TAMPERE UNIVERSITY OF TECHNOLOGY

Processing Data for Visual Network Analysis

Citation
Huhtamäki, J., Russell, M. G., & Still, K. (2016). Processing Data for Visual Network Analysis. In E. Bendoly, & S. Clark (Eds.), *Visual Analytics for Management: Translational Science and Applications in Practice* (1 ed., pp. 56-71). [5] Routledge.

Year
2016

Version
Peer reviewed version (post-print)

Link to publication
TUTCRIS Portal (http://www.tut.fi/tutcris)

Published in
Visual Analytics for Management

Copyright
This is an Accepted Manuscript of a book chapter published by Routledge in Visual Analytics for Management - Translational Science and Applications in Practice on 25 November 2016, available online: https://www.taylorfrancis.com/books/e/9781315640891

Take down policy
If you believe that this document breaches copyright, please contact cris.tau@tuni.fi, and we will remove access to the work immediately and investigate your claim.

Download date:18.11.2019

Chapter 5

Processing data for visual network analytics: innovation ecosystem experiences

Jukka Huhtamäki
Tampere University of Technology

Martha G. Russell
mediaX, Human Sciences Technology Advanced Research Institute, Stanford University

Kaisa Still
VTT Technical Research Centre of Finland

The ecosystem concept has its roots in biology. Collins English Dictionary defines an ecosystem as "a system involving the interactions between a community of living organisms in a particular area and its nonliving environment." In business and innovation literature, the ecosystem concept is used both as a metaphor (e.g. Russell et al. 2011; Russell et al. 2015; Hwang and Horowitt 2012) and as a business strategy artifact (Moore 1993), as well as to refer to system-level analysis (cf. Pentland 2015).

To make a distinction between ecosystems of business and innovation in the context of this discussion, we point to their expected outputs. When the key objective in business ecosystems is to organize value creation and value appropriation in an interdependent and dynamic setting, we see the main benefit of innovation ecosystems as the increase of information flow and collaboration and therefore the creation of new business-relevant knowledge, ideas and technologies that may lead to new products, processes and companies. Innovation ecosystems survive through a constant idea flow, re-configuration, and evolution (cf. Pentland 2015).

Key characteristics of ecosystems are interconnectedness, interdependency, co-evolution, value co-creation, and co-opetition (Huhtamäki et al. 2011; Järvi and Kortelainen 2016). Actors in business ecosystems and innovation ecosystems are loosely interconnected: Iansiti and Levien (2004) stress that "like their biological counterparts, business ecosystems are characterized by a large number of loosely interconnected participants who depend on each other for their mutual effectiveness and survival." Success of a given innovation often relies on the success of focal companies' environments, i.e. companies are interdependent with each other (Adner and Kapoor 2010). Thomas and Autio (2012) state that co-evolving ecosystem actors "develop over time sympathetically with the other participants in order to maintain stability and health of the ecosystem in the face of change." Ramaswamy (2009) claims that value co-creation is an emerging business and innovation paradigm that leads to the need of "changing the very nature of engagement and relationship between the institution of management and its employees, and between them and co-creators of value–customers, stakeholders, partners and other employees." Finally, Ritala and Hurmelinna-Laukkanen (2009) note that co-opetition, i.e. collaboration with competitors, is in some cases "an effective way of creating both incremental and radical innovations, especially in high-tech industries."

Russell et al. (2011) take an even broader scope and define innovation ecosystem as an "inter-organizational, political, economic, environmental and technological systems of innovation through which a milieu conducive to business growth is catalyzed, sustained and supported." They note that in an ecosystem, individual relationships form a network structure "through which information and talent flow through systems of sustained value co-creation." The organic role of the network structure adds to the importance of the network as a visual idiom in visual analytics. The network view allows for making use of the Gestault laws of grouping described in Chapter 1. Importantly, Russell et al. (2011) include organizational investors and individual people–founders, advisors, and business angels–as innovation ecosystem actors (and therefore potential units of analysis in investigations) (cf. Huhtamäki et al. 2011). Importantly, from visual analytics viewpoint, all of these actors approach innovation ecosystems from their individual viewpoint, each of them with various biases in sensemaking.

