeframe should be considered. Rather than providing exact date range controls, we use a set of predefined date ranges (3 months, 6 months, 1 year, 3 years). Sources determine what data source types should be included. Our system has access to six different broad source types including Business & Technology News, IT Analyst Blogs, Industry Authority Blogs, National & Global News, US Regional News, International Regional News. By default, we use "All" but all other source types (described above) can be selected individually. Language allows the specification of the languages. By default, we consider English-based documents only. Other languages include Spanish, French, French, German, Italian, Japanese, Portuguese, Russian, Spanish, Chinese, and Dutch.

4.5. Continuous downloading

Given the scale of potential search results (100,000+) we implemented a continuous downloading function to buffer the size and provide quick results. Depending on the search order criteria selected, each search query provides either the most relevant (or most recent) 100 documents. Additional documents, in 100 document increments, can be brought in by using the play button. In doing so we can manage both the scale and relevance of the underlying document corpus.

4.6. Entity identification and extraction

Each query identifies and extracts entities from each of the documents. Some entity types are domain-agnostic, while others are specific to a given domain (see Table 1).

4.7. Visualization panel

The visualization panel (Fig. 2D) contains the multi-entity visualization of our search query. Given the nature of our data and focus of ecosystem intelligence tasks of understanding the relationship between entities, a network representation approach was deemed most appropriate as suggested by previous studies (e.g., [47]). A network visualization includes the visualization of relationships (edges or links) between data elements (nodes). In our study, nodes represented one of the entity types listed in Table 1 and edges between nodes corresponded to if the entities co-occurred in a document. We color-encode nodes by entity type. We scale both nodes and edges by the frequency of occurrence. The resulting network is thus an undirected, weighted multi-partite network.

Since many entities can co-occur frequently, the resulting network can be very dense (highly connected). We implemented several network pruning algorithms to reduce some of this density without losing any structural information and providing a readable network graph. Specifically, we implemented three network pruning algorithms: (1) maximum spanning tree, (2) raw degree, and (3) none. An illustrative example of the three pruning options for the same query and search result is shown in Fig. 3.

A maximum spanning tree only keeps those strongest edges that minimally connect the network into a tree without a cycle. The algorithm for implementing the maximum spanning tree layout is shown in the Appendix (S1) [80]. Let $G = (V, E)$ denote the entity co-occurrence graph where V is a set of nodes (or vertices) and E is a set of edges (or links). In our case, entities extracted are nodes and co-occurrences are edges. Edges are weighted by the number of co-occurrences between the two entities. Given this co-occurrence graph, the maximum spanning tree algorithm extracts a tree graph that does not contain any cycles. By doing so, the resulting graph contains $|V|$ nodes and $|V| - 1$ edges. We allow users to select among several edge options: TF-IDF (default), weighted co-occurrence, and co-occurrence.

The raw degree pruning option considers the degree of each extracted entity and uses a rank ordered set of edges. This reduces the total number of edges in the graph. The last option is not to prune at all, thus showing the complete network graph. This provides the largest number of edges and likely results in a very dense and unreadable network.

For all network visualizations, we implement a force-based layout algorithm. A force-based layout is based on the idea that network entities are shaped by mechanical laws, assigning repulsive forces between nodes and attraction forces between endpoints of edges. The use of a force-based layout is particularly appealing when the motivating issue is to identify central or prominent nodes, peripheral nodes, or clusters in a network graph.

4.8. Visualization controls

To facilitate readability of the resulting network visualization, we implemented several important controls following prior work [59,64,76], including a force on/off (allowing a graph to freeze in position and turn off the force-directed power), labels on/off (to remove potential label clutter), edges on/off (to see nodes only), zoom and pan capabilities (to move in/out the visualizations), and fit-to-screen (for a rapid reset and to see the whole graph). To find specific nodes, we also implemented a quick highlight function with partial text matching capability.

4.9. Filters

To facilitate dynamic network exploration, the filter panel contains controls that allow users to specify desired ranges of important node and edge-level characteristics as well as time ranges. By default all elements for the full time range are shown.

4.9.1. Node-level characteristics

At the node level, there are five characteristics that users can control. Users can specify which node categories (see Table 1) to include in the visualization using multi-select checkboxes, what degree (number of edges) range the nodes should have, whether to include orphaned nodes or not, the range of eigenvector centrality (measure of the influence of a node in a network), and TF-IDF (measure of how important a word is to a document in a collection or corpus).

(a) Maximum Spanning Tree (default)
•TF-IDF (default)
•Weighted Co-Occurrence
•Co-Occurrence

(b) Raw Degree

(c) None

Fig. 3. Network pruning options.

4.9.2. Edge-level characteristics

At the edge, level, there are three characteristics that users can control. Users can specify the weight of edges (how frequent two nodes appear), the range of betweenness centrality (the number of geodesic paths that run along edges), and TF-IDF.

4.9.3. Temporal focus

Users can also specify the temporal range for the documents (start date, end date) using a slider control. The Play function replays the network over time and provides an evolutionary perspective through animation. By default the whole time range (earliest to latest date) is selected.

4.10. Control options

Our system also offers capabilities to export the resulting data and visualization for post-processing. We enabled this functionality as users would like to be able to embed the visualization into presentations or onto social media posts or apply more sophisticated network visualization layouts for further processing. We allow users to download the resulting graph either as an image (.png format) or as network files (Gephi's .gexf [81] or Ecoxight [59]).

As we mentioned earlier, users can select one of three network pruning algorithms, namely maximum spanning tree, raw degree, or none. One of these options can be selected from the control options menu. Users can select among three: Maximum Spanning Tree Edge Option with TF-IDF, Co-Occurrence, Weighted Co-Occurrence. One of the parameters in force-directed network layouts is how strongly the nodes should repel or attract each other. This parameter, which we operationalized as network spread, can influence how close nodes or separated nodes are from each other and thereby determine readability and aesthetic to a certain extent. Users can specify the degree of network spread using a control slider from 1 (very loose) to 10 (tightly compacted).

4.11. Results panel

The results panel consists of two main tabs: the Network tab and the Node/Link tab (see Fig. 2C). The Network tab contains information about the whole network, providing summary on network characteristics (# of nodes, # of edges, and a list of top nodes ranked by degree) and a paginated list of all the documents that have been retrieved. Documents can be sorted by Title, Relevance, and Date and a Quick Document Search allows users to quickly scan documents for relevant information. Similar to a Google search result, the listing contains the title of the document, hyperlinked to the original source that can be opened in a new window, and relevant information (including relevance score, publisher, date, and a short description). The Node/Link tab contains information on a selected edge and all the associated documents with it.

