

ecoxight: Discovery, Exploration, and Analysis of Business Ecosystems Using Interactive Visualization

RAHUL C. BASOLE and ARJUN SRINIVASAN, Georgia Institute of Technology
 HYUNWOO PARK, The Ohio State University
 SHIV PATEL, Georgia Institute of Technology

The term *ecosystem* is used pervasively in industry, government, and academia to describe the complex, dynamic, hyperconnected nature of many social, economic, and technical systems that exist today. Ecosystems are characterized by a large, dynamic, and heterogeneous set of geospatially distributed entities that are interconnected through various types of relationships. This study describes the design and development of *ecoxight*, a Web-based visualization platform that provides multiple coordinated views of multipartite, multiattribute, dynamic, and geospatial ecosystem data with novel and rich interaction capabilities to augment decision makers ecosystem intelligence. The design of *ecoxight* was informed by an extensive multiphase field study of executives. The *ecoxight* platform not only provides capabilities to interactively explore and make sense of ecosystems but also provides rich visual construction capabilities to help decision makers align their mental model. We demonstrate the usability, utility, and value of our system using multiple evaluation studies with practitioners using socially curated data on the emerging application programming interface ecosystem. We report on our findings and conclude with research implications. Collectively, our study contributes to design science research at the intersection of information systems and strategy and the rapidly emerging field of visual enterprise analytics.

CCS Concepts: • **Human-centered computing** → **Information visualization**; **Visual analytics**; • **Information systems** → **Decision support systems**; • **Social and professional topics** → **Computing and business**;

Additional Key Words and Phrases: Network visualization, temporal network, multipartite graph, geospatial data

ACM Reference format:

Rahul C. Basole, Arjun Srinivasan, Hyunwoo Park, and Shiv Patel. 2018. *ecoxight: Discovery, Exploration, and Analysis of Business Ecosystems Using Interactive Visualization*. *ACM Trans. Manage. Inf. Syst.* 9, 2, Article 6 (April 2018), 26 pages.
<https://doi.org/10.1145/3185047>

1 INTRODUCTION

The term *ecosystem* is used pervasively in industry, government, and academia (Adner 2017). Drawn from the biological and ecological sciences, it is a metaphor used to describe the complex, dynamic, hyperconnected nature of many social, economic, and technical systems that exist

Authors' addresses: R. C. Basole, A. Srinivasan, and S. Patel, Georgia Institute of Technology, School of Interactive Computing, 85 Fifth Street, Atlanta, GA 30332; emails: {basole, arjun010, shivpatel}@gatech.edu; H. Park, The Ohio State University, 2100 Neil Ave, Columbus, OH 43210; email: park.2706@osu.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2018 ACM 2158-656X/2018/04-ART6 \$15.00

<https://doi.org/10.1145/3185047>

today (Moore 1993; Iansiti and Levien 2004). Examples of information system (IS)-centric ecosystem contexts are plentiful, including developer ecosystems (e.g., Linux, R, Ruby-on-Rails), platform ecosystems (e.g., Apple, Google, Facebook), and innovation ecosystems (e.g., Silicon Valley).

Ecosystems are typically characterized by a large, heterogeneous set of often geospatially distributed entities that are interconnected through various types of relationships (Teece 2016). 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. It is this scale, complexity, and emergent dynamism that makes systematic understanding of ecosystems particularly challenging (Basole 2014; Basole et al. 2016; Iyer and Basole 2016).

With the remarkable increase in open, socially curated, and proprietary digital data, however, the opportunity to understand and analyze complex ecosystems has increased substantially (Basole et al. 2013, 2015b). Although existing business intelligence tools provide rich reporting functionalities, they generally provide static dashboards and lack the ability to interactively explore and analyze the interconnected structure of a business ecosystem. Moreover, existing tools are primarily geared for analysts with significant computer and analytical skills. Intuitive and easy-to-use visual analytic tools have yet to enter the so-called decision maker “C-suite” (Lohr 2014).

We followed a design science approach (Hevner et al. 2004) to develop, implement, and evaluate *ecoxight*, a Web-based system for visual exploration of the structure and dynamics of complex ecosystems. From a design science knowledge classification perspective, *ecoxight* contributes to a relatively novel and emerging problem (i.e., ecosystem structure and dynamics) with an innovative, new solution (i.e., multiple coordinated visualizations) (Goes 2014). Specifically, *ecoxight* is geared toward handling the nuances of ecosystem data, namely multipartite (i.e., different types of entities), multiplex (i.e., multiple edges between nodes), multivariate (i.e., quantitative and qualitative attributes), dynamic (i.e., node/edges and their attributes change over time), and geospatial data. Furthermore, based on a comprehensive observation of current ecosystem analysis workflows, *ecoxight* provides users with multiple analysis modes (bottom-up, top-down, and middle-out) across coordinated views, dynamic filters, flexible visual encodings, and an ecosystem playback function. We provide guidelines to our approach and present findings from our evaluation study with real-world corporate decision makers.

The remainder of the article is structured corresponding to the design science schema advocated by Gregor and Hevner (2013). Section 2 provides an overview of the problem description, a motivating context that is used as a running example, and the overarching design requirements derived from our field interviews. Section 3 presents the design and implementation of *ecoxight*, including a description of the data model, the user interface, the visual encodings used to portray ecosystems, the core interaction techniques to explore the data, and the different visualization perspectives implemented in our system. We evaluate our system using multiple user studies and present our results in Section 4. Section 5 discusses the implications of our work for ecosystem intelligence practice and theory and identifies future research opportunities. We conclude in Section 6.

2 PROBLEM DESCRIPTION

In this section, we first describe the emergence of ecosystem analysis followed by a contemporary example of an ecosystem of increasing interest. We then present findings from our field interviews with executives at global enterprises with whom we collaborated to understand their current workflows for ecosystem analysis and identify key design requirements.

2.1 From Business to Ecosystem Intelligence

Visual business intelligence has long been a topic of interest to decision makers (Jern 1999; Chen et al. 2012). Much of the solutions have focused on supporting operational issues, such as supply chain performance, employee productivity, and portfolio management. One reason for this was that data was primarily collected for internal aspects of the enterprise. Enterprises did not see the value to reveal much about themselves or share it with others. Moreover, whatever data was available was either costly to obtain or often not in a digestible form for analysis.

Today, much has changed. Digitized data about virtually every sphere of an enterprise and its competitive landscape is now readily available, either through press releases, social media, or on-line databases. Even previously hard to access data sources, such as patents or product information, have been opened up to the public. With this tsunami of digital data, decision makers now have unprecedented access to data at their fingertips. With increased competition, globalization, and hyperconnectivity, enterprises must find ways to leverage this data and make sense of their collaborative and competitive landscape (Basole 2014; Basole et al. 2017). We refer to ecosystem intelligence as the process of the defining, gathering, analyzing, and distributing of intelligence about partners, competitors, products, and any aspects of the environment needed to support strategic decision making of an organization (Basole 2017).

What differentiates ecosystem analysis from “traditional” forms of business intelligence is the complexity of the underlying data and the tempo with which decisions (real time vs. strategic) must be made. First, it is well understood that no single data source will provide complete insight into an ecosystem. Almost always, multiple data sources need to be integrated. Data sources are often disjointed, offered by different parties, highly heterogeneous, and generally have different levels of quality and resolution. Second, ecosystem analysis often requires an understanding of a heterogeneous set of entities and connections between them. Entities may include organizations, people, technologies, products, and places; relationships may be partnerships, alliances, litigations, or product offerings. The ability to identify and then connect these different entities and relationships in meaningful ways is challenging. Third, ecosystems are ultimately dynamic organisms. Entities and relationships within ecosystems change over time. The attributes associated with them evolve as well. To truly make sense of ecosystems requires decision makers to get a handle on the dynamic forces shaping them. All of these issues are domain agnostic and can be found in diverse settings, including software (Messerschmitt et al. 2005; Jansen et al. 2013), information and communication technology (Basole 2009; Basole et al. 2015a), and biotechnology/bioinformatics (Powell et al. 1996; Holzinger and Jurisica 2014), among others.

Collectively, these data characteristics provide significant challenges and opportunities for visualization research. Successful solutions must take into consideration the multiplex, multivariate, and dynamic nature of ecosystem data. It must also take into consideration the diverse user profiles that could be using the system (ranging from technically savvy analysts to nontechnical strategic managers) and facilitating different mental maps and analysis modes.

Although some solutions have been proposed (Basole et al. 2013; Heymann and Le Grand 2013), many focus only on single aspects of ecosystem analysis, such as structure, behavior, geography, or evolution. To the best of our knowledge, no one has considered all of the underlying characteristics of ecosystem data (i.e., multipartite, multivariate, multiplex, temporal, and geospatial) in an integrative, intuitive, easy-to-use, and scalable way.

2.2 Illustrative Example: The Application Programming Interface Ecosystem

We briefly introduce a real-world ecosystem, which we will use as a running example throughout this article: the application programming interface (API) ecosystem.

One of the most transformative forces shaping the business landscape today is the API, a technology that allows firms to interact and share informational assets and services with other firms (Evans and Basole 2016). APIs allow firms to scale their operations, access new markets, and enable novel services. APIs can also be combined to create new digital services called *mashups*. It has been argued that these digital relationships are increasingly replacing traditional corporate alliances as the primary means of interfirm partnerships (Iyer and Subramaniam 2015). Today, there are more than 25,000 APIs and mashups offered globally by firms in all types of industries, handling a staggering amount of API calls every day (Basole 2016; Huhtamäki et al. 2017).

Although many decision makers have realized the economic importance of APIs (some companies derive more than 50% of their revenue from APIs), the explosive growth and complexity of this ecosystem is hard to grasp without appropriate ecosystem intelligence tools. Following conversations with decision makers, executives, and investors, it is evident that there is a great desire to understand who is playing a central role, what sectoral differences may exist, what the global footprint may be, who is funding it, and how it is evolving. The API ecosystem is thus an ideal example for a contemporary ecosystem.

We draw on two publicly available data sources—ProgrammableWeb (PW)¹ and Crunchbase (CB)²—to create the API ecosystem. PW contains detailed information on 16,000+ APIs, including initial offering and deprecation date, description, developers, followers, protocols, and categories. It also contains information on nearly 9,000+ mashups that have been created using these APIs. CB is also a socially curated database that contains detailed information on entrepreneurial and investment activities of more than 50,000+ companies and 10,000+ investors worldwide, primarily in technology-oriented industries. We link these two databases using company names and URLs to generate a multiattribute, multipartite, multiplex, multivariate, geospatial temporal network. Nodes represent APIs, companies, and investors. Edges represent API offerings (company↔API), mashups between APIs (API↔API), and investments into companies (investor↔company).