In this chapter, we answer the call made in Chapter 3 to develop interactive visualization tools and collaborative visual investigation processes that enable the emergence of shared vision between innovation ecosystem actors and stakeholders (Russell et al. 2011.) This process encourages the voicing of different perspectives that can “give rise to subsequent visual consideration, more targeted evaluations, additional requests to redevelop and enhance the visual platforms that the organization leverages" and facilitates scrutinizing, countering, reinforcing and even dismissing those perspectives. We introduce a new visual and modeling idiom, namely the network, for empirical innovation ecosystem investigations in which the narrative is guided by scatterplots of country or industry-level aggregate metrics constructed from established KPIs (cf. Still et al. 2012). More broadly, this investigation contributes to the development of the overall the overall Information System artifact for innovation ecosystem analytics including i) technological artifact, ii) information artifact, and iii) social artifact (Lee, Thomas, and Baskerville 2015).

Ostinato Model: Visual network analytics for innovation ecosystems

Information visualization is an integral part of the network analysis methodology. In reviewing existing work on visual network analysis, Freeman (2000) notes that visualization both allows investigators to observe and identify patterns in social structures and also supports sharing these findings with others. Visual analytics (Wong and Thomas 2004; Keim et al. 2010) extends information visualization to include the processes of visualization-centric investigations. Computational approaches permit automating various phases of the analysis process.

Data-driven visual network analytics is a key method for supporting management and decision-making through investigations of structural patterns latent in innovation ecosystems and social media phenomena. The Ostinato Model, a process model for data-driven visual network analytics, defines a structured process for investigating a network phenomenon in a data-driven manner (Huhtamäki et al. 2015). The Model was developed through several rounds of Action Design Research taking place at different levels of abstraction and complexity in innovation ecosystems. The Model allows taking a visual analytics approach investigate innovation ecosystems through networks using a computational data-driven method. It builds on several streams of literature on information visualization (Card, Mackinlay, and Shneiderman 1999), data-driven visualization pipelines (Nykänen et al. 2008), visual analytics (Wong and Thomas 2004), interactive visual analysis (Heer and Shneiderman 2012) and visual network analysis (Freeman 2000). The Model combines and extends several process models including the information visualization reference model, the visual analytics model (Keim et al. 2010), and most importantly the Network Analysis and Visualization (NAV) process model (Hansen et al. 2012).

[INSERT FIGURE 5.1 HERE]

Figure 5.1. Ostinato Model for visual network analytics (adapted from Huhtamäki et al. 2015)

Our overall perspective embodied by the Ostinato Model is in sync with Bendoly (2016). We affirm that visualization is of utmost utility in supporting the full investigative process from validation of data sources and feedback mechanism for data cleaning and aggregation, through exploration and discovery, to sharing the findings with others in a way that allows them to contextualize according to their worldviews, articulate their biases and conduct group-level sensemaking.

Key principles of the Ostinato Model are transparency, continuous data collection, exploration, loose coupling, interoperability, reproducibility, automation, enabling manual steps and low entry barrier. The model has two phases: Data Collection and Refinement; and Network Creation and Analysis. The Data Collection and Refinement phase is further divided into Entity Index Creation, Web/API Crawling, Scraping, and Data Aggregation. The Network Construction and Analysis phase is composed of Filtering in Entities, Node and Edge Creation, Metrics Calculation, Node and Edge Filtering, Entity Index Refinement, Layout Processing and Visual Properties Configuration. A cycle of exploration and automation characterizes the Model and is embedded in each phase.

Another key principle of the Ostinato Model is enabling process transparency and easy access for all members of an interdisciplinary team to data (e.g. tabular representation) in all possible phases of the analytical process. The use of standard data processing tools is recommended, including spreadsheet processors for accessing the data and prioritizes implementing many types of diagnostic and explorative analyses to support the collaborative sensemaking processes of multiple investigators as co-creators of insights.