4.12. Interactions

To facilitate exploratory analysis, our system offers several interaction techniques. Hovering over a node highlights that node and its incident nodes and edges, while fading the remaining network elements. This provides the user not only feedback on what nodes are selected but also provides a way of browsing the neighborhood network of a node. Hovering over an edge highlights the edge and two incident nodes.

Clicking on a node freezes the visualization and filters out the document list to only those documents that contain the node. Similarly clicking on an edge switches to the node/link tab and shows only documents in which the two incident nods are contained. To facilitate multiple analysis entry points, users can hover over documents and the corresponding nodes and/or edges contained in these documents are highlighted. Double-Clicking a node makes that entity the focal search term and triggers a new query.

4.13. Implementation

Our system is implemented as a web-based ecosystem intelligence tool using d3.js, CSS, and angular.js using the NorthernLight Millie API.

5. Illustrative use cases

To demonstrate the value and utility of our system, we describe three illustrative use cases for prototypical real-world and very contemporary ecosystem analysis use cases. Each use case aims to highlight a different aspect of our system. The first use case focuses on the recent narrative around an organization; the second use case focuses on the narrative of a specific technology over the past three months; and the last use case focuses on both the narrative of a technology in a given context for a time period in the past. Since our system uses an open search term as a starting point, any common query could be used, including the ability to learn about a specific company, a particular technology, business issue, or combination thereof.

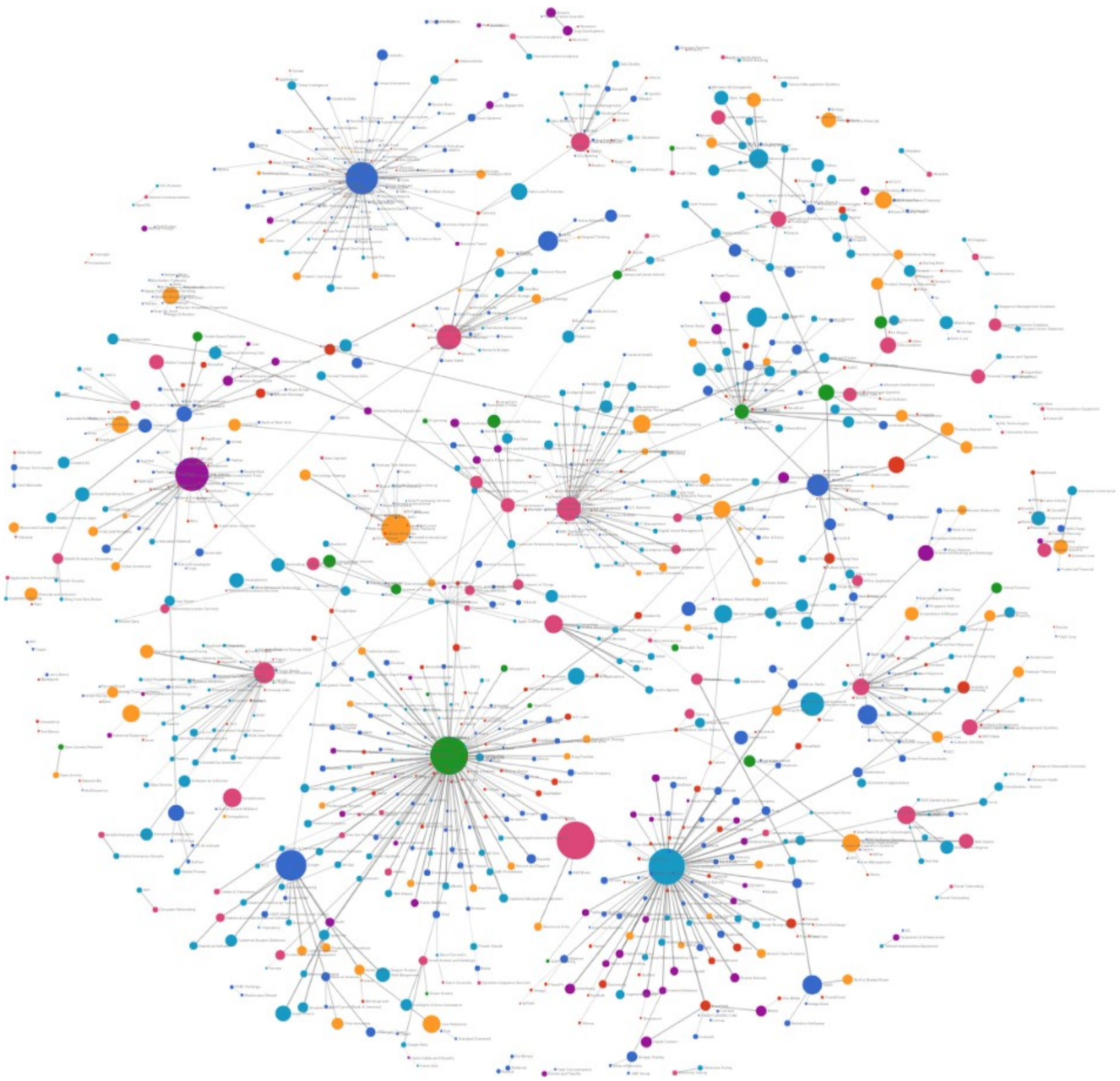

Fig. 4. Multi-entity view of the OpenAI Ecosystem with Company, Venture, Business Issue, Trend, Industry, Information Technology, and IT Market nodes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5.1. Use case 1: Organization focused

OpenAI is a private U.S.-based AI research and deployment company that has been credited with much of the hype around generative AI after their launch of groundbreaking products such as DALL-E (image generator), Whisper (speech recognition system) and ChatGPT (conversational chatbot). The field of generative AI is moving at lightning speed, with new companies, offerings, and technologies emerging rapidly, and getting a holistic view of what is shaping it can be challenging.

We searched for ["OpenAI"] in the full-text documents for the past three months (February 2022-April 2023) to begin our inquiry. The search led to the identification of 13,260 unique documents, which after complete download and a maximum spanning tree filter resulted in 3503 ecosystem entities and 3502 edges between them (see Fig. 4).

The force-directed visualization of the resulting prominent entity set shows that several prominent companies (large blue nodes) dominate the narrative, including Microsoft, Amazon, Google, and Apple. Yet, there are also smaller company nodes that stand out such as Intel and Nivida, two important players in the AI acceleration space. The two largest trend entities (green) include Cognitive Computing and Cloud Computing; the most discussed industry (purpose) is Web Search. Overall, when exploring the network visualization interactively, the results show that a diverse set of topics shape the OpenAI ecosystem.