2.3 Design Requirements

The design of *ecoxight* was guided by in-depth field interviews with seven global executives (three female and four male, all with 25+ years of experience), primarily from the technology, software, and manufacturing industry, who professed to struggle with making sense of the complex business ecosystems in which they are operating.³

All seven respondents commented that existing tools lack the ability to depict the multifaceted nature of ecosystems. One respondent from the semiconductor industry for instance commented: “[O]ur world is fast-clock speed and global. We are asked to create insights quickly with whatever data we have. I don’t want just a slice of the big picture, but the whole picture.” Another executive noted: “I use market research reports, Google, and online databases, like Angel.co and Crunchbase, on a weekly basis to keep abreast with what’s happening in my industry, but I’m always limited to doing individual searches and then stitching the results together manually.” In one-on-one meetings, we discovered that most executives either periodically receive static summary reports from their internal analyst groups or subscribe to market research firms for their ecosystem intelligence. However, with either approach, they felt that there were substantial limitations.

It appeared that one of the key drawbacks of these two approaches (market research or internal analysis) was that once executives had discovered something interesting in the report, they had to formulate a new query, which in turn had to be sent back to the analysts for further investigation.

¹<https://www.programmableweb.com>.

²<https://www.crunchbase.com>.

³Due to signed nondisclosure agreements, we are not allowed to reveal the identity and affiliation of our respondents.

In general, this was viewed a time-intensive process and can, in the best-case scenario, take a few hours but more commonly takes multiple days to complete pending the complexity of the query. As one VP of strategy commented: “Many of our [ecosystem] decisions are strategic or more long-term, and it’s okay to wait, but given the pace of our industry there is something to be said about getting a quick and timely response to pressing questions.” Although these comments refer to limitations of the current ecosystem intelligence process, they also reveal broad desirable ecosystem tool characteristics, including responsiveness, interactivity, and flexibility.

One important observation we made during our field interviews is that five out of seven participants perceived to have copious data at their disposal, either collected internally or accessible through third parties. They all agreed that the data was immense and complex, and what they particularly lacked was the ability to make “systemic” sense of it. One corporate strategist, for instance, commented that “the real challenge is to mash up our different data sources to gain insights.” The key term that appeared to emerge from our conversations was that executives wanted “grounded, triangulated insight” into ecosystem activities—in other words, they wanted evidence that is validated and supported by data. Of course, this is not new, as data-driven decision making is growing in many organizations. However, most of today’s decisions supported by data are operational in nature and not related to more strategic or ecosystemic data. Across all of our interviews, executives identified three overarching analysis problems: the inability to analyze multivariate ecosystem relationships, explore patterns over time, and easily import new data.

When inquiring about current tools, all executives responded that they (and their teams) used off-the-shelf statistical (e.g., R or SAS) or visualization software (e.g., Excel, Tableau, and Qlik). In the case of statistical software, they lamented that these tools primarily provide means to examine aggregate ecosystem data. Interestingly, executives who use visualization tools commented that they had “to do a lot of data wrangling before it was in any usable form.” Two executives who have had more exposure to visualization tools said that their teams used open source platforms like Gephi to depict ecosystem structure but feel frustrated with the inability to dynamically explore the data on the fly and primarily have to work with predefined visual snapshots.

Collectively, the comments revealed many challenges that decision makers face with ecosystem analyses, as well as desirable and must-have features. Using an iterative validation approach, we ultimately identified five design requirements:

- *R1: Provide “triangulated” insight.* Corporate decision makers want to make informed data-driven decisions. An ideal visualization tool should thus provide the ability to see the underlying multipartite ecosystem data from multiple perspectives (Scherr 2008).
- *R2: Understand multiplex relationships.* One of the key issues raised by decision makers is the ability to understand the nature of relationship between entities. In many instances, relationships are multiplex (i.e., multiple relations) and contain a rich set of attributes. Hence, an ideal visual ecosystem analysis tool should provide appropriate mechanisms to explore such relationships (Renoust et al. 2015).
- *R3: Understand temporal ecosystem dynamics.* A critical requirement articulated by all corporate decision makers was the need to understand changes and temporal dynamics in an ecosystem. This includes an understanding how the ecosystem grows and shrinks; what relationships are created, modified, or deleted; and which entities gain or lose prominence. In other words, a visualization tool for ecosystem analysis should support appropriate techniques to explore dynamic networks (Kerracher et al. 2014; Beck et al. 2014).
- *R4: Facilitate multiple modes of inquiry.* Corporate decision makers use several different modes of inquiry. From our observations of existing workflows, they either start with what they know and then expand their inquiry (bottom-up), would like to see the big picture and

then drill down (top-down), or choose a specific subset of the data and then either drill down or expand their analysis (middle-out). A visual ecosystem analysis tool should provide any of these three modes along with rich dynamic filtering and querying capabilities (Ahlberg and Shneiderman 1994).

- *R5: Easy-to-use and familiar design.* Although most UI developments stress the need for ease of use, it was particularly striking how concerned corporate decision makers were about the usability and familiarity of the tool (Holzinger 2005). An ideal visual ecosystem analysis tool should thus build, perhaps even mimic, best practice UI design using familiar interactions found on the Web.

3 ECOXIGHT

In this section, we present the design and implementation of *ecoxight*, a novel Web-based ecosystem visualization tool targeted for corporate executives. Rather than providing a separate related work section for each of the relevant research threads, we provide justification and connection to prior work for each of the design choices we made within the corresponding sections.

3.1 Data Model

The *ecoxight* platform stores ecosystem data as a network. Node and edge files are represented as hash maps indexed by *NodeIDs*.

3.1.1 Nodes. For each node, the default attributes include *NodeID*, *Label*, *PrimaryCategory*, *AllCategories*, and *NodeType*. *NodeID* is a unique identifier for each ecosystem node, *Label* is a textual label for each node, *PrimaryCategory* refers to the primary category class to which a node belongs (eCommerce, Search, Mapping, etc.), *AllCategories* is a list of all categories to which the node belongs, and *NodeType* defines the type of a node (API, investor, company etc.). In addition to the default attributes, nodes can also have start and end dates, geographic coordinates (defined by latitude and longitude), or any other user-defined attributes.

3.1.2 Edges. Each edge has an *EdgeID*, and a *Source*, *Target*, and *Weight* (by default, it is set to 1). Edges also have a *StartDate* and an *EndDate*. If no *StartDate* is specified, the default is set to the earliest date in the dataset. Similarly, if the *EndDate* is undefined, the default is set to the present day. In other words, an edge without dates is considered to exist “forever” with respect to the dataset. In addition to the default attributes, edges can also have *URL*, *Title*, and *Description* attributes. In addition, users can also define edge types.

3.1.3 Data Import. The *ecoxight* platform contains a data import tool that validates and processes node and edge files in CSV format and converts them into indexed hash maps that can be consumed by the system. (Note: We purposefully chose a simple CSV format for input, as many of the target users were most familiar with creating data in spreadsheet format.) If any error is encountered during data validation (missing values, an edge having a start date after a node’s end date, etc.), the tool points out the row number and column name in the node or edge file causing the error. In addition, the tool creates a metadata file with lists for node and edge attributes (both default and user defined). The metadata file also stores the dataset’s date ranges, data type of the node and edge attributes, and a flag to indicate if geographic coordinates are available for any of the nodes. This metadata file is used to dynamically populate the UI and enable or disable the geography view.

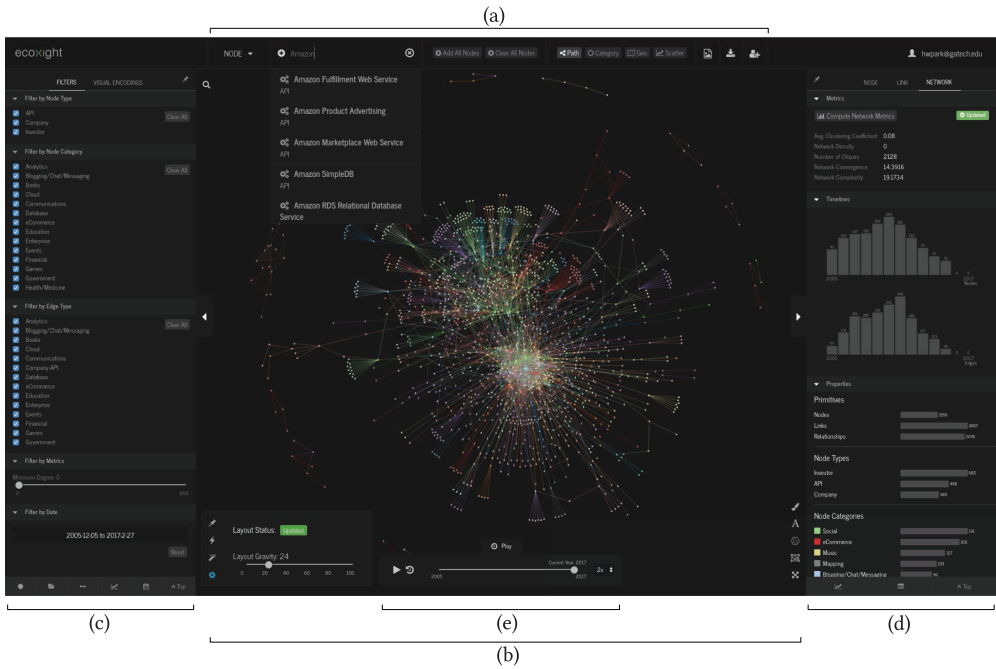


Fig. 1. The ecovisight platform UI.

3.2 User Interface

A key requirement articulated by corporate decision makers was the design of a simple, intuitive, and clean user interface (R5). We turned to industry-leading software products for inspiration, including those offered by Adobe, Microsoft, and Google, to get a sense of current best practices in UI layout and design. We then sketched multiple UI designs on paper and whiteboards and implemented wireframes to get a sense of the trade-offs we would be making in terms of the different UI elements. The researchers openly discussed the different design alternatives and sought feedback from potential users, HCI experts, and Web designers. Critical design decisions included consideration in how to organize the browser window and where to place elements including view controls, data controls, filtering capabilities, information panel, and search. The fundamental idea was to maximize the visual “workspace,” provide as many controls on demand, and accelerate use.

Following several design iterations, ecovisight’s final UI consists of five components (Figure 1): the *Data & View Selection* bar (a), the *Visualization Workspace* (b), a collapsible *Filter & Encodings* panel (c), a collapsible *Details* panel (d), and a collapsible *Playback* panel (e). We discuss each of these components in the sections that follow.

3.2.1 Data & View Selection Bar. The Data & View Selection bar (see Figure 1(a)) is the primary access point of adding data to the visualization workspace and switching between views. The data section contains a dropdown list to specify the data type of interest, namely individual nodes (Google Maps, SV Angel, etc.), node categories (eCommerce, Social, Analytics, etc.), or node types (API, company, investor, etc.). The textbox next to it provides a scrollable autocomplete list that is populated as users’ type in their query. Selecting a data entity from this list is then added to the visualization workspace. Users can also add [⊕] or remove [⊗] all nodes to the workspace. These

functionalities facilitate the multimode analysis (bottom-up, top-down, middle-out) requested by corporate decision makers in (R4).

The ecoclight platform contains four coordinated views: the *path* view [🔗], the *category* view [🔍], the *geography* view [🌐], and the *Scatternet* view [📊]. A view can be selected by clicking on the corresponding tab. The active view is indicated by a darker shading of the tab. The path view is the default view when launching the tool.