5.1. Appreciating the Audience and Context

Innovation ecosystems are open and, more specifically, complex adaptive systems (Thomas and Autio 2012). These properties yield analytic requirements that are very difficult to measure and benchmark. To operate in this complex domain, we subscribe to Critical Realist philosophy to establish an epistemological and ontological platform for the investigations, and we agree with Dobson (2001) in acknowledging the importance of taking a philosophical stance in establishing a common platform for investigative work. In short, Critical Realist research combines realist ontology with interpretive epistemology (Bygstad and Munkvold 2011). This assumes that generalizable structures and mechanisms exist in social phenomenon. In order to identify these mechanisms and structures, however, investigators must move beyond superficial statistical measurements and case-specific qualitative observations to apply several complementing methods, both qualitative and quantitative, in the investigations. More pragmatically, we point to Bygstad and Munkvold (2011) who introduce Critical Realism as an approach to data analytics and claim that analytics processes should serve the objective of identifying the structures and mechanisms that exist within phenomenon and surface as observable events. We further agree with Dobson (2001) and Archer (1995) that social structure should be observed as dynamic, and therefore, that studying its evolution over time is important, that social structure is both a key driver of social activity, and that social activity is a key driver of the evolution of social structure.

Further, to contribute to the body of knowledge in visual analytics of innovation ecosystems and more specifically to dig deeper into the data processing requirements for the Ostinato Model, we take an Action Design Research (ADR) approach (Sein et al. 2011) to develop new analytics IT and information artifacts that support the investigative processes of the innovation ecosystems. In ADR, artifact development takes place in between researchers and the organization to whose requirements the artifacts are developed to satisfy. Guided emergence is the core process of ADR, i.e., the artifact emerges through intensive interaction rather than being developed first and then evaluated. Guided emergence takes place in Build-Intervene-Evaluate (BIE) cycles, in which concurrent evaluation is an organic part. (Sein et al. 2011).

One focal application context for our approach has been EIT ICT Labs, currently operating as EIT Digital, “a leading European open innovation organisation.” (http://www.eitdigital.eu/about-us/overview/). We joined with EIT ICT Labs management in the beginning of their operation to investigate the existing connections between the then six co-location cities (Still et al. 2014) and again after several years of activities has transpired. The Innovation Ecosystems Network Dataset (IEND), a socially constructed set of data on growth companies, their key individuals and investors was used to as the sole source of data (Rubens et al. 2010).

[INSERT FIGURE 5.2. HERE]

Figure 5.2. Interconnections between EIT ICT Labs co-locations (adapted from Still et al. 2014)

In network visualization in Figure 5.2, the EIT ICT Labs co-location cities are represented as nodes and connected to all companies having their main office in that city. The companies are further connected to each key individual in the dataset with whom they are or have been affiliated. Moreover, the companies are connected to their investors. Acquisitions form connections between individual companies. Betweenness centrality is used to size the nodes in the visualization, and nodes are laid out with Force Atlas 2 (Bastian, Heymann, and Jacomy 2009). Figure 5.3 shows the Ostinato Model variation for the EIT ICT Labs investigation.

[INSERT FIGURE 5.3 HERE]

Figure 5.3. Ostinato Model for EIT ICT Labs investigation

Using betweenness centrality as the key metric and force-driven algorithm for laying out the nodes, we were able to highlight the nodes that showed mobility in between the co-location centers. The key connecting tissue between the co-location centers was further highlighted through filtering in nodes according to their betweenness centrality value. Two key insights were achieved through the investigation. First, investors emerged as the key source of mobility in the existing EIT ICT Labs innovation ecosystem and, therefore, ways to engage the investors close to EIT ICT Labs operations needed to be investigated. Second, on basis of the visualization the inclusion of the San Francisco Bay Area as the then hypothetical seventh node of EIT ICT Labs in Figure 5.4, it was argued that the Bay Area could be, in fact, the most important connector among the six European cities in which EIT ICT Labs was operating at the time.