5.2. Use case 2: Technology focused

The Metaverse is a vision of what many technology pundits believe to be the next iteration of the Internet, a collective virtual shared space, created by the convergence of virtually enhanced physical and digital reality. The Metaverse has been proposed to fundamentally shape our economic and social life over the next decade. While the early hype was high, it appears to have gone through a cooling off phase recently. What industries are primarily discussed in association with the Metaverse today?

To explore this question, we begin our search with the query of ["Metaverse"] only found in the titles of documents found in the IT News data repository for the past six months (November 2022-April 2023). We choose this time period to focus in on the more recent trends. The search results in 3919 unique documents. After continuous

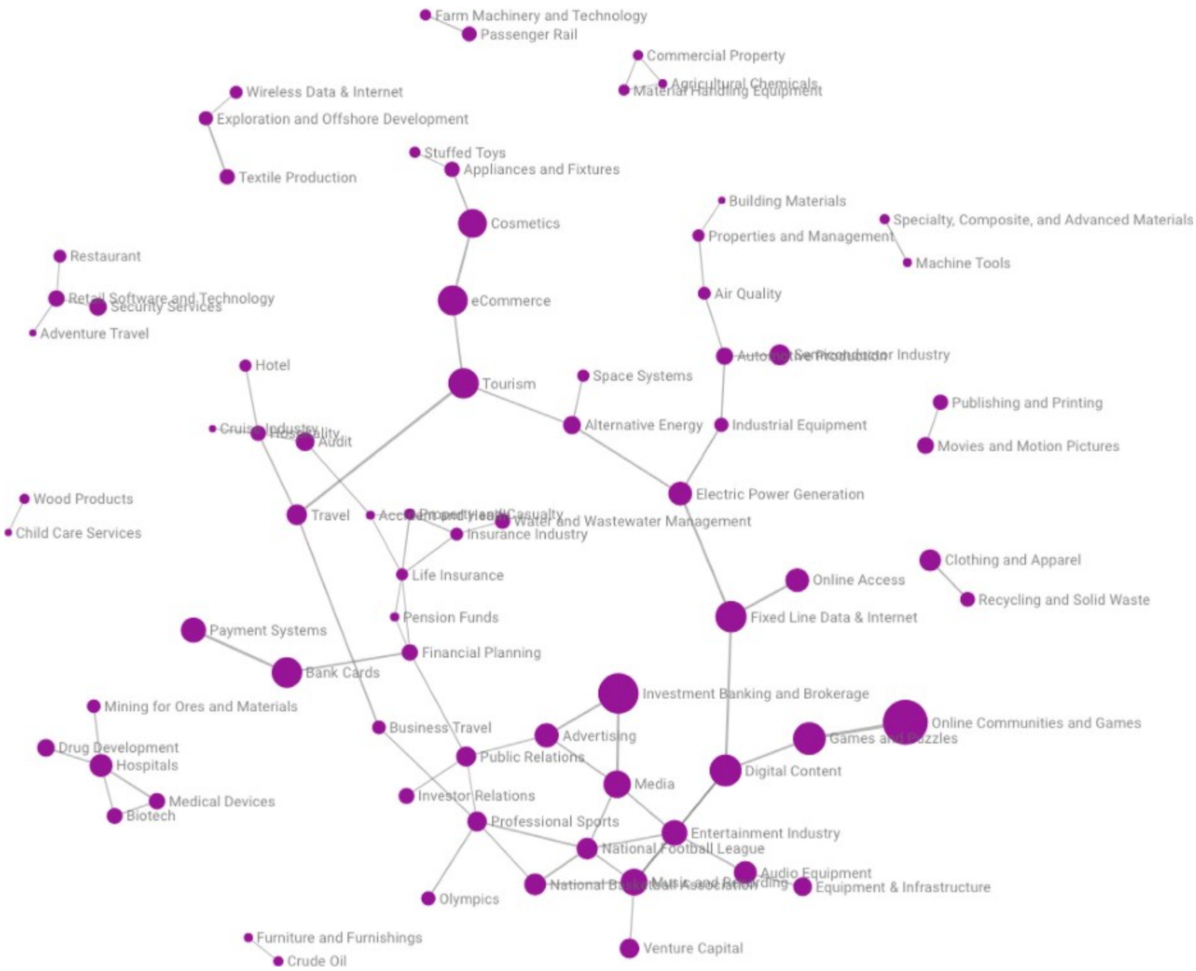

Fig. 5. Industry-centric view of the metaverse ecosystem.

download of all articles and entity extraction, the resulting ecosystem map consists of 1555 node entities and 4437 edges between them.

Focusing only on Industry entities, we identify that while of the most mentioned industries continue to be Online Communities, Games, and Digital Content, several other interesting interconnected industry clusters appear (see Fig. 5). The first is the cluster surrounding the Entertainment Industry, which includes Media, Digital Content, Music and Recording, and as well Sports leagues. The second interesting cluster surrounds Tourism, Travel, and Business Travel. Both of these clusters and the documents that shape them relate to the experience economy. The Metaverse thus is clearly being discussed in relation to making new experiences happen.

5.3. Use case 3: Technology + Context focused

The emergence of artificial intelligence promises to fundamentally transform many industries, including healthcare. The potential use of AI in the fight against global pandemics, such as COVID-19, is of particular interest to a wide range of stakeholders, including decision makers, regulators, and end consumers. But what was the narrative around AI during the early stages of COVID-19? What were the trends or issues and what companies, technologies, or initiatives were associated with it?

To explore this question, we begin by searching for ["Artificial Intelligence" AND COVID] in the full texts of documents in Business & Technology news sources in the Healthcare data repository over the first three months (February-April 2020) when COVID energed. Our relatively broad search query (in less than 2 s) results in 4503 unique documents. After a continuous download of all articles and entity extraction, nearly 2000 pertinent entities and 5137 edges are identified. The resulting network visualization (see Fig. 6), filtered down to only companies (blue) and ventures (red) reveals a core–periphery structure, with prominent technology, life science firms at the center. Several important startup ventures, sized by prominence can be identified near the core, but mostly on the periphery. Hovering over a firm highlights its adjacent neighbors while fading out the rest of the network for readability. News articles corresponding to a selection are shown in the documents pane.

In order to understand pertinent general business issues associated with AI and COVID-19, we use the filter panel on the left and select only the business issue entity type (yellow) (see Fig. 7). Several issues stand out, including New Product Development, Value Chains, and Strategic Partnerships. Selecting one of the boundary spanning nodes – Drug Pipeline – reveals its co-occurring relation with core competence, intellectual property, and product strategy and roadmaps.