Although not directly related to data or views, the top bar also contains two additional control icons. First, we provide users the capability of downloading a screenshot of the active visualization workspace by clicking on the download icon [📥]. Second, users can select a different dataset, access their user account, or sign out by clicking on the user icon [👤].

3.2.2 Visualization Workspace. The visualization workspace (or canvas) (see Figure 1(b)) is the area where the currently selected (active) view is rendered. The visualization workspace occupies the largest space of the UI, as it reflects where the analysis and exploration is done. The two side panels can be collapsed to provide even more working space.

3.2.3 Filter & Encodings Panel. The left Filter & Encodings panel (see Figure 1(c)) contains two tabs: one for filters [🔍] and the other for visual encodings [🎨].

The filters tab presents a complete list of all qualitative, quantitative, and temporal filters. These are organized by node type, node category, edge type, network metrics, and date. The content for the filters tab is populated dynamically when the data is loaded. By default, all filtering check boxes are selected. Deselecting any of the check boxes or adjusting filter sliders dynamically updates all views in the visualization workspace. Filter selections are global and thus apply to all views.

The visual encodings tab contains a list of all visual encoding options. These include node-level encodings such as color, size, shape, opacity, and edge-level encodings such as shape, color, and opacity.

A persistent navigation bar is provided at the bottom of the panel to allow users to quickly jump to a section of the panel if the set of entries gets too long.

3.2.4 Details Panel. The Details panel (see Figure 1(d)) contains three tabs, one corresponding to each ecosystem analysis level (node, link, and network). The node tab is populated when a node is selected in the visualization workspace. It shows textual summary details about that node (type, description, location, etc.), a horizontal histogram of category distribution and connection summary, a timeline bar chart to depict relationships over time, a date-ordered list of detailed connection descriptions, and a summary of node-level metrics. The link tab shows a connection summary, timeline, and detailed description of all relationships between two nodes. The network tab shows details about the active network in the visualization workspace. This includes network-level metrics (including network density, average clustering coefficient, convergence, complexity, etc.), node and link timelines, distribution histograms for node and edge types, as well as node categories.

The content in all tabs in the Details panel is directly linked to any filters applied. Hence, any visible changes made to the canvas by any of the filters are reflected in the Details panel as well. We use brushing and linking to connect any entries in the histograms and timelines to data elements in the visualization workspace.

Similar to the left-hand panel, a persistent navigation bar is provided at the bottom of the panel to allow users to quickly jump to a section of the Details panel if set of entries gets too long.

3.2.5 Playback Panel. The Playback panel (see Figure 1(e)) contains a common timeline slider with play [▶], pause [⏸], and repeat [🔄] buttons, allowing users to navigate through time and, through using animation (Heer and Robertson 2007; Bederson and Boltman 1999; Misue et al.

1995; Eades et al. 1991), examine how the ecosystem has evolved (R3) (Ahn et al. 2011). We provide multiple speed options (i.e., 0.5x, 1x, 2x) to facilitate different evaluation tempos. The playback function applies to any of the views, and thus a multiview evolutionary perspective is provided (R1). The content in the Details panel is tied to the playback feature, providing connected insight into a structural view of ecosystems along with the summary metrics.

3.3 Visual Encodings

Following graph visualization convention, and as mentioned in Section 3.1, we represent ecosystems as networks consisting of nodes and edges. This encoding scheme corresponds to the typical mental map that corporate decision makers have of ecosystems (Basole et al. 2015b). However, as many of our prototypical users are noncomputer scientists, and thus could struggle interpreting visualizations with predefined visual mappings (Grammel et al. 2010), we follow guidelines by Huron et al. (2014) to provide users the capability to dynamically change visual encodings of the underlying data elements. Similar capabilities are found in many existing graph visualization tools (e.g., Bastian et al. (2009)).

Nodes. The ecovis platform allows users to change node shape, color, and size. Nodes can be assigned to one of five shape types (circle, cross, diamond, square, or triangle). The shape of a node is determined by the *NodeType* attribute, thus supporting our ability to visualize multipartite networks (R1). If there are more than five node types in the dataset, shapes are repeated. By default, a node's color is determined by its *PrimaryCategory*. The ecovis platform uses the Tableau 20 color palette (Stone 2012). Nodes can be colored by various node attributes, as selected through a dropdown, such as *NodeType*, *ModularityClass* (dynamically derived attribute), and *Country*. If there are more than 20 values for the selected coloring attribute, colors are repeated. Node size is defined by one of the quantitative structural or user-defined node attributes. Structural node attributes include partner count, total link count, or counts for different edge types. The node shape, color, size, and opacity encodings can be changed dynamically by changing the corresponding values in the Visual Encodings panel in the node encodings section.

In addition to these visual encodings, nodes also are visually encoded differently based on their "states." These states denote whether a node is selected/highlighted, expanded/collapsed, or active/inactive. We found these node states to be critical to give users feedback and guide them in their analysis. Figure 3(b) shows a state transition diagram for the different available active node states and the corresponding visual encoding. As you can see, all states are bidirectional to facilitate an "undo" equivalent functionality. A node is said to be inactive when it is not a part of the currently focused subnetwork or no interaction can be performed with the node. Inactive nodes are de-emphasized by fading and setting their color to a light gray.

As represented in Figure 3(b), when adding nodes to the visualization canvas, nodes have their category fill color and no stroke [a]. Users can either select the node [b], expand the node to bring its connections [c], or pin the node [d] to the canvas.

When nodes are explicitly added or expanded, their labels are preserved for emphasis and readability purposes. Otherwise, the default behavior shows node labels when a node is hovered on or a partner node is hovered upon/selected and then disappears when the selection/hover is deactivated. Labels are also preserved when a node is expanded and are removed when the node is collapsed again. Users can toggle between displaying/hiding all node labels [A] using the global controls option (see Figure 1(f)) or individual node labels using a context menu that appears by right clicking a node in any view.

Edges. By default, edges are drawn as straight lines. Users can modify the shape to a curved line in the Visual Encodings panel. We provide this option to users given the various benefits (e.g.,

aesthetics, readability) found for different contexts (Xu et al. 2012). Furthermore, by default, the edge color is the interpolation between the colors of the two nodes it connects. We provide users the ability to color edges light gray (to emphasize nodes) or change the opacity of the edges. The thickness of an edge is determined proportionally to the cumulative edge *weight*. The edge shape, color, and opacity encodings can be changed dynamically in the Visual Encodings panel in the edge encodings section.

Temporal synchronization. Both node and edge sizes change dynamically as nodes are added or removed, or as filters are modified. For example, if nodes are sized by their partner counts or the user modifies the date range, or filters out a certain category of nodes, the size of the nodes and edges that are affected by those filters will adjust. The sizing scale updates itself dynamically such that at any instance, the largest node or the thickest edge highlights the most prominent node or edge in the network. By explicitly handling temporal visual encodings, users can much easier see the evolutionary changes in the network (R5).

3.4 Interactions

Following Lam (2008), we define interaction as actions from users that cause visible changes in the visualization and interaction techniques as those actions.

Given the inherent complexity of the data and ecosystem analyses, we considered several ways of reducing interaction costs. One approach we considered was to simplify or limit the number of interactions. However, none of our proposed solutions resonated with corporate decision makers, who felt that their exploration capabilities would be severely limited. Another approach we considered was to organize interactions and make them available (incrementally) based on task context. We drew inspiration for our interaction organization from different graph task taxonomy models (Lee et al. 2006; Kerracher et al. 2015). Most importantly, we strived to make interaction consistent across all views to minimize the gulf of execution (Lam 2008).

After several design iterations with our collaborators, we ultimately agreed on three primary interactive exploration “modes” in ecoxight, namely the *Network*, *Ego Network*, and *Relationship* exploration mode.

The *Network* exploration mode lets users explore the ecosystem network, add or remove nodes, expand or collapse existing nodes, pin or unpin nodes to a position on the canvas, and navigate paths. Broadly, it can be thought of as the overview or browse mode of ecoxight. An example of the this mode is shown in Figure 1.

The *Ego Network* exploration mode is accessed when a node is selected. In this mode, a node and its direct connections (i.e., ego network) are highlighted. The selected node is considered the focus node. Nodes and edges outside of the ego network are faded into the background for de-emphasis. When activating the *Ego Network* exploration mode, the node tab in the Details panel is opened and populated with information about the selected node. If the link tab is active, hovering over any partner node shows information about the focus node and partner dyad in the link tab. We explicitly implemented this action on hover, as we found that corporate decision makers often wanted to explore different node pairs. Rather than clicking and selecting different pair of nodes, this approach facilitates quick browsing.

Last, we have the *Relationship* exploration mode. The mode is activated by selecting a focal node and then right clicking on one of its connected partner nodes. The link tab is opened and shows details about the relationship between the nodes. When there is only one relationship, the visual depiction remains the same. However, most ecosystems have multiplex relationships. One of the crucial design considerations that we thus had to make is how to best represent multiplex edges (R2). Adding them into the view by default, either through edge bundling (Holten and Van Wijk

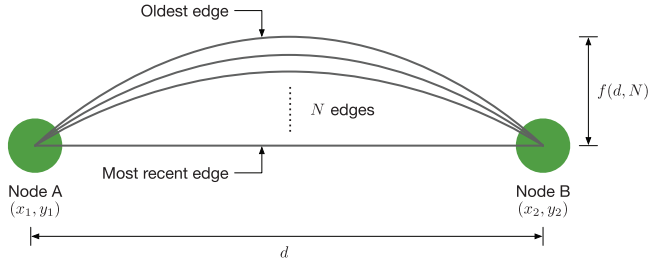


Fig. 2. Harp view. A conceptual representation of visualizing multiplex relationships.

2009) or other techniques (Riche et al. 2012), was not deemed ideal. At the same time, we were clear that whatever design alternative we ultimately chose should provide the multiplex representation in the visualization workspace rather than just in the side panel. We considered several design alternatives and drew inspiration from prior work (e.g., Renoust et al. (2015)). One option was to use multiple parallel lines spaced out either sequentially or temporally between two perpendicular axes. Although it had some appeal, the main disadvantage was the loss of a direct link to the two nodes. It also added visual overhead. To overcome this, we considered using a stepped anchor approach, but that also was not effective, as the visual did not flow smoothly and sharp angles caused users not being able to follow links. We also considered a magic lens that would appear when hovering over an edge and would show multiplex details. However, this approach was not that intuitive, and we anticipated that it would fail when the network would get dense and large.

To maintain a natural visual flow of edges between nodes, we drew an modified analogy to tree rings. Tree rings are concentric circles to depict growth of a tree. By using curved lines between two nodes, akin to an arc diagram, we provide a clean traceable link between nodes. Although we initially considered spacing edges out based on time, we realized that this may not work for node pairs with high multiplexity and maintain consistent temporal spacing for all node pairs. We instead opted for constant spacing and depicting the oldest link as the most curved (thereby indicating the greatest length) and the most recent as a straight edge. Note that this is inverse to tree rings. Moreover, rather than using curved lines on both sides, we limited it to one side to ensure that there was no confusion on the order of the edges. We refer to this technique as the *harp view*.