[INSERT FIGURE 5.4 HERE]

Figure 5.4. What if San Francisco Bay Area would be the 7th city? (adapted Still et al. 2014)

In another example, several different datasets were federated to provide a multi-dimensional perspective. Investigation of the Finnish innovation ecosystem using three different datasets in parallel and as a federated dataset added complexity to the data processing of the Ostinato Model (Still et al. 2013). Each of the three datasets - Thomson Reuters SDC, IEND Executives and Finance, and IEND Angels and Startups - addressed a different part of the innovation ecosystem, yet with some overlap. To allow for international comparability, we chose not to include data on the funding that many of the Finnish companies had received from Finnish Funding Agency for Innovation Tekes, a governmental agency that supports a significant portion of Research and Development in Finnish companies. The Ostinato Model variation in Figure 5.5 shows the process applied in the Finnish innovation ecosystem investigation. A separate data collection and refinement process was used (marked as a deck of three distinct processes) for each dataset. Moreover, the network representations were also created through separate processes. Analysis using an aggregation of the datasets was conducted only after the network representations had been constructed and validated.

The resulting multiscopic views allowed for a number of confirmatory and novel insights on the Finnish Innovation Ecosystem. Nokia's role was visible in all the different ecosystem viewpoints. A handful of key individuals played integral roles, and a key source of venture capital consisted of government-based organizations. Student-based Startup sauna, a non-institutional entity, also had a prominent role as a source of new activity.

[INSERT FIGURE 5.5. HERE]

Figure 5.5. Ostinato Model for Finnish Innovation Ecosystem investigation

5.2. Design Principles Applied

The innovation ecosystem investigations we have conducted, including the two examples described in the previous section, have required making a number of design decisions related to collecting and processing the transactional microdata, from its original form to network representations. In this section, we describe and discuss a number of data-processing related design decisions and their impact on the investigative process. For tractability, we pin the issues to the various phases of the Ostinato Model. The design issues and related steps of the Ostinato Model are included in Table 5.1.

[INSERT TABLE 5.1. HERE]

Table 5.1. Data-processing related design issues and their use in Ostinato Model phases

Boundary specification

There are several ways that investigative teams can to specify the boundaries of the innovation ecosystem under investigation (cf. Basole et al. 2012). Boundary specification indirectly impacts all the different phases of the investigative process. Importantly, the investigators should be able to experiment with various options for the boundaries to better understand their effects. Moreover, we note that in country-level investigations, we have chosen to follow the national boundaries in a fashion similar to more traditional ways of measuring innovation–through survey data, while the computational approach comes with reduced workload for crossing the boundaries.

Source data format

Useful data sources for innovation ecosystem investigations are available in various formats from websites that require crawling and scraping to machine-readable spreadsheets ready to be used for computational analysis. This chapter does not cover the wide variety of practices for data access that are available. It is, however, worth mentioning that in most investigations, we have relied on using MongoDB as means for implementing a data proxy. Key virtues of MongoDB, a NoSQL database are the flexibility of the data schema and ability to search the data.

Tabular data representation

A key principle for enabling process transparency and easy access to all members of interdisciplinary team is using tabular data representation in all possible phases of the process. This importantly allows the use of standard data processing tools, including spreadsheet processors for accessing the data, and implementing the kinds of diagnostic and explorative analysis that support investigators' sensemaking processes. While this observation or principle may seem obvious, we would like to point out the plethora of options that a skilled data scientist has available for managing the data from NoSQL and relational databases to graph databases and big data technologies, including Apache Hadoop and Spark.

Temporal data is imperative

We do want to stress the importance of the availability of transactional microdata on innovation ecosystem actors. With transactional microdata, we refer to data that includes timestamps of the activities taking place in between explicitly identified innovation ecosystem actors. Examples of activities include individuals founding a startup, serial entrepreneur joining the board, startups raising a round of venture capital from specified angel and organizational investors, firms acquiring a startup, and firms announcing an Initial Public Offering in a specified stock exchange.

Graph-based filtering

A major exception to the default rule for representing data in tabular format is cases in which boundary specification is being conducted with k-step rules (Basole et al. 2012), i.e. rules that: a) specify inclusion filter for a set of nodes; and b) further include nodes that are k-steps away from the included nodes. While we have so far implemented this part of the analysis pipeline by crawling entities on the basis of recursive inclusion rules, we point to practices available in graph databases–the likes of Cypher for Neo4j and SparQL for Resource Description Framework–to specify filtering rules based on network topology. Also, self-service tools such as Gephi implement topology-based filters that allow expressing simple boundary specification rules.