Shifting to healthcare specific concepts and issues, we select the Healthcare entity type (see Fig. 8). The maximum spanning tree layout reveals three separate clusters: one surrounding hospitals, another around telemedicine, and a third around health systems. Exploration of each of the nodes and its documents reveals that the Hospital cluster is primarily associated with the use of AI as it relates to patients, patient care, patient-related data, and patient safety. The Telemedicine cluster appears to be related to medical centers and healthcare facilities. Lastly, the health systems cluster appears to examine its connection to regulatory entities, such as the FDA, and broader health issues such as behavioral and mental health.

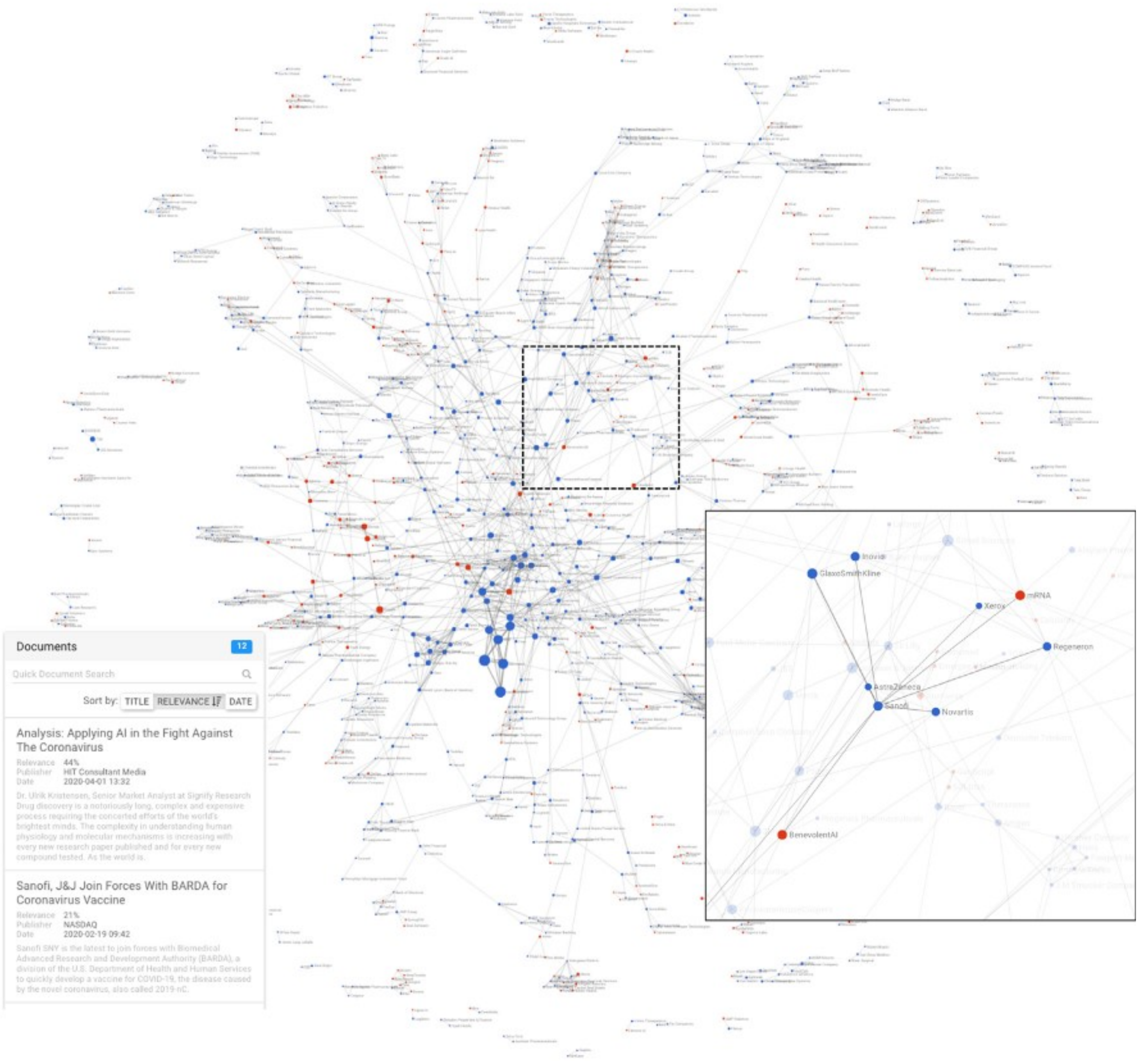

Fig. 6. Company and venture network visualization (with an interactive focused selection on one firm). Relevant documents to the selection are provided in the details panel. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 7. Business issues associated with AI and COVID19. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

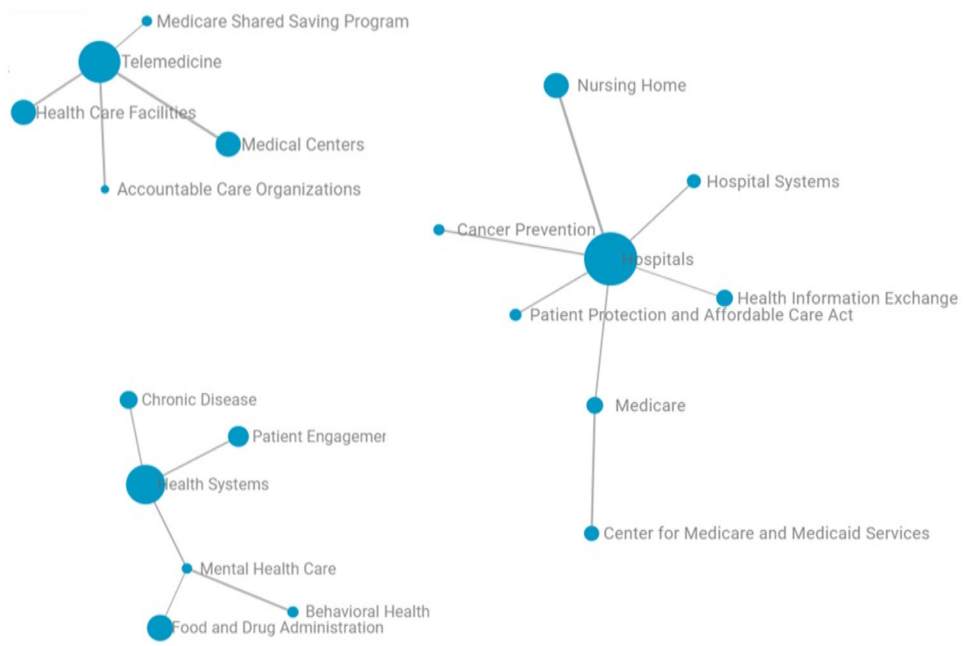

Fig. 8. Healthcare entities associated with AI and COVID19.