More formally, denote the coordinates of two nodes by (x_1, y_1) and (x_2, y_2) . Suppose that the two nodes have N edges that are sorted by date. We create N elliptical arcs between the two endpoints for all edges. An ellipsis for edge i is defined by two radii denoted by $r_{x,i}$ and $r_{y,i}$ defined as follows.

$$r_{x,i} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}, \quad r_{y,i} = r_{x,i} \left(1 - \frac{i+1}{N}\right) \quad (1)$$

Note that $r_{x,i}$ and $r_{y,i}$ are proportional to the distance between the two nodes by construction, and the set of elliptical arcs scale naturally with interactively changing node positions. We intentionally chose to use the *distance* between the two nodes as the longer *radius* of the ellipsis so that the shape does not balloon up like a circle. Note also that $r_{y,i}$ is always set to be less than $r_{x,i}$; more precisely, $r_{y,i}$ asymptotically approaches to be $r_{x,i}$ as N grows. This treatment ensures that $r_{x,i}$ is the longer radius of the ellipsis. Since $r_{x,i}$ remains constant for all i so that all elliptical arcs share the same longer-side radius. In addition, $r_{y,i}$ ranges from $r_{x,i}(\frac{N-1}{N})$ to 0 as i iterates from 0 to $N-1$. Thus, in the harp view, the straight line represents the most recent link between the two nodes, whereas the most curved elliptical arc represents the oldest link. Figure 2 illustrates the mathematical considerations behind the harp view.

Table 1. Summary of Interactions in Ecoxight

Exploration Mode	Object	Interaction	Result
<i>N EN R</i>	<i>ND EG C</i>	<i>H SC DC RC</i>	
•	•	•	Highlights a node and its ego network.
•	•	•	Enters the <i>Ego Network</i> mode with the clicked node as the focus node.
•	•	•	Expands a collapsed node (or vice versa).
•	•	•	Opens the view-specific node context menu.
•	•	•	If the link tab is open in the Details panel, it populates it with relationship information between the focus node and the hovered node.
•	•	•	Opens context menu with the “Expand Link” option. This allows entering the <i>Relationship</i> exploration mode.
•	•	•	Highlights the edge and shows a tool tip with details about the edge.
•	•	•	Highlights and freezes the selected edge.
•	•	•	Opens context menu with the “Collapse Link” option. This allows going back to the <i>Ego Network</i> exploration mode.
•	•	•	Clears all selections and system goes back into the <i>Network</i> exploration mode.

N, network; *EN*, ego network; *R*, relationship; *ND*, node; *EG*, edge; *C*, canvas; *H*, hover; *SC*, single click; *DC*, double click; *RC*, right click.

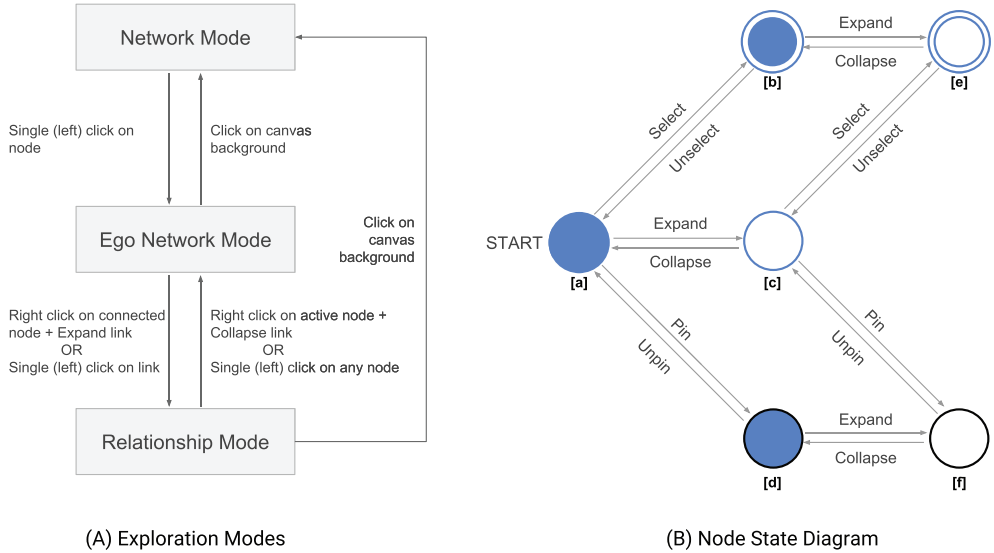


Fig. 3. The ecoxight platform provides multiple exploration modes, organized by graph task groups, to minimize the gulf of execution. Node glyphs also have varying graphical properties in terms of color and stroke to indicate different node states.

When in the *Relationship* exploration mode, hovering over any edge provides a tooltip with multivariate details and description on that edge. The same information is provided in the connected entry in the Details panel, which is linked to the multiplex edges.

The three exploration modes are complementary and provide an important differentiation of interaction contexts. Table 1 provides a summary of all node interactions supported in ecoxight, the object for which these interactions are intended, different modes in which the interactions work, and the result of each interaction. Figure 3(b) provides a conceptual representation of these three modes and their relation to each other.

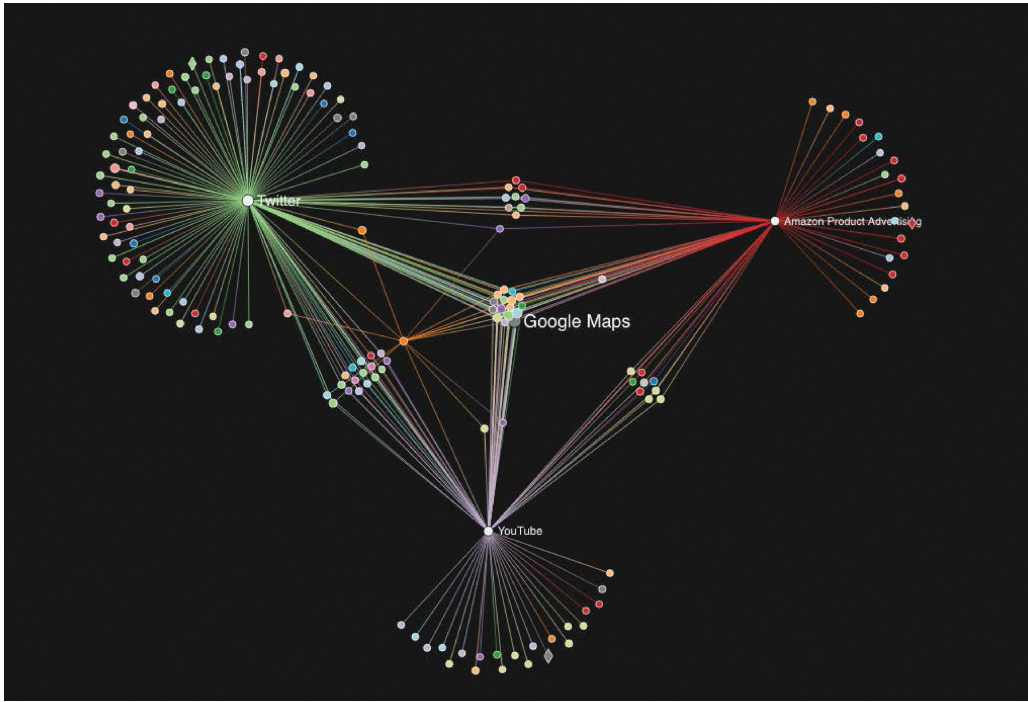


Fig. 4. The path view allows users to expand/collapse the ecosystem along a path of interest. Focal firms are pinned to the canvas while a force-directed layout arranges the remaining nodes.

3.5 Views

A key design requirement for *ecoxight* is the ability to provide “triangulated” insight, as a single view of the data is deemed insufficient for systemic analysis (R1). Building on prior work (Basole et al. 2013), *ecoxight* also provides multiple coordinated views (path, category, geography, and ScatterNet). Each of these views provide a unique, complementary topological perspective of the ecosystem, helping users explore various aspects of the underlying data. All interaction techniques described earlier, as well as zoom and pan, are available in each of the views, thereby providing a consistent user experience. A quick search function allows users to find nodes swiftly within the active view (see Figure 1(h)).

The *ecoxight* platform differentiates between view-agnostic and view-specific controls. These are placed at the right and left-bottom side of the visualization canvas, respectively (see Figure 1(f) and (g)). View-agnostic controls include showing/hiding labels [A], showing/hiding the color legend [●], showing/hiding edges [⚙], selecting subset of nodes (for inclusion or exclusion) [☐], and fitting the view to the screen [⌕]. View-specific controls are revealed by clicking on the control icon [⚙]. We describe these within each of the view sections that follow.

3.5.1 Path View. The path view is a topology-focused connectivity view, emphasizing nodes and their groups of partners, as well as shared connections across nodes (Figures 1(b) and 4). By default, it uses a cluster-emphasizing force-directed layout, positioning nodes on the workspace with approximately equal edge lengths and minimum edge crossings (Fruchterman and Reingold 1991).

In contrast to other network visualizations, our approach allows users to explicitly pin any number of nodes to any position in the workspace. Similar to organizing items on a desk, this

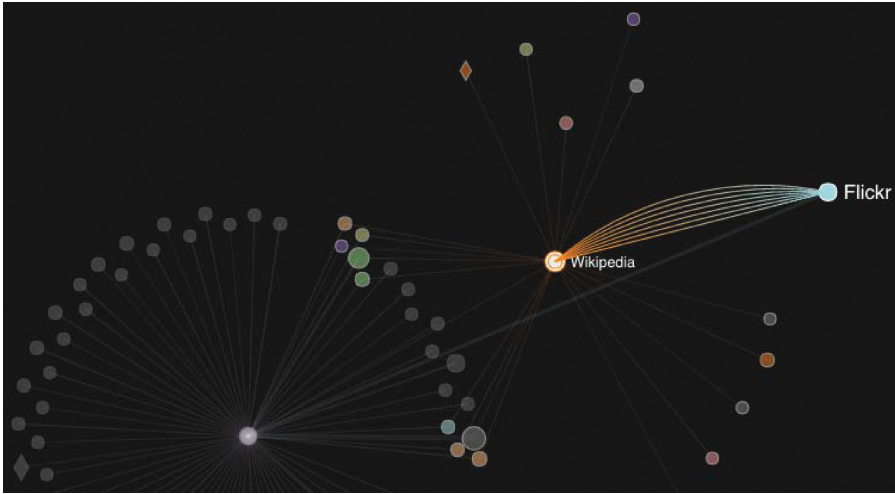


Fig. 5. Multiplex edges can be expanded to show details of each edge type using the harp view technique.

method allows users to arrange nodes spatially based on their analysis preference. The physics-based layout then adjusts the nonpinned nodes relative to the location of pinned nodes. If the active network gets too large or “unruly,” users can either fit the visualization to the workspace or dynamically adjust the layout gravity parameter (specified in the force-directed layout) through a slider in the view-specific settings panel.