Volumetrically big data

Data that is truly big in volume does require special arrangements for its management. Additionally, a key restriction in visual network analysis is, however, the amount of pixels available for provisioning the network visualization. According to our experience, a manageable network size for visual analytics is in the scale of thousands or tens of thousands of nodes. From a technical viewpoint it is important to note that the limit of data becoming big in volume depends heavily on the selection of analytical technology including spreadsheet processors and other stock tools for exploration used by the investigators.

Multiple data sources

In the investigation of the Finnish Innovation Ecosystem (Still et al. 2013) we used complementing sets of data–both in parallel and, importantly, as an aggregate–to create a multiscopic view of the innovation ecosystem. To create the multiscopic view, we first created dataset-specific representations of the innovation ecosystem. The aggregate set of data was then composed through a process of finding the entities–in this investigation companies, investors, and individuals–that appeared in more than one dataset and creating a unique identifier for each, which was used across the datasets. We used a semi-manual process to create the identifiers for this particular investigation.

The use of multiple sets of data in parallel is for us the key venue for future development. At best, the datasets are aggregated as the very first step of the analysis process. This allows for boundary specification in a consistent way that takes into account cross-dataset connections between entities. Aggregating full sets of volumetric data, however, insists on full automation of the process. Approaches for achieving this call for applying machine learning techniques from string matching (Navarro 2001) to named entity recognition (Finkel, Grenager, and Manning 2005).

Balancing between aggregation and expressivity for filtering

The mantra "system visualizations require systems of visuals" introduced in Chapter 1 applies fully to the domain of innovation ecosystem analytics. To support decision making among individuals with divergent perspectives and objectives, the visualizations can facilitate the development of a shared vision (Russell et al. 2015), which provides an "intelligent fit" and guides decisions.

Processing data for a static snapshot representation of innovation ecosystem structure, i.e. for a static picture, introduces requirements that are completely different from those of creating a network representation that allows for interactive exploration–important in discovering the "intelligent fit"–and analysis of the innovation ecosystem. While both snapshot and interactive approaches are specific to downstream tools in the analysis process, this is particularly the case for interactive analytics. Importantly, from information visualization process model to visual analytics principles, we stressed the importance permitting the visualization users to interact with all the different phases of the process from data collection to cleaning to transformation to analysis and visualization.

One-mode or multi-mode networks

Across a wide spectrum of innovation ecosystem visual representations ranging from from a local innovation ecosystem engager Demola Tampere (Huhtamäki et al. 2013), to continent-wide, global-reaching EIT ICT Labs, we have designed one-mode and multi-mode networks. Some of the design processes were implemented within the research team per se, and in others, additional innovation ecosystem actors joined the design process. For reasons of data availability and application objectives, we have usually relied on multimode networks to represent innovation ecosystem structures. From a quantitative network analysis viewpoint, this is not optimal; one-mode directed networks would allow us to use the widest possible set of node-level metrics for identifying actors' structural roles. However, one-mode representation requires reducing the complexity of the innovation ecosystem; consequently, this leads into heightened need for investigators' awareness and understanding of the data processing done "under the hood"–more sophistication demanded for "intelligent fit".

In addition to the design issues described above, we have identified a set of overarching themes that should be considered as part of the design rationale of the data-processing architecture. These include decisions on the extent to which the organic complexity of the innovation ecosystem should be reduced, the data sourcing approach between in-house and external data, and the use of graph-based data management practices throughout the data-processing architecture.