5.4. Summary

An important precursor to ecosystem decision making is the gathering, digesting, and understanding of relevant data to a question at hand. The ecosystem intelligence cycle thus involves the process of turning raw information into relevant intelligence for analysts and executives to use in subsequent decision making. The three use cases illustrate this from different vantage points.

The focus of use case 1 was to understand the narrative around an organization of interest. The intelligence inquiry was thus guided by the question: What companies, offerings, and technologies are shaping the narrative surrounding OpenAI? By identifying relevant documents, extracting relevant entities, and visually mapping their relationships, this case study was able to highlight that large hyperscalers are prominent actors in the OpenAI ecosystem as well as important enabling companies. Executives and strategic decision makers are often interested in organization-focused questions to gain an understanding of strategic actions and positioning as well as competitive behaviors.

The focus of use case 2 was to understand the narrative around a technology of interest and industries that are primarily associated with it. The intelligence inquiry was thus guided by the question: What industries are primarily discussed in association with the metaverse today? By identifying relevant documents, extracting relevant entities, and visually mapping their relationships, this case study was able to confirm common perceptions market and technology analysts may have about relevant industries and also guide investors in considering new emerging areas. Specifically, the results showed that in addition to games and digital content, the broader entertainment industry as well as the travel and tourism industry are significant participants in the metaverse discussion today.

The focus of use case 3 was to consider the narrative associated with both a technology and a specific context, namely artificial intelligence as it relates to COVID-19. This question can be of interest to both technology and healthcare executives as well as policy makers. The intelligence inquiry was thus motivated by the following question: What are key trends or issues and what companies, technologies, or initiatives are associated with AI and COVID-19? By identifying relevant documents, extracting relevant entities, and visually mapping their relationships, this case study was able to reveals different topic and entity clusters of interest, including those how AI has helped drug development, impacted healthcare systems and organizations, and its connection to regulatory entities.

6. Implications

The combination of text mining and visual analytics for business intelligence has many exciting opportunities for both research and practice. The system presented in this paper provides merely a step in this direction. In the following section we discuss implications and provide an overview of potential extensions of our work.

6.1. Visualization for business intelligence

The use of visualization as an empirical method is actually surprisingly limited, not just in information systems research, but management and organization science research in general. While early studies have advocated the use of visualizations to study socio, economic, and technical systems [82], most studies have used it primarily to communicate and report summarized results (e.g., graphs and charts). Yet, with the advent of big data, characterized as high volume, variety, velocity, and veracity, and advances in user interfaces, interactive visualizations can play an equally important role in the early exploration, discovery, sensemaking, and analysis process. In fact, visualizations can serve both as hypothesis confirming and generating.

One of the most striking shortcomings is that the majority of visualizations in management research continue to be of static nature. Static graphical representations prevent users to probe, explore, or manipulate data interactively. They are often served as a means to communicate or report a finding or story, thus limiting the consumers ability to "dig deeper". We should note that we do not necessarily advocate that all visualizations should be interactive, far from it, but we do believe that there is a significant potential for advancing information systems research for the introduction of interactive visual artifacts.

Recent research in design-science driven information systems has started to address this issue. Our study is one example. Yet, most interactive visual analytic studies of business data come from the information visualization community [58]. One reason is that the design and development of such artifacts, rather than reliance on off-the-shelf systems, may require competencies that fall outside traditional IS education. However, as the need for visual decision support continues to grow, the need for designing and developing appropriate and potentially novel computational and graphical techniques is warranted. Rather than evaluating these decision support systems after the fact, there is a tremendous opportunity for researchers to help influence the design and development of such systems early on.

6.2. Tuning the text mining results

The value of any visual analytic system is largely driven by the appropriateness and quality of the data. It is thus reasonable to state that if text mining results could be further improved, the value of our system is likely to increase as well. There are several ways we can imagine extending our text mining approach and results.

6.2.1. Granularity of entity co-occurrence

In the current implementation, we consider two entities to be co-occurring if they appear within the same document. Yet, it may be prudent to allow the granularity of the co-occurrence to vary from the document (highest) to paragraph or perhaps even sentence (lowest) level. Allowing users to dynamically specify whether the co-occurrence should appear at sentence, paragraph, or document level could provide greater fidelity. The more granular you go, however, the more likely it is that there will be insufficient matches given the over-constraint. Best practice in text mining suggests that if there are sufficient number of documents, co-occurrence at the document level does provide sufficiently high quality results. What may be missed, and could be captured at lower levels, is the true relation between two entities, rather than just potentially coincidental co-occurrence.

6.2.2. Identification of co-occurrence type and strength of relationships

In our current approach, we use the frequency of entity co-occurrence to determine the strength (thickness) of links between them. At the same time, however, one could imagine a much richer description of the nature of the links and the attributes that it contains beyond the frequency of co-occurrence. For instance, links between firms could be depicted by the type of relationship that has been identified. If the link has a quantitative nature, say investment, the link could be scaled by the amount of investment/funding, the knowledge that is exchanged, etc. In fact, relationships between firms are likely highly multi-attribute in nature. An exciting and challenging research area would be to specify, identify, and extract specific relationship of interest (e.g., "who is partnering with [Y] on innovation?") In order to do this effectively, novel text analytic heuristics may have to be developed. If pursued successfully, RDF-triples (subject–predicate–object) may provide the foundation for even richer ecosystem intelligence analyses.

6.2.3. Specification of data sources and scenarios

Specification of specific data sources, rather than entire data repositories may also allow for more granular analysis. Potential future research opportunities associated with ecosystem intelligence include the specification of pre-defined analysis scenarios, rather than open query explorations. Decision makers may have periodic tasks to complete, such as updating the competitive landscape every quarter or identifying a particular issue for different entities (e.g. "What technologies are most commonly associated with [X]?"). An ability to specify and/or select such questions from a pre-populated menu can accelerate time to insight.