Should users want to return to a force-directed layout, they can either individually unpin nodes (using a context menu populated by right clicking on a node) or unpin all nodes using a view-specific control [⚡]. The node context menu provides several additional functions, including center+focus (which focuses and zooms in on a node), node label visibility, radial reordering of its connections by category, two-step network expansion, and generation of a concentric layout (which places the focal node at the center and connecting nodes on concentric circles).

In the path view, as well as all other views, multiplex edges can be viewed by selecting a focal node and then clicking on the edge. This leads to a fan-out of the edges using the harp view approach (see Figure 5). Hovering over one of the edges provides more information in the tooltip.

3.5.2 Category View. The category view (Figure 6) positions nodes relative to their categories. Categories are represented as color-encoded arc bands. Nodes are positioned based on the centroids of the categories to which they belong. Nodes are placed positioned based on how strong the pull between a category and a node is. The center of the pull is mapped to the centroid of the category arcs. A node belonging to multiple categories will be pulled toward multiple centroids and will position itself based on the strength of the pull (determined by PrimaryCategory and AllCategories attributes). Overlapping nodes are avoided by adding a jitter to a node’s position if required. In contrast to other systems, ecobox’s category view provides users the ability to (1) sort arcs in ascending or descending order by label or by any quantitative metric and (2) size the arc by a quantitative metric such as node count (default), relationship count, total funding amount, and number of API calls (Figure 6(b)). Similar to systems such as VisAware (Livnat et al. 2005), StarGate (Ma 2008), or dotlink360 (Basole et al. 2013), we believe that the category view provides an effective visual representation of multiple category membership and node relationships.

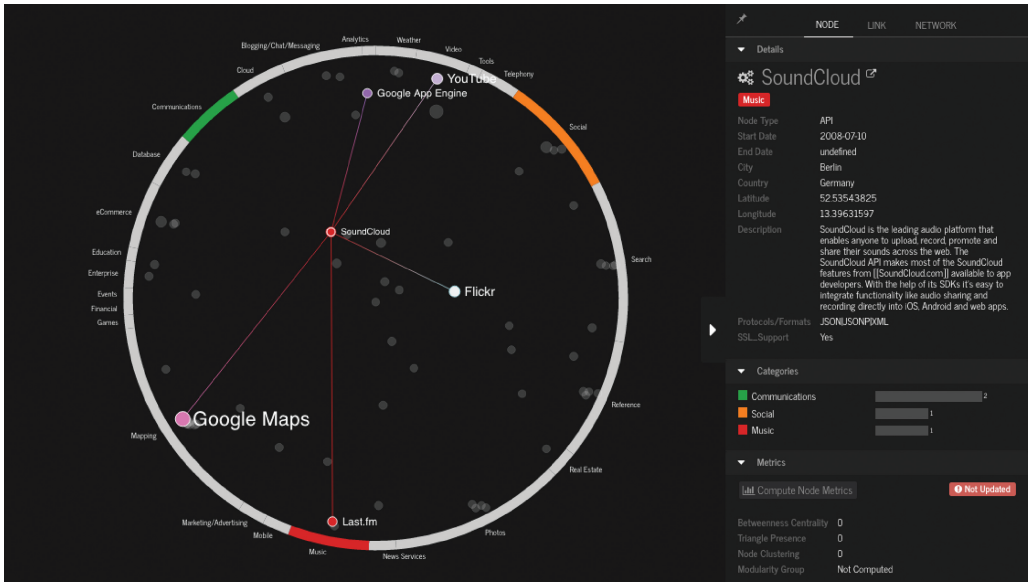


Fig. 6. The category view depicts categories as arcs. Nodes are placed inside respective to their categories. Hovering a node highlights its corresponding categories. Arcs can be reordered and sized dynamically.

Corresponding to the interaction techniques described previously, hovering over a node in the category view highlights all of its direct connections and corresponding categories while de-emphasizing others. Hovering on an arc highlights all nodes that belong to that category. Double clicking an arc brings in all nodes of all types with that category. This action is similar to using the data type dropdown in Data & View Selection bar. Filtering the network by node-level attributes or dates changes the positions of the nodes since the arc sizes dynamically change. In doing so, we provide a rich depiction of the temporal evolution of node-category memberships (R3).

3.5.3 Geography View. Many ecosystems have a geospatial component, such as location of companies and people. The geography view thus places nodes on an interactive world map corresponding to their latitude and longitude (see Figure 7). If nodes do not have any geolocation data (e.g., APIs), they either inherit this data from related nodes (e.g., companies) or are not shown in the view at all. Similar to the other views, all of the same node interactions apply. Users can zoom in/out, recenter, and pan the map by dragging anywhere on the canvas. In the case that no geolocations are available in the dataset, the geography view is not provided at all and the icon [📍] is removed from the Data & View Selection bar.

3.5.4 ScatterNet. Given that ecosystem data contains a wide variety of quantitative information, ecovisht includes an enhanced ScatterNet view, which integrates the basic scatterplot with network functionalities (see Figure 8). The benefits of attribute-related node positioning have been illustrated in systems such as PivotGraph (Wattenberg 2006), Semantic Substrates (Shneiderman and Aris 2006), and GraphDice (Bezerianos et al. 2010). Nodes in ScatterNet are positioned corresponding to their x - and y -values. A range of axis variables, either user defined or derived from the network, can be selected from the dropdown list. ScatterNet provides zoom and pan capabilities, as well as selection of axes scale (linear vs. logarithmic). Similar to the other views, all of the same node interactions apply.

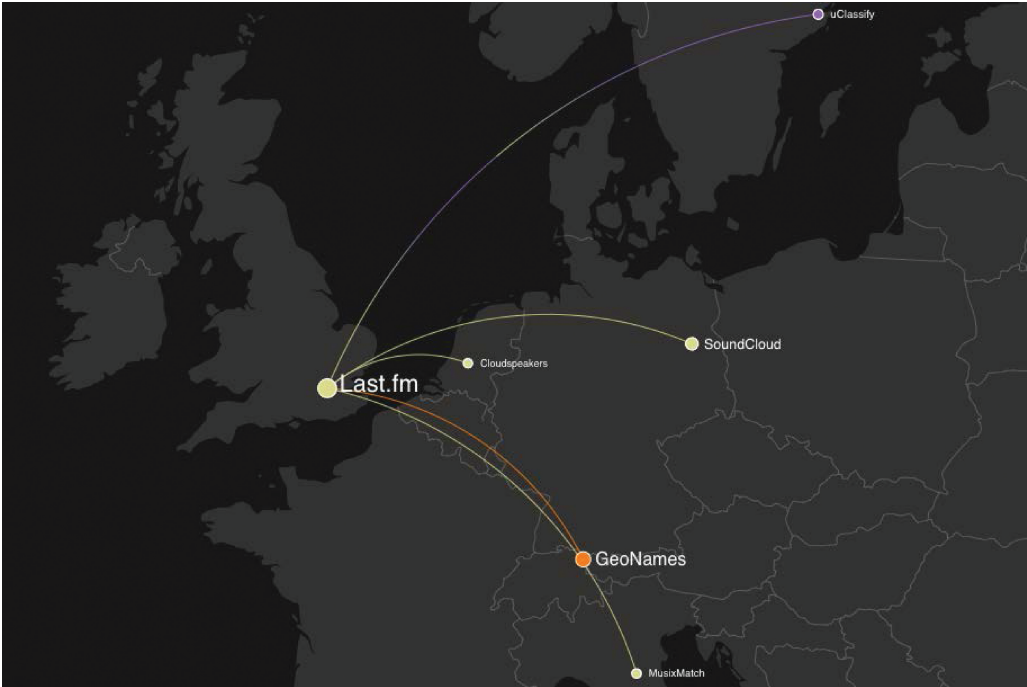


Fig. 7. The geography view provides an interactive world map. Selecting a node displays more information in the Details panel.

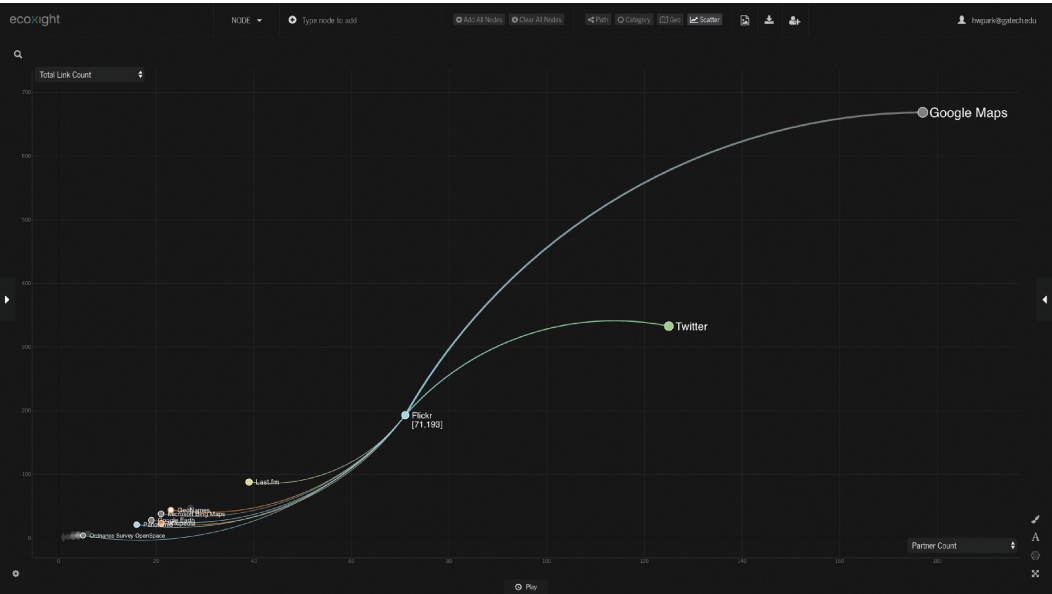


Fig. 8. ScatterNet overlays a network representation onto a traditional scatterplot. Users can zoom and pan and adjust the scale of the axes.

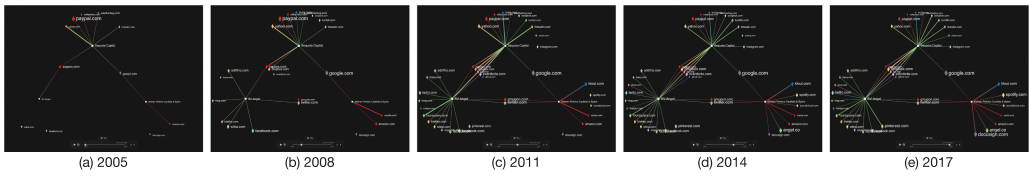


Fig. 9. The ecobox platform provides a multispeed playback mode that provides a dynamic perspective into ecosystem evolution. In this figure, we visualize the unique and shared company investments of three major venture capital firms (SV Angel, Sequoia Capital, and KPCB) over a sample of five time periods (1996–2016). Note that position (fixed for this illustration) and size/thickness of nodes, edges, and labels change dynamically.