Complexity versus clarity

Out of the four categories of Cynefin model–simple, complicated, complex and chaotic (Snowden and Boone 2007)–innovation ecosystems are either complex or chaotic. This complexity further highlights the importance to take into account the biases of human cognition discussed in Chapter 1. We claim that this nature of the phenomenon should be visible in its visual representation especially when a newly composed team investigates a previously unexplored innovation ecosystem. In other words, more specific representations are needed in subsequent steps – to drill down (Chapter 2.3) into the ways that individual mechanisms contribute to system-level behavior. Our experience shows that innovation ecosystem actors representing governmental organizations and early-phase startups often share very little in their worldviews. We believe that aligning divergent worldviews, by discussing the structure and mechanisms underlying an innovation ecosystem is imperative. We look forward to future investigations to explore the process of enacted sensemaking (Weick, Sutcliffe, and Obstfeld 2005) as detailed in Chapter 3.3.

From in-house to external data

When building up their operations, EIT ICT Labs invited us to join with them in developing a visual analytics artifact to support their work in exploring the existing interconnection between at the time six co-location centers. We started the exploration with internal data collected to represent their engineered activities. Together, we soon learned to realize that the insights afforded by this data were in most cases already known to the EIT ICT Labs orchestrators. Therefore, we chose to source data for analytics from another source – the IEN Dataset, as mentioned before.

5.3. Lessons for Future Development

Why not use graph representation from the beginning of the process, then? This is approach is indeed worth considering. A few observations are in order, however.

First, should graph representation of the source data be used, all members of the investigative team need to be able to access the data through an interactive exploratory user interface, perhaps operated with a Web browser. Second, it should be possible to export full sets of source data, both actor and transaction data, in tabular format for further investigations with spreadsheet processors and other self-service analytics tools. Third, it is important to realize that building a graph database with the data collected from various sources is effectively the process of building a network representation of the innovation ecosystem, an iterative development process for which the Ostinato Model can be used.

Another major improvement to add expressivity to the analytics process would be creating network-shaped projections of network data through queries and filters that are based on combinations of network topology and properties of actors and transactions. This capability would greatly benefit the creation of different views of full network representations of innovation ecosystems. For example, through network representation of the existing six co-location centers and related actors in the EIT ICT Lab ecosystem, we visually observed that Paris and Berlin are close to each other in terms of network topology. The green belt between the two cities suggests that a large number of investors operate in the two cities adding to their interconnectedness. When San Francisco Bay Area is added as the seventh EIT Digital node, however, the relative interconnectedness of Berlin and Paris became less significant than the interconnectedness of the Bay Area. Alternative explanations were possible. To explore whether this was due to the fact that the investors in the first visualization were largely also operating San Francisco Bay Area would require filtering the data. An expressive way to do this could be to use filters that are based on network topology and properties of actors and their interconnections.

We foresee two key streams of future work. The Ostinato Model provides a firm platform for further innovation ecosystem investigations conducted using a data-driven visual network analytics approach; we have a firm basis for future investigations. This allows continued experiments with the Action Design Research approach and Critical Realist mindset. We want to echo Freeman’s (2000) call for further work in developing infrastructure for visual network analysis–i.e. to develop systems of visual for system visualization (Chapter 1): “We can look forward to similar progress in developing database programs designed to facilitate the storage and retrieval of social network data. But the real breakthrough will occur when we develop a single program that can integrate these three kinds of tools into a single program.” While easy-access tools such as NodeXL and commercial tools including Palantir and Quid are available, a major need exists for component-based lightweight analysis processes that allow for data-driven visual network analytics investigations in the context of business and innovation ecosystems and beyond.

Second, at a more conceptual level, we encourage an analysis of the data-driven visual network analytics process to identify and describe in detail the three subsystems within the overall Information System artifact: i) technological artifact, ii) information artifact, and iii) social artifact (Lee, Thomas, and Baskerville 2015). We expect that this analysis will allow for further specificity in future rounds of Action Design Research, in which new artifacts for innovation ecosystem visual analytics are created through guided emergence. We are thrilled to observe the development steps that the editors, authors and readers of the book at hand will make in developing new systems of visuals to support collaborative visual analytics and visualization-supported enacted sensemaking in organizations.

References

Adner, Ron, and Rahul Kapoor. 2010. “Value Creation in Innovation Ecosystems: How the Structure of Technological Interdependence Affects Firm Performance in New Technology Generations.” *Strategic Management Journal* 31 (3) (March): 306–333. doi:10.1002/smj.821.