6.3. Generative AI

The emergence of large language models and generative AI technologies, such as ChatGPT, promise to create some exciting new directions for unstructured data-centered ecosystem intelligence [83]. Generative pre-trained transformer (GPT) models are trained on large variety of data including Common Crawl, webtexts, books, images, and Wikipedia. GPT-4 for instance is capable of answering complex questions, producing narratives, generating computer code, translating between languages, and performing calculations, among many other things. Given its tremendous power and potential promise, domain-specific GPT models are also starting to emerge, including BloombergGPT, a LLM trained on comprehensive financial data by Bloomberg. One of the challenges for true competitive intelligence, however, is that to the best of our knowledge these LLMs are often trained on data up to a given time in the past and struggles to accurately reason on current events and activities. Yet, this seems like only a temporary issue.

Generative AI capabilities for ecosystem intelligence could for instance enhance the search process by allowing for more natural language queries. Rather than using specific keywords only, it could understand the semantics of questions or act as a co-pilot to iteratively recommend relevant questions to an inquiry. Generative AI capabilities could provide fine-tuned summaries of both retrieved data sources at both the individual document as well as entire corpus level. If trained on visualization layouts, generative AI models could recommend different visual representations and tune layout algorithm parameters to adjust the visual. Visualizations could be enhanced through complementary textual annotations. Some of these capabilities are already emerging; many others are likely to follow. Unquestionably, generative AI presents exciting opportunities for competitive intelligence and are a natural future extension of our system.

7. Concluding remarks

Business ecosystem intelligence is increasingly viewed as an essential aspect of corporate strategy and a topic of growing interest to both scholars and practitioners. The ability to rapidly make sense of companies, industries, markets, and trends requires the design and development of new tools that can leverage the wealth of textual data and computationally transform these into human-actionable insights. In this paper, we present an interactive visual analytics system that leverages text mining and visual analytics to provide analysts, decision makers, and scholars an intuitive UI to explore and discover the public sentiment around free-from queries of interest. Methodologically, we demonstrate that visual analytics, an emerging field that fuses information visualization and data analytics, combined with artificial intelligence can significantly boost human cognition for disentangling patterns from unstructured data.

CRediT authorship contribution statement

Rahul C. Basole: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing – original draft, Supervision, Project administration, Funding acquisition. **Hyunwoo Park:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Writing – original draft. **C. David Seuss:** Resources, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgment

The authors would like to thank the editor and two anonymous reviewers for providing constructive feedback. We would like to thank Ms. Yeeun Choi for her assistance in refining figures and formatting the text in the revision stage. Hyunwoo Park received funding acknowledged as follows. This work was supported by the New Faculty Startup Fund from Seoul National University. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2022R1C1C1011888).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.dss.2023.114133.