3.6 Implementation

The ecobox platform is implemented as a Web application using the Bootstrap framework.⁴ Data imported using ecobox’s data import tool is stored in a custom NoSQL graph database. An API layer is used to extract data from the database and make it available for operations such as addition and expansion of nodes and retrieving relationship details. To ensure smooth transitions between stages of network evolution and filtering, the active network is stored in memory using hash maps indexed by *NodeIDs* during runtime. The UI elements are laid out using HTML5 and jQuery. All visualizations are developed using the D3.js library (Bostock et al. 2011).

4 USER STUDIES

In addition to the continuous feedback we received from our industry collaborators during the design of ecobox, we also conducted three formal user studies with the objective to assess both the short- and long-term usefulness and usability of ecobox for the analysis of ecosystems. A secondary goal was to assess if the design requirements described in Section 2.3 were satisfied. We chose the same aforementioned API ecosystem as our real-world study context. We opted for a predominantly qualitative value-driven user study (Stasko 2014), complemented by a survey, as we believed that broad and deep insights would be best attained through observation in combination with an open interview format.

We conducted our user evaluation in three separate stages. The first was an on-site laboratory-based user study. Users were asked to answer specific questions about the API ecosystem and conduct an open-ended analysis. The second stage was conducted entirely online and “in the wild.” Users were asked to freely use ecobox and provide qualitative comments via email or over Skype interviews. The third stage repeated the second stage but allowed users to freely use the system for an extended period of time along with using their own datasets. We opted for this triple approach to get a more complete picture of the usability and value of ecobox. The following sections present the protocol and results of each stage.

4.1 Stage 1: User Study (Laboratory)

4.1.1 Participants. Eight prototypical users (three female and five male) participated in the study. All participants had extensive experience (25+ years) as corporate decision makers in venture financing, software, technology, and research. All users had used one or more analytical tools for their work (e.g., Excel, SAS, R, Tableau). Although not required, all participants were familiar with the API economy, startups, and venture capital. All participants were potential users

⁴<http://getbootstrap.com/>.

of ecobox for their current work and were familiar with ecosystem concepts and issues. None of them had seen or tried ecobox prior to the study. Participants did not receive any compensation for completing the study.

4.1.2 Tasks and Procedure. Each participant's session was conducted individually and consisted of three phases. The first phase (tutorial+practice) focused on learning about the system. We provided an overview of ecobox and a brief description of the underlying data sources. Rather than investigators leading, we asked each participant to drive the tutorial to get familiar with the interface and click interactions. We performed a step-by-step walk-through of the various interface components and their basic functionalities. We allowed users to comment on any issues and ask questions if they did not understand a particular function. We reminded users of functions when it appeared that they had forgotten about them. This phase typically lasted between 20 and 30 minutes.

For the second phase (evaluation), we asked participants to complete a set of seven questions, with each question growing in difficulty. For example, one question was "Considering only APIs in the Enterprise and Mapping categories, identify a path between the following two APIs: DocuSign Enterprise and Mapbox." Users were allowed to use any view to answer the questions. We asked users to think aloud to understand their analysis process and note their answer on a piece of paper. Three investigators took extensive notes; we also logged click data for corroboration. After each question, we asked users to reset the views to have a fresh visualization canvas. The goal was to understand if participants had learned about the view and their operations, and if they could determine which would best help solve the problem. This phase typically lasted between 30 and 40 minutes.

In the third phase (open-ended use), we asked participants to freely explore the tool. Since all participants were familiar with the dataset and study context, each gravitated to exploring different aspects of the data. For this phase, we also asked users to think aloud, describing what they did, why they did it, and potentially also that would like to do. This phase lasted typically between 15 and 25 minutes.

We concluded the user study by asking participants to complete short five-point Likert scale (1 = strongly disagree to 5 = strongly agree) survey about their impressions of the system, and we asked them a series of interview questions about their opinions. This allowed us to corroborate their performance with their subjective assessment of the tool. Overall, each evaluation study typically lasted between 55 and 90 minutes.

4.1.3 Results. The results and feedback we received from our user study were both encouraging and helpful about the overall design of ecobox. The overarching assessment from our expert participants was that the system is highly usable and valuable for practitioners, as shown in Figure 10 and summarized in Table 2. Although we were not interested in individual views but rather the value of the system as a whole, we were particularly encouraged by the mean assessment of insight (4.3/5.0), essence (4.3/5.0), and confidence (4.0/5.0).

The "think aloud" approach and poststudy interview with the participants provided some nuanced view of opinions and also shed light into some of the unexpected findings.

Ease of use and learnability. Overall, ease of use and learnability were rated reasonably high but lower than expected. One participant commented that "ecobox is a full-fledged system, similar to Photoshop. Once I got a handle on all the features, it was really powerful. But it took a bit of time." Another noted that "[A] 15-minute tutorial really doesn't do justice to the many things ecobox can do. Had I used this for an hour more, I would have answer the questions much faster and accurately." One participant in fact said this: "The tricky part is to make systems easy to learn and

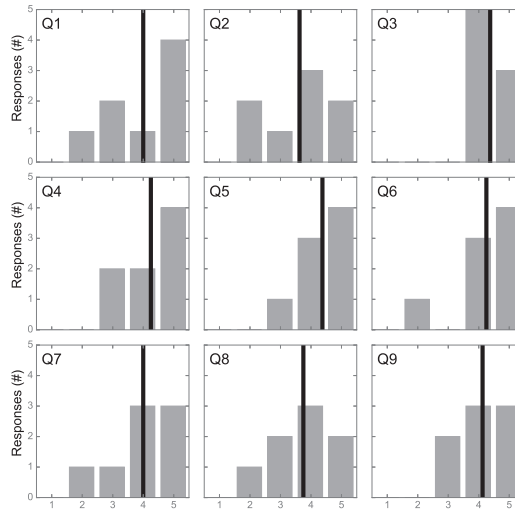


Fig. 10. Histogram of user responses (Stage 1). Black vertical bar refers to the mean. Refer to Table 2 for questions.

Table 2. User Study Results (Stage 1)

Number	Category	Question	Mean	Std. Dev.	Min.	Max.
Q1	Ease of use	ecoxight was easy to use.	4.00	1.12	2	5
Q2	Ease of use	ecoxight was easy to learn.	3.62	1.11	2	5
Q3	Insight	ecoxight enabled me to discover insights about the data.	4.38	0.48	4	5
Q4	Insight	ecoxight enabled me to ask insightful questions about the data.	4.25	0.83	3	5
Q5	Essence	ecoxight helped me generate knowledge about the data.	4.38	0.70	3	5
Q6	Essence	ecoxight conveyed an overall essence (or take-away sense) of the data.	4.25	0.97	2	5
Q7	Confidence	ecoxight helped me generate confidence about the data.	4.00	1.00	2	5
Q8	Speed	ecoxight helped me complete the given tasks quickly.	3.75	0.97	2	5
Q9	Speed	ecoxight helped me complete the given tasks effectively.	4.12	0.78	3	5

powerful. While I rated ease of use lower, it really reflects my initial use. But I don't see this as an issue long term, because I prefer a system that keep me engaged long-term. ecoxight definitely does that." Together, these findings strongly suggest that additional time is needed for mastery. All participants felt that the UI was well designed and organized and thus allowed them to easily navigate between views and panels.

Table 3. Task List (Stage 1)

Task	Correct	Accuracy
1. How many APIs does amazon.com offer?	5/8	63%
2. In what year did the Twitter API form the most connections?	6/8	75%
3. Name one mashup that uses the Facebook API and the foursquare API.	6/8	75%
4. In what segments does Kleiner Perkins Caufield & Byers invest?	6/8	75%
5. In which countries does spotify.com have investors?	4/8	50%
6. Consider only APIs in the “Enterprise” and the “Mapping” category. Tell us a path between the following two APIs “DocuSign Enterprise” and “Mapbox.”	5/8	63%
7. What APIs have mashup connections with the following three APIs: YouTube, Netflix, and Google Maps?	5/8	63%

Task performance. The evaluation phase revealed that participants were able to successfully answer most of the questions. Table 3 summarizes the results. It can be observed that the accuracy levels are lower than desired for Q1 and Q5 through Q7. One reason is that we did not include partially correct answers, which would have raised the average. We attribute the low accuracy level to Q1 to getting familiar with the system. It appeared that successfully answering some of the latter questions required more experience than was afforded by the tutorial and practice phase.

Value for ecosystem analysis. All participants unanimously agreed that ecobox was incredibly valuable for ecosystem analysis. The ability to browse companies, investors, and product (e.g., APIs) and their relationship revealed novel insights into the complexity of the ecosystem. All agreed that ecobox is really a “analysis accelerator” augmenting a decision-makers’ ecosystem intelligence and reducing time to insight. All participants requested access for their own use and identified opportunities and contexts in which ecobox may prove valuable, including patent landscapes, university ecosystems, and R&D laboratories.

4.2 Stage 2: Short-Term User Study (Field)

4.2.1 Participants. Three participants (one female and two male) participated in the field study. All three participants were senior executives with significant ecosystem experience (ranging from 15 to 30 years). Titles included managing director, vice president, and chief executive officer. All three participants work in the venture capital and startup world and were thus intimately familiar with the study context. All participants were experienced users with tools like Tableau and Quid. One of the participants had seen but not used dotlink360. None of the participants had seen or tried ecobox prior to the study. Participants did not receive any compensation for the study.

4.2.2 Tasks and Procedure. Participants were invited by email. Each participant was sent a link to the Web-based system along with a username and password. We provided an online tutorial of ecobox. Participants were asked to follow it and send us any issues or clarifying questions. We responded either via email or Skype. We then asked each participant to use the system freely for 2 weeks and provide us feedback via email.

4.2.3 Results. Participants used ecobox on average 13.5 hours during the 2-week span. The average session lasted between 1 and 2 hours. Participants used all views but heavily favored the path and segment views. Overall, participants were very enthusiastic about ecobox and its potential to transform ecosystem intelligence in a wide variety of fields. One of the striking comments we received from the participants was that ecobox helped augment and expand their working ecosystem analysis “memory,” enabling them to “investigate parts of the API ecosystem that [they]

had not considered before.” One participant, for instance, explored the interconnections between the mapping, social, and reference categories and was surprised to learn that although Google Maps is one of the mapping leading APIs, more interesting mashups were being formed with other ones. Another participant was surprised to see that almost all of the API action occurs in the United States but some prominent emerging investors are coming from Asia, suggesting a strong geographic divide. The ecovight platform also afforded participants to systematically examine the evolution of the entire or parts of the ecosystem. One participant, for instance, said this: “I was mesmerized to see visually how the growth in venture capital firms between 2008-2010 triggered an onslaught of new companies and APIs. With barcharts I only see [cumulative] growth, with this I actually see structural patterns [emerge].” The ecovight platform’s ability to provide different analysis starting points also appeared to resonate with the participants. One participant enthusiastically voiced his pleasure of being able to do the following: “[E]ither start with what [I] know and expand my horizon or start at the top and eliminate things as I explore. This aligns well with my different mode of operandi.” A concern raised by all participants, however, was the ability to continuously update the data. They all understood that new data can be imported, but for it to truly make an impact on analysis workflows, it should be tied to content pipelines or a data warehouse.