Archer, Margaret S. 1995. *Realist Social Theory: The Morphogenetic Approach*. Cambridge: Cambridge University Press.

Basole, Rahul C., Martha G. Russell, Jukka Huhtamäki, and Neil Rubens. 2012. “Understanding Mobile Ecosystem Dynamics: A Data-Driven Approach.” In *Proceedings of the 2012 International Conference on Mobile Business (ICMB 2012)*, 17–28.

Bastian, Mathieu, Sebastien Heymann, and Mathieu Jacomy. 2009. “Gephi: An Open Source Software for Exploring and Manipulating Networks.” In *Proceedings of the Third*

International AAAI Conference on Weblogs and Social Media, May 17-20, 2009, San Jose, California, USA.

Bendoly, Elliot. 2016. "Fit, Bias and Enacted Sensemaking in Data Visualization: Frameworks for Continuous in Operations and Supply Chain Management Analytics." *Journal of Business Logistics*.

Bygstad, Bendik, and Bjørn Erik Munkvold. 2011. "In Search of Mechanisms. Conducting a Critical Realist Data Analysis." In *Thirty Second International Conference on Information Systems, Shanghai 2011*.

Card, Stuart, Jock Mackinlay, and Ben Shneiderman. 1999. *Readings in Information Visualization: Using Vision to Think*. 1st ed. San Francisco, California: Morgan Kaufmann.

Dobson, Philip J. 2001. "The Philosophy of Critical Realism—An Opportunity for Information Systems Research." *Information Systems Frontiers* 3 (2) (June): 199–210. doi:10.1023/A:1011495424958.

Finkel, Jenny Rose, Trond Grenager, and Christopher Manning. 2005. "Incorporating Non-Local Information into Information Extraction Systems by Gibbs Sampling." In , 363–370. ACL '05. Stroudsburg, PA, USA: Association for Computational Linguistics. doi:10.3115/1219840.1219885.

Freeman, Linton C. 2000. "Visualizing Social Networks." *Journal of Social Structure* 1 (1): [np]. http://www.cmu.edu/joss/content/articles/volume1/Freeman.html.

Hansen, Derek L., Dana Rotman, Elizabeth Bonsignore, Nataa Milic-Frayling, Eduarda Mendes Rodrigues, Marc Smith, and Ben Shneiderman. 2012. "Do You Know the Way to SNA?: A Process Model for Analyzing and Visualizing Social Media Network Data." In *Proceedings of the 2012 International Conference on Social Informatics*, 304–313. IEEE. doi:10.1109/SocialInformatics.2012.26.

Heer, Jeffrey, and Ben Shneiderman. 2012. "Interactive Dynamics for Visual Analysis." *Communications of the ACM* 55 (4): 45–54.

Huhtamäki, Jukka, Ville Luotonen, Ville Kairamo, Kaisa Still, and Martha G. Russell. 2013. "Process for Measuring and Visualizing an Open Innovation Platform: Case Demola." In *17th International Academic MindTrek Conference 2013: "Making Sense of Converging Media", October 1-3, Tampere, Finland*. ACM. http://urn.fi/URN:NBN:fi:tty-201312201533.

Huhtamäki, Jukka, Martha G. Russell, Neil Rubens, and Kaisa Still. 2015. "Ostinato: The Exploration-Automation Cycle of User-Centric, Process-Automated Data-Driven Visual Network Analytics." In *Transparency in Social Media: Tools, Methods and Algorithms for Mediating Online Interactions*, edited by Sorin Adam Matei, Martha G. Russell, and Elisa Bertino, 197–222. Springer International Publishing Switzerland. http://link.springer.com/chapter/10.1007/978-3-319-18552-1_11.

Huhtamäki, Jukka, Martha G. Russell, Kaisa Still, and Neil Rubens. 2011. "A Network-Centric Snapshot of Value Co-Creation in Finnish Innovation Financing." *Open Source Business Resource* (March): 13–21. http://www.osbr.ca/ojs/index.php/osbr/article/view/1288/1234.