References

[1] R.C. Basole, Visual analytics for entrepreneurship research, in: Big Data Directions in Entrepreneurship Research: Researcher Viewpoints, 2021, p. 9.
[2] R.C. Basole, D. Armbruster, N. Cortez, B. Barnett, F. Guilak, K. Kempf, Computational modeling of business ecosystem dynamics, in: Proceedings of the 56th Hawaii International Conference on System Sciences, IEEE Computer Society, 2023, pp. 1334–1338.
[3] B.R. Iyer, R.C. Basole, Visualization to understand ecosystems, Commun. ACM 59 (11) (2016) 27–30.
[4] M.G. Jacobides, In the ecosystem economy, what's your strategy? Harv. Bus. Rev. 97 (5) (2019) 128–137.
[5] G. Parker, M.V. Alstyne, X. Jiang, Platform ecosystems: How developers invert the firm, MIS Q. 41 (1) (2017) 255–266.
[6] W.B. Rouse, Modeling and Visualization of Complex Systems and Enterprises, Wiley, 2015.
[7] R.C. Basole, Visual analytics for innovation and R&D intelligence, Res.-Technol. Manag. 66 (3) (2023) 38–50.
[8] P. Weill, S.L. Woerner, Is your company ready for a digital future? MIT Sloan Manag. Rev. 59 (2) (2018) 21–25.
[9] R.C. Basole, Understanding ecosystem data, in: Proceedings of the 53rd Hawaii International Conference on System Sciences, IEEE Computer Society, 2020.
[10] C. Holsapple, A. Lee-Post, R. Pakath, A unified foundation for business analytics, Decis. Support Syst. 64 (2014) 130–141.
[11] V.B. Kobayashi, S.T. Mol, H.A. Berkers, G. Kismihók, D.N. Den Hartog, Text mining in organizational research, Organ. Res. Methods 21 (3) (2018) 733–765.
[12] W.H. Inmon, A. Nesavich, Tapping Into Unstructured Data: Integrating Unstructured Data and Textual Analytics Into Business Intelligence, Pearson Education, 2007.
[13] R.C. Basole, Visual business ecosystem intelligence: Lessons from the field, IEEE Comput. Graph. Appl. 34 (5) (2014) 26–34.
[14] B.F.K. Norambuena, T. Mitra, C. North, A survey on event-based news narrative extraction, ACM Comput. Surv. 55 (14s) (2023) 1–39.
[15] P.T. Roundy, M.A. Bayer, Entrepreneurial ecosystem narratives and the micro-foundations of regional entrepreneurship, Int. J. Entrep. Innov. 20 (3) (2019) 194–208.
[16] S. Hubner, F. Most, J. Wirtz, C. Auer, Narratives in entrepreneurial ecosystems: drivers of effectuation versus causation, Small Bus. Econ. 59 (2022) 211–242.
[17] H.P. Abbott, The Cambridge Introduction to Narrative, Cambridge University Press, 2020.
[18] G. Phillips-Wren, M. Daly, F. Burstein, Reconciling business intelligence, analytics and decision support systems: More data, deeper insight, Decis. Support Syst. 146 (2021) 113560.
[19] F. Burstein, C. W Holsapple, B. Zhu, H. Chen, Information visualization for decision support, in: Handbook on Decision Support Systems 2: Variations, Springer, 2008, pp. 699–722.
[20] E. Dimara, J. Stasko, A critical reflection on visualization research: Where do decision making tasks hide? IEEE Trans. Vis. Comput. Graphics 28 (1) (2021) 1128–1138.
[21] S. Bina, T. Kaskela, D.R. Jones, E. Walden, W.B. Graue, Incorporating evolutionary adaptions into the cognitive fit model for data visualization, Decis. Support Syst. 171 (2023) 113979.
[22] J. Baker, D. Jones, J. Burkman, Using visual representations of data to enhance sensemaking in data exploration tasks, J. Assoc. Inf. Syst. 10 (7) (2009) 2.
[23] M. Gentzkow, B. Kelly, M. Taddy, Text as data, J. Econ. Lit. 57 (3) (2019) 535–574.
[24] W. Fan, L. Wallace, S. Rich, Z. Zhang, Tapping the power of text mining, Commun. ACM 49 (9) (2006) 76–82.
[25] M.L. Jensen, J.M. Averbeck, Z. Zhang, K.B. Wright, Credibility of anonymous online product reviews: A language expectancy perspective, J. Manage. Inf. Syst. 30 (1) (2013) 293–324.
[26] A. Moreno, C. Terwiesch, Doing business with strangers: Reputation in online service marketplaces, Inf. Syst. Res. 25 (4) (2014) 865–886.
[27] P.V. Singh, N. Sahoo, T. Mukhopadhyay, How to attract and retain readers in enterprise blogging? Inf. Syst. Res. 25 (1) (2014) 35–52.
[28] A. Ghose, P.G. Ipeirotis, B. Li, Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content, Mark. Sci. 31 (3) (2012) 493–520.
[29] S. Stieglitz, L. Dang-Xuan, A. Bruns, C. Neuberger, Social media analytics, Bus. Inf. Syst. Eng. 6 (2) (2014) 89–96.
[30] Y. Pan, P. Huang, A. Gopal, Storm clouds on the horizon? New entry threats and R&D investments in the US IT industry, Inf. Syst. Res. 30 (2) (2019) 540–562.
[31] R.C. Basole, T. Clear, M. Hu, H. Mehrotra, J. Stasko, Understanding interfirm relationships in business ecosystems with interactive visualization, IEEE Trans. Vis. Comput. Graphics 19 (12) (2013) 2526–2535.
[32] W. Li, H. Chen, J.F. Nunamaker Jr., Identifying and profiling key sellers in cyber carding community: AZSecure text mining system, J. Manage. Inf. Syst. 33 (4) (2016) 1059–1086.
[33] S. Ludwig, K. De Ruyter, D. Mahr, M. Wetzels, E. Brüggen, T. De Ruyck, Take their word for it, MIS Q. 38 (4) (2014) 1201–1218.
[34] A.H. Huang, R. Lehavy, A.Y. Zang, R. Zheng, Analyst information discovery and interpretation roles: A topic modeling approach, Manage. Sci. 64 (6) (2018) 2833–2855.
[35] T. Loughran, B. McDonald, Textual analysis in accounting and finance: A survey, J. Account. Res. 54 (4) (2016) 1187–1230.
[36] T.R. Hannigan, R.F. Haans, K. Vakili, H. Tchalian, V.L. Glaser, M.S. Wang, S. Kaplan, P.D. Jennings, Topic modeling in management research: Rendering new theory from textual data, Acad. Manag. Ann. 13 (2) (2019) 586–632.
[37] M. Wedel, P. Kannan, Marketing analytics for data-rich environments, J. Mark. 80 (6) (2016) 97–121.
[38] L.W. Lee, A. Dabirian, I.P. McCarthy, J. Kietzmann, Making sense of text: Artificial intelligence-enabled content analysis, Eur. J. Mark. (2020).
[39] Y. Jin, Y. Matsuo, M. Ishizuka, Extracting inter-firm networks from worldwideweb, in: The 9th IEEE International Conference on E-Commerce Technology and the 4th IEEE International Conference on Enterprise Computing, E-Commerce and E-Services (CEC-EEE 2007), IEEE, 2007, pp. 635–642.
[40] G. Hoberg, G. Phillips, Product market synergies and competition in mergers and acquisitions: A text-based analysis, Rev. Financ. Stud. 23 (10) (2010) 3773–3811.
[41] G. Hoberg, G. Phillips, Text-based network industries and endogenous product differentiation, J. Polit. Econ. 124 (5) (2016) 1423–1465.
[42] O. Netzer, R. Feldman, J. Goldenberg, M. Fresko, Mine your own business: Market-structure surveillance through text mining, Mark. Sci. 31 (3) (2012) 521–543.
[43] J.F. Moore, The Death of Competition: Leadership and Strategy in the Age of Business Ecosystems, HarperCollins, 1996.
[44] D.J. Teece, Business ecosystem, in: The Palgrave Encyclopedia of Strategic Management, Springer, 2016, pp. 1–4.
[45] R.C. Basole, Visualization of interfirm relations in a converging mobile ecosystem, J. Inf. Technol. 24 (2) (2009) 144–159.
[46] R.C. Basole, W.B. Rouse, Complexity of service value networks: Conceptualization and empirical investigation, IBM Syst. J. 47 (1) (2008) 53–70.
[47] R.C. Basole, J. Huhtamäki, K. Still, M.G. Russell, Visual decision support for business ecosystem analysis, Expert Syst. Appl. 65 (2016) 271–282.
[48] R. Alt, Managing AI is managing complexity - An interview with Rahul C. Basole, Electron. Mark. (2022) 1–7.
[49] R. Adner, Ecosystem as structure: An actionable construct for strategy, J. Manag. 43 (1) (2017) 39–58.
[50] M. Bogers, J. Sims, J. West, What is an ecosystem? Incorporating 25 years of ecosystem research, in: Academy of Management Proceedings, 2019, pp. 1–29.
[51] M.G. Jacobides, C. Cennamo, A. Gawer, Towards a theory of ecosystems, Strateg. Manag. J. 39 (8) (2018) 2255–2276.
[52] A. Shipilov, A. Gawer, Integrating research on interorganizational networks and ecosystems, Acad. Manag. Ann. 14 (1) (2020) 92–121.
[53] E. Autio, L. Thomas, Innovation ecosystems, in: The Oxford Handbook of Innovation Management, Oxford University Press, Oxford, UK, 2014, pp. 204–288.
[54] K. Järvi, S. Kortelainen, Taking stock of empirical research on business ecosystems: A literature review, Int. J. Bus. Syst. Res. 11 (3) (2017) 215–228.
[55] P.K. Senyo, K. Liu, J. Effah, Digital business ecosystem: Literature review and a framework for future research, Int. J. Inf. Manage. 47 (2019) 52–64.
[56] M. Tsujimoto, Y. Kajikawa, J. Tomita, Y. Matsumoto, A review of the ecosystem concept: Towards coherent ecosystem design, Technol. Forecast. Soc. Change 136 (2018) 49–58.
[57] M. Iansiti, R. Levien, The Keystone Advantage: What the New Dynamics of Business Ecosystems Mean for Strategy, Innovation, and Sustainability, Harvard Business Press, 2004.