4.3 Stage 3: Long-Term User Study (Field Study)

Although the first two studies confirmed the usability and overall usefulness of ecovight, it is pertinent to understand the long-term value of our system to ecosystem intelligence tasks. To examine this issue, we conducted an extended in-the-wild study over 6 months (two quarters).

4.3.1 Participants. Four participants (one female and three male) participated in the long-term study. All four participants had significant work experience in strategy, technology, and innovation (ranging from 20 to 30 years) and conduct ecosystem analysis on a regular basis. Participants came from North America and Europe. Titles included executive director, VP strategy, director of strategy and innovation, and managing director. All of the participants had extensive experience using data-driven tools and were consumers of traditional market research reports. All participants had used ecovight prior to the study. Participants did not receive any compensation for the study.

4.3.2 Tasks and Procedure. Participants were invited by email. Each participant was sent a link to the Web-based system along with a username and password. Since all participants had used ecovight, we made the tutorial optional. We made multiple datasets (API ecosystem, startup ecosystem, and alliance networks) available to each participant and also provided template files to allow them to upload their own data. We then asked participants to use the system freely. If any questions or issues came up, they were asked to contact the investigators. At the end of the study, we requested participants to provide their feedback and comments, either via email or through a Skype call.

4.3.3 Results. Participants used ecovight on a weekly basis, with the average session lasting between 45 and 90 minutes. The session length was often longer when participants used their own datasets. Participants used all views but primarily utilized the path and scatterplot views. The playback as well as download image/graph file functions were used extensively. All participants used at least one of the predefined datasets during the study term. Overall, participants were pleased with the functionality and robustness of the system to explore ecosystem data. One participant made the following comment: “Ecosystem analysis is complex. It requires a lot of data and different perspectives. [ecovight] accelerated the process from going from a spreadsheet of data that I collected to seeing the actual structure of my ecosystem. This is powerful.” The ability to quickly

upload and visualize their own datasets resonated with all participants. Participants created 8 to 10 different datasets of varying size and complexity. One participant particularly liked the ability to add custom node attributes and dynamically select them as different visual encodings during the analysis: “[M]any firm characteristics can influence my understanding. I really liked the ability to add variables to my dataset and then size or color them for subsequent exploration. It rapidly enhanced my holistic understanding of the ecosystem.” One of the interesting observations was that participants often switched back and forth between the path and scatterplot views. In part, this confirmed our prior finding that users frequently revert back to the view with which they are most familiar (Basole et al. 2016). It also underlines the fact that each view serves their own purpose. As one participant commented: “The path view allowed me to understand complex and potentially hidden connections and identify actors of interest. Switching to the scatterplot then allowed me to see correlations between key performance indicators related to these firms.” From a usage frequency perspective, it was somewhat surprising to see that the geography view was the least used, given that ecosystems have a global footprint. However, it was evident that this was really a limitation of the geographic data specification as lat/long. Most participants had city, state, and country information but then were unable to quickly convert these into lat/long. This suggests that additional functionality in converting data (e.g., geolocation) should be considered. It was also interesting to learn that ecobox really was a starting point for many analyses. All users frequently downloaded the resulting visualizations either as an image or a graph file. The former point highlights the fact that visualizations are really used for communication purposes and are ultimately embedded in some form of presentation. The latter point underlines the fact that the implemented network layout algorithms may not be sufficient enough for some. One participant particularly knowledgeable with graph visualization tools explained this: “[W]hen networks get too large and I want to emphasize specific structural aspects, the standard force-directed algorithms fail. I need to be able to finetune them.” This feedback highlights an important point that visualizations really need to be continuously improved. Perhaps the most “liked” feature that most participants did not see before was the playback function (see Figure 9). All participants used the playback function to reinforce their understanding of the evolutionary path of their ecosystem data, as well as communicate to their colleagues the temporal nature of the context. The long-term study confirmed that ecobox can serve an important role in contemporary ecosystem analysis and fits into the overall workflow. Yet it also revealed that refinements to visual layouts and functionalities are necessary.

5 DISCUSSION AND FUTURE WORK

In this article, we have largely focused on ecobox’s interface and the system’s value derived from lab and in-the-wild usage studies. However, our learning from designing and implementing ecobox coupled with user feedback has helped us identify several other areas for further investigation.

5.1 Ecosystem Data Generation and Maintenance

From stages two and three of the user study, we identified two primary data management related tasks that need to be supported better to enhance ecobox as a more complete ecosystem intelligence platform. These include *data (network) modeling* and *data maintenance*.

Ecobox, similar to most existing network visualization tools, assumes that the data imported is in the form of network (commonly represented using node and edge tables). However, this may not always be the case. Many datasets, depending on the collection method (Web scraping, user surveys, etc.), are not inherently a network but are typically in the form of a tabular flat file. Examples of this might be a dataset of investments from a Web site like CB⁵ or a dataset of financial and

⁵<https://crunchbase.com/>.

accounting metrics from Bloomberg.⁶ In such cases, the data is typically provided as a flat CSV file and users are tasked with creating networks (typically requiring them to use programming tools) from the CSV file. This process of creating networks from such raw data is commonly referred to as network modeling (Liu et al. 2014; Heer and Perer 2014) and requires users to specify nodes and relationships in terms of data columns. This process can be resource intensive. To facilitate rapid data onboarding of such type, we are working on adding an external interface for network modeling that allows importing user created networks into ecobox's visualization canvas. Additionally, to avoid asking users to switch between network modeling and visualization interfaces, we are also exploring techniques to allow users to update an existing network model through interactions within the visualization canvas itself (Srinivasan et al. 2018).

Another critical observation from our user study was the importance of keeping data up to date and the ability to integrate with other data sources. This is a general challenge for many visual analysis platforms. One short-term solution we foresee is to link ecobox's data import interface to online data tables stored in the cloud. Although this is supported by the existing setup, it may not be very practical for users to update their data tables each time before they use the tool, as this might require them to recollect the data, remove duplicates, and so forth. A more practical solution to this challenge may be that of taking an API-based approach to establish an online data pipeline that feeds into ecobox. We have started exploring this option and are working on a separate interface that facilitates Web-based data generation and update.

5.2 Exploring Other Input Devices and Modalities

The ecobox platform is currently primarily designed for browsers running on conventional desktop and laptop contexts. However, an increasingly popular theme within the visualization research community, as well as corporate environments, as phrased by Lee et al. (2012), is to "[go] beyond the mouse and keyboard." Frequent users of ecobox have inquired about making the system more touch and mobile friendly. Although there have been systems that explore network visualizations on touch-based devices (Frisch et al. 2009; Schmidt et al. 2010), touch-based interaction for network visualization and analysis remains an open area for research (Holzinger et al. 2014). Along these lines, we are exploring the pros and cons of various elements and current interaction techniques (e.g., lack of hover in most touch devices (Jacko 2012)) in ecobox. Additionally, recent work within the visualization community has also explored the use of multimodal interaction leveraging both touch and speech for exploring network visualizations (Srinivasan and Stasko 2018). Keeping these advances and user feedback in mind, another area for future work is to expand ecobox's design and explore more naturalistic user interaction techniques with network visualizations across a variety of devices.

5.3 Supporting Multiple Exploration Modes at Scale

As highlighted in the design requirements, users wanted the capability to explore ecosystems using a top-down, bottom-up, and middle-out approach. Most existing network visualization tools, such as Gephi (Bastian et al. 2009) and Linkurious,⁷ typically support top-down or bottom-up approaches. Few, if any, existing network visualization tools facilitate an interactive middle-out exploration strategy. One may argue that effective filtering capabilities can permit a middle-out exploration strategy in a tool that supports top-down exploration. However, to support fast performance and, more importantly, interaction, effective data models must be used and decisions such as which nodes are added or removed and when nodes are added, expanded, collapsed, or filtered

⁶<https://www.bloomberg.com/>.

⁷<https://linkurio.us/>.

from the visualization need to be made. As the size of the data grows, in addition to rendering, managing in-memory data becomes an important challenge. With our current data model and loading the visualized subset of the network in-memory, we have been able to facilitate seamless interaction with up to 4,000 nodes and 12,000 edges. However, managing larger datasets, and improving the underlying data models and memory management techniques, remains an open area for improvement.

6 CONCLUSION

This study describes the design, development, and evaluation process of *ecoxight*, a visual ecosystem analysis tool that helps decision makers make sense of complex ecosystems. Our tool is ecosystem agnostic and facilitates import of any ecosystem dataset. The system provides capability for handling multipartite, multivariate, geospatial, and temporal data with rich node and edge interaction techniques. We believe that *ecoxight* has significant practical relevance for corporate decision makers, as evidenced by feedback we received through multiple user study stages. Our system overcomes many of the limitations of prior work, including handling of temporal data, dynamic filters, and Web-based access. Our design process and user evaluation also revealed promising directions for future research. Suggestions include integration with search engines, automated ecosystem creation using text analytics, deployment in collaborative and distributed environments, and interaction using novel modalities such as speech and gestures. It is our hope that this study will stimulate further research in visual ecosystem intelligence and computational enterprise science in general.

REFERENCES

- Ron Adner. 2017. Ecosystem as structure an actionable construct for strategy. *Journal of Management* 43, 1, 39–58.
- Christopher Ahlberg and Ben Shneiderman. 1994. Visual information seeking: Tight coupling of dynamic query filters with starfield displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 313–317.
- Jae-Wook Ahn, Meirav Taieb-Maimon, Awalyn Sopan, Catherine Plaisant, and Ben Shneiderman. 2011. Temporal visualization of social network dynamics: Prototypes for nation of neighbors. In *Social Computing, Behavioral-Cultural Modeling and Prediction*. Springer, 309–316.
- Rahul C. Basole. 2009. Visualization of interfirm relations in a converging mobile ecosystem. *Journal of Information Technology* 24, 2, 144–159.
- Rahul C. Basole. 2014. Visual business ecosystem intelligence: Lessons from the field. *IEEE Computer Graphics and Applications* 5, 26–34.
- Rahul C. Basole. 2016. Accelerating digital transformation: Visual insights from the API ecosystem. *IT Professional* 18, 6, 20–25.
- Rahul C. Basole. 2017. *Visual Business Ecosystem Intelligence: Framing the Challenges and Possible Solutions*. Working Paper. Georgia Institute of Technology.
- Rahul C. Basole, Trustin Clear, Mengdie Hu, Harshit Mehrotra, and John Stasko. 2013. Understanding interfirm relationships in business ecosystems with interactive visualization. *IEEE Transactions on Visualization and Computer Graphics* 19, 12, 2526–2535.
- Rahul C. Basole, Jukka Huhtamäki, Kaisa Still, and Martha G. Russell. 2016. Visual decision support for business ecosystem analysis. *Expert Systems With Applications* 65, 271–282.
- Rahul C. Basole, Timothy Major, and Arjun Srinivasan. 2017. Understanding alliance portfolios using visual analytics. *ACM Transactions on Management Information Systems* 8, 4, 12.
- Rahul C. Basole, Hyunwoo Park, and Brandon C. Barnett. 2015a. Coopetition and convergence in the ICT ecosystem. *Telecommunications Policy* 39, 7, 537–552.
- Rahul C. Basole, Martha G. Russell, Jukka Huhtamäki, Neil Rubens, Kaisa Still, and Hyunwoo Park. 2015b. Understanding business ecosystem dynamics: A data-driven approach. *ACM Transactions on Management Information Systems* 6, 2, 6.
- Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. 2009. Gephi: An open source software for exploring and manipulating networks. In *Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM'09)*. 361–362.

- Fabian Beck, Michael Burch, Stephan Diehl, and Daniel Weiskopf. 2014. The state of the art in visualizing dynamic graphs. *EuroVis STAR* 2, 1–21.
- Benjamin B. Bederson and Angela Boltman. 1999. Does animation help users build mental maps of spatial information? In *Proceedings of the 1999 IEEE Symposium on Information Visualization (Info Vis'99)*. IEEE, Los Alamitos, CA, 28–35.
- Anastasia Bezerianos, Fanny Chevalier, Pierre Dragicevic, Niklas Elmqvist, and Jean-Daniel Fekete. 2010. GraphDice: A system for exploring multivariate social networks. *Computer Graphics Forum* 29, 863–872.
- Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D³ data-driven documents. *IEEE Transactions on Visualization and Computer Graphics* 17, 12, 2301–2309.
- Hsinchun Chen, Roger H. L. Chiang, and Veda C. Storey. 2012. Business intelligence and analytics: From big data to big impact. *MIS Quarterly* 36, 4, 1165–1188.
- Peter Eades, Wei Lai, Kazuo Misue, and Kozo Sugiyama. 1991. *Preserving the Mental Map of a Diagram*. International Institute for Advanced Study of Social Information Science, Fujitsu Limited.
- Peter C. Evans and Rahul C. Basole. 2016. Revealing the API ecosystem and enterprise strategy via visual analytics. *Communications of the ACM* 59, 2, 26–28.
- Mathias Frisch, Jens Heydekorn, and Raimund Dachselt. 2009. Investigating multi-touch and pen gestures for diagram editing on interactive surfaces. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*. ACM, New York, NY, 149–156.
- Thomas M. J. Fruchterman and Edward M. Reingold. 1991. Graph drawing by force-directed placement. *Software: Practice and Experience* 21, 11, 1129–1164.
- Paulo B. Goes. 2014. Editor's comments: Design science research in top information systems journals. *MIS Quarterly* 38, 1, iii–viii.
- Lars Grammel, Melanie Tory, and Margaret Anne Storey. 2010. How information visualization novices construct visualizations. *IEEE Transactions on Visualization and Computer Graphics* 16, 6, 943–952.
- Shirley Gregor and Alan R. Hevner. 2013. Positioning and presenting design science research for maximum impact. *MIS Quarterly* 37, 2, 337–355.
- Jeffrey Heer and Adam Perer. 2014. Orion: A system for modeling, transformation and visualization of multidimensional heterogeneous networks. *Information Visualization* 13, 2, 111–133.
- Jeffrey Heer and George G. Robertson. 2007. Animated transitions in statistical data graphics. *IEEE Transactions on Visualization and Computer Graphics* 13, 6, 1240–1247.
- Alan R. Hevner, Salvatore T. March, Jinsoo Park, and Sudha Ram. 2004. Design science in information systems research. *MIS Quarterly* 28, 1, 75–105.
- Sébastien Heymann and Bénédicte Le Grand. 2013. Visual analysis of complex networks for business intelligence with Gephi. In *Proceedings of the 2013 17th International Conference on Information Visualisation*. IEEE, Los Alamitos, CA, 307–312.
- Danny Holten and Jarke J. Van Wijk. 2009. Force-directed edge bundling for graph visualization. *Computer Graphics Forum* 28, 983–990.
- Andreas Holzinger. 2005. Usability engineering methods for software developers. *Communications of the ACM* 48, 1, 71–74.
- Andreas Holzinger and Igor Jurisica. 2014. Knowledge discovery and data mining in biomedical informatics: The future is in integrative, interactive machine learning solutions. In *Interactive Knowledge Discovery and Data Mining in Biomedical Informatics*. Springer, 1–18.
- Andreas Holzinger, Bernhard Ofner, and Matthias Dehmer. 2014. Multi-touch graph-based interaction for knowledge discovery on mobile devices: State-of-the-art and future challenges. In *Interactive Knowledge Discovery and Data Mining in Biomedical Informatics*. Springer, 241–254.
- Jukka Huhtamäki, Rahul C. Basole, Kaisa Still, Martha Russell, and Marko Seppänen. 2017. Visualizing the geography of platform boundary resources: The case of the global API ecosystem. In *Proceedings of the 50th Hawaii International Conference on System Sciences*.
- Samuel Huron, Sheelagh Carpendale, Alice Thudt, Anthony Tang, and Michael Mauere. 2014. Constructive visualization. In *Proceedings of the 2014 Conference on Designing Interactive Systems*. ACM, New York, NY, 433–442.
- Marco Iansiti and Roy Levien. 2004. *The Keystone Advantage: What the New Dynamics of Business Ecosystems Mean for Strategy, Innovation, and Sustainability*. Harvard Business Press.
- Bala Iyer and Mohan Subramaniam. 2015. Corporate alliances matter less thanks to APIs. *Harvard Business Review*. Retrieved March 15, 2018, from <https://hbr.org/product/corporate-alliances-matter-less-thanks-to-apis/H024G7-PDF-ENG>.
- Bala R. Iyer and Rahul C. Basole. 2016. Visualization to understand ecosystems. *Communications of the ACM* 59, 11, 27–30.
- Julie A. Jacko. 2012. *Human Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications*. CRC Press, Boca Raton, FL.
- Slinger Jansen, Michael A. Cusumano, and Sjaak Brinkkemper. 2013. *Software Ecosystems: Analyzing and Managing Business Networks in the Software Industry*. Edward Elgar Publishing.

- Mikael Jern. 1999. Visual intelligence—turning data into knowledge. In *Proceedings of the 1999 IEEE International Conference on Information Visualization*. IEEE, Los Alamitos, CA, 3–8.
- Natalie Kerracher, Jessie Kennedy, and Kevin Chalmers. 2014. The design space of temporal graph visualisation. In *Proceedings of the Eurographics Conference on Visualization (EuroVis'15)*, Vol. 14. 7–11.
- Natalie Kerracher, Jessie Kennedy, and Kevin Chalmers. 2015. A task taxonomy for temporal graph visualisation. *IEEE Transactions on Visualization and Computer Graphics* 21, 10, 1160–1172.
- Heidi Lam. 2008. A framework of interaction costs in information visualization. *IEEE Transactions on Visualization and Computer Graphics* 14, 6, 1149–1156.
- Bongshin Lee, Petra Isenberg, Nathalie Henry Riche, and Sheelagh Carpendale. 2012. Beyond mouse and keyboard: Expanding design considerations for information visualization interactions. *IEEE Transactions on Visualization and Computer Graphics* 18, 12, 2689–2698.
- Bongshin Lee, Catherine Plaisant, Cynthia Sims Parr, Jean-Daniel Fekete, and Nathalie Henry. 2006. Task taxonomy for graph visualization. In *Proceedings of the 2006 AVI Workshop on Beyond Time and Errors: Novel Evaluation Methods for Information Visualization*. ACM, New York, NY, 1–5.
- Zhicheng Liu, Shamkant B. Navathe, and John T. Stasko. 2014. Ploceus: Modeling, visualizing, and analyzing tabular data as networks. *Information Visualization* 13, 1, 59–89.
- Yarden Livnat, Jim Agutter, Shaun Moon, and Stefano Foresti. 2005. Visual correlation for situational awareness. In *Proceedings of the 2005 IEEE Symposium on Information Visualization (INFOVIS'05)*. IEEE, Los Alamitos, CA, 95–102.
- Steve Lohr. 2014. Making big data think bigger. *New York Times*. Retrieved March 15, 2018, from <http://bits.blogs.nytimes.com/2014/09/20/making-big-data-think-bigger/>.
- Kwan-Liu Ma. 2008. Stargate: A unified, interactive visualization of software projects. In *Proceedings of the 2008 IEEE Pacific Visualization Symposium (PacificVIS'08)*. IEEE, Los Alamitos, CA, 191–198.
- David G. Messerschmitt and Clemens Szyperski. 2005. *Software Ecosystem: Understanding an Indispensable Technology and Industry*. MIT Press, Cambridge, MA.
- Kazuo Misue, Peter Eades, Wei Lai, and Kozo Sugiyama. 1995. Layout adjustment and the mental map. *Journal of Visual Languages and Computing* 6, 2, 183–210.
- James F. Moore. 1993. Predators and prey: A new ecology of competition. *Harvard Business Review* 71, 3, 75–83.
- Walter W. Powell, Kenneth W. Koput, and Laurel Smith-Doerr. 1996. Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. *Administrative Science Quarterly* 41, 1, 116–145.
- Benjamin Renoust, Guy Melancon, and Tamara Munzner. 2015. Detangler: Visual analytics for multiplex networks. *Computer Graphics Forum* 34, 321–330.
- Nathalie Henry Riche, Tim Dwyer, Bongshin Lee, and Sheelagh Carpendale. 2012. Exploring the design space of interactive link curvature in network diagrams. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*. ACM, New York, NY, 506–513.
- Maximilian Scherr. 2008. Multiple and coordinated views in information visualization. *Trends in Information Visualization* 38, 1–8.
- Sebastian Schmidt, Miguel A. Nacenta, Raimund Dachsel, and Sheelagh Carpendale. 2010. A set of multi-touch graph interaction techniques. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*. ACM, New York, NY, 113–116.
- Ben Shneiderman and Aleks Aris. 2006. Network visualization by semantic substrates. *IEEE Transactions on Visualization and Computer Graphics* 12, 5, 733–740.
- Arjun Srinivasan, Hyunwoo Park, Alex Endert, and Rahul C. Basole. 2018. Graphiti: Interactive specification of attribute-based edges for network modeling and visualization. *IEEE Transactions on Visualization and Computer Graphics* 24, 1, 226–235.
- Arjun Srinivasan and John Stasko. 2018. Orko: Facilitating multimodal interaction for visual exploration and analysis of networks. *IEEE Transactions on Visualization and Computer Graphics* 24, 1, 511–521.
- John Stasko. 2014. Value-driven evaluation of visualizations. In *Proceedings of the 5th Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*. ACM, New York, NY, 46–53.
- Maureen Stone. 2012. In color perception, size matters. *IEEE Computer Graphics and Applications* 2, 8–13.
- David J. Teece. 2016. *Business Ecosystem*. Palgrave Macmillan UK, London, England. DOI: http://dx.doi.org/10.1007/978-1-349-94848-2_724-1
- Martin Wattenberg. 2006. Visual exploration of multivariate graphs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 811–819.
- Kai Xu, Chris Rooney, Peter Passmore, Dong-Han Ham, and Phong H. Nguyen. 2012. A user study on curved edges in graph visualization. *IEEE Transactions on Visualization and Computer Graphics* 18, 12, 2449–2456.

Received March 2017; revised November 2017; accepted January 2018