[58] R.C. Roberts, R.S. Laramee, Visualising business data: A survey, Information 9 (11) (2018) 285.
[59] R.C. Basole, A. Srinivasan, H. Park, S. Patel, Ecoxight: Discovery, exploration, and analysis of business ecosystems using interactive visualization, ACM Trans. Manag. Inf. Syst. 9 (2) (2018) 1–26.
[60] A. Faber, S.-V. Rehm, Towards design principles for visualizing business ecosystems, in: International Conference on Enterprise Information Systems, Springer, 2019, pp. 299–319.
[61] M. Wattenberg, Visualizing the stock market, in: CHI'99 Extended Abstracts on Human Factors in Computing Systems, 1999, pp. 188–189.
[62] C.S. Merino, M. Sips, D.A. Keim, C. Panse, R. Spence, Task-at-hand interface for change detection in stock market data, in: Proceedings of the Working Conference on Advanced Visual Interfaces, 2006, pp. 420–427.
[63] R.C. Basole, Topological analysis and visualization of interfirm collaboration networks in the electronics industry, Decis. Support Syst. 83 (2016) 22–31.
[64] H. Park, M.A. Bellamy, R.C. Basole, Visual analytics for supply network management: System design and evaluation, Decis. Support Syst. 91 (2016) 89–102.
[65] P.C. Evans, R.C. Basole, Revealing the API ecosystem and enterprise strategy via visual analytics, Commun. ACM 59 (2) (2016) 26–28.
[66] R.C. Basole, Accelerating digital transformation: Visual insights from the API ecosystem, IT Prof. 18 (6) (2016) 20–25.
[67] R.C. Basole, H. Park, R.O. Chao, Visual analysis of venture similarity in entrepreneurial ecosystems, IEEE Trans. Eng. Manage. 66 (4) (2018) 568–582.
[68] R.C. Basole, S.S. Patel, Transformation through unbundling: Visualizing the global FinTech ecosystem, INFORMS Serv. Sci. 10 (4) (2018) 379–396.
[69] R.C. Basole, H. Park, Interfirm collaboration and firm value in software ecosystems: Evidence from cloud computing, IEEE Trans. Eng. Manage. 66 (3) (2018) 368–380.
[70] R.C. Basole, Visualizing interfirm collaboration in the microservices ecosystem, in: Proceedings of the 52nd Hawaii International Conference on System Sciences, IEEE Computer Society, 2019.
[71] R.C. Basole, The ecosystem of machine learning methods, in: Proceedings of the 54th Hawaii International Conference on System Sciences, 2021, pp. 5871–5881.
[72] R.C. Basole, Visualizing the evolution of the AI ecosystem, in: Proceedings of the 54th Hawaii International Conference on System Sciences, IEEE Computer Society, 2021, pp. 6193–6203.
[73] R.C. Basole, H. Park, B.C. Barnett, Coopetition and convergence in the ICT ecosystem, Telecommun. Policy 39 (7) (2015) 537–552.
[74] A. Faber, A. Hernandez-Mendez, S.-V. Rehm, F. Matthes, Visualizing business ecosystems: Applying a collaborative modelling process in two case studies, in: Proceedings of the 2018 Australasian Conferences on Information Systems, 2018, pp. 1–29.
[75] R.C. Basole, T. Major, A. Srinivasan, Understanding alliance portfolios using visual analytics, ACM Trans. Manag. Inf. Syst. 8 (4) (2017) 1–21.
[76] H. Park, R.C. Basole, Bicentric diagrams: Design and applications of a graph-based relational set visualization technique, Decis. Support Syst. 84 (2016) 64–77.
[77] A. Srinivasan, H. Park, A. Endert, R.C. Basole, Graphiti: Interactive specification of attribute-based edges for network modeling and visualization, IEEE Trans. Vis. Comput. Graphics 24 (1) (2017) 226–235.
[78] T. Major, R.C. Basole, Graphicle: Exploring units, networks, and context in a blended visualization approach, IEEE Trans. Vis. Comput. Graphics 25 (1) (2018) 576–585.
[79] R.C. Basole, Mining logomaps for ecosystem intelligence, in: Proceedings of the 54th Hawaii International Conference on System Sciences, 2021, pp. 1081–1091.
[80] J.B. Kruskal, On the shortest spanning subtree of a graph and the traveling salesman problem, Proc. Amer. Math. Soc. 7 (1) (1956) 48–50.
[81] M. Bastian, S. Heymann, M. Jacomy, Gephi: an open source software for exploring and manipulating networks, in: Third International AAAI Conference on Weblogs and Social Media, 2009.
[82] S.E. Markham, The scientific visualization of organizations: A rationale for a new approach to organizational modeling, Decis. Sci. 29 (1) (1998) 1–23.
[83] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017).

Rahul C. Basole is Managing Director and Global Lead for Visualization and Interaction Science at Accenture Data & AI, focusing on new capabilities, growth strategy, and talent development. He is a globally recognized thought leader in visualization, enterprise analytics, and AI strategy, was named to Stanford University's Global List of Top 2% Scientists for both single-year and overall career impact, and his award-winning work has been published in leading management, computer science, and engineering journals and conferences. In his prior role, he was a tenured professor in the College of Computing at the Georgia Institute of Technology, a Fellow of the Batten Institute at the Darden School of Business, and a Visiting Faculty at Stanford University. He is a senior member of the ACM and IEEE. He holds a B.S. in industrial and systems engineering from Virginia Tech, a M.S. in industrial and operations engineering from the University of Michigan, and a Ph.D. in industrial and systems engineering from Georgia Tech.

Hyunwoo Park is Vice Dean for Academic Affairs and Associate Professor in the Graduate School of Data Science at Seoul National University. Before joining SNU, he was an Assistant Professor in Management Sciences at the Fisher College of Business and a Core Faculty for the Translational Data Analytics Institute (TDAI) at The Ohio State University. Before OSU, he was a postdoctoral fellow at the Tennenbaum Institute at Georgia Tech. He holds a Ph.D. in Industrial Engineering from Georgia Tech, a Master of Information Management and Systems from UC Berkeley, and a B.S. in Electrical Engineering from Seoul National University. His research interests include business and data analytics with an emphasis on visualization, supply chain management from the network perspective, and technology and innovation management in the presence of digital platforms. His research has been published in leading journals, and he has won several awards and nominations from major conferences, including INFORMS and the Academy of Management.

David Seuss is CEO of Northern Light, a Boston-based company that provides machine learning-powered knowledge management platforms for market research and competitive intelligence to global enterprises. Prior to Northern Light, Seuss was founder and CEO of Spinnaker Software Corporation, which he led from inception to a public, NASDAQ listed company. Before starting Spinnaker, Seuss was a consultant and manager for the Boston Consulting Group, running teams working with global companies on issues of corporate strategy. David has an industrial engineering degree from Georgia Tech and an MBA with High Distinction from the Harvard Business School.