Hwang, Victor W., and Greg Horowitt. 2012. *The Rainforest: The Secret to Building the Next Silicon Valley*. 1.02 ed. Los Altos Hills, California: Regenwald.

Iansiti, Marco, and Roy Levien. 2004. *The Keystone Advantage: What the New Dynamics of Business Ecosystems Mean for Strategy, Innovation, and Sustainability*. Boston, Mass: Harvard Business Review Press.

Järvi, Kati, and Samuli Kortelainen. 2016. "Taking Stock of Empirical Research on Business Ecosystems: A Literature Review." *International Journal of Business and Systems Research*: Forthcoming.

Keim, Daniel, Jörn Kohlhammer, Geoffrey Ellis, and Florian Mansmann. 2010. *Mastering the Information Age - Solving Problems with Visual Analytics*. Eurographics Association. http://www.vismaster.eu/book/.

Lee, Allen S., Manoj Thomas, and Richard L. Baskerville. 2015. "Going back to Basics in Design Science: From the Information Technology Artifact to the Information Systems Artifact." *Information Systems Journal* 25 (1) (January): 5–21. doi:10.1111/isj.12054.

Moore, James F. 1993. "Predators and Prey: A New Ecology of Competition." *Harvard Business Review* 71 (3): 75–86.

Navarro, Gonzalo. 2001. "A Guided Tour to Approximate String Matching." *ACM Comput. Surv.* 33 (1) (March): 31–88. doi:10.1145/375360.375365.

Nykänen, Ossi, Jaakko Salonen, Matti Haapaniemi, and Jukka Huhtamäki. 2008. "A Visualisation System for a Peer-to-Peer Information Space." In , 76–85. Tampere, Finland: Tampere University of Technology.

Pentland, Alex "Sandy." 2015. *Social Physics: How Social Networks Can Make Us Smarter*. New York, New York, USA: Penguin Books.

Ritala, Paavo, and Pia Hurmelinna-Laukkanen. 2009. "What's in It for Me? Creating and Appropriating Value in Innovation-Related Coopetition." *Technovation* 29 (12) (December): 819–828. doi:10.1016/j.technovation.2009.07.002.

Rubens, Neil, Kaisa Still, Jukka Huhtamäki, and Martha G. Russell. 2010. "Leveraging Social Media for Analysis of Innovation Players and Their Moves." Innovation Ecosystems Consortium.

Russell, Martha G., Jukka Huhtamaki, Kaisa Still, Neil Rubens, and Rahul C. Basole. 2015. "Relational Capital for Shared Vision in Innovation Ecosystems." *Triple Helix: A Journal of University-Industry-Government Innovation and Entrepreneurship (THJI)*.

Russell, Martha G., Kaisa Still, Jukka Huhtamäki, Camilla Yu, and Neil Rubens. 2011. "Transforming Innovation Ecosystems through Shared Vision and Network Orchestration." In . Stanford, California, USA.

Sein, Maung, Ola Henfridsson, Sandeep Purao, Matti Rossi, and Rikard Lindgren. 2011. "Action Design Research." *MIS Quarterly* 35 (1): 37–56.

Snowden, David J., and Mary E. Boone. 2007. "A Leader's Framework for Decision Making." November.

Still, Kaisa, Jukka Huhtamäki, Martha G. Russell, Rahul C. Basole, Jaakko Salonen, and Neil Rubens. 2013. "Networks of Innovation Relationships: Multiscopic Views on Finland." In *Proceedings of the XXIV ISPIM Conference – Innovating in Global Markets: Challenges for Sustainable Growth, 16-19 June 2013, Helsinki, Finland*, 15.

Still, Kaisa, Jukka Huhtamäki, Martha G. Russell, and Neil Rubens. 2012. "Paradigm Shift in Innovation Indicators: From Analog to Digital." In *Proceedings of the 5th ISPIM Innovation Forum, 9-12 December, 2012, Seoul, Korea*. Seoul, Korea.

———. 2014. "Insights for Orchestrating Innovation Ecosystems: The Case of EIT ICT Labs and Data-Driven Network Visualisations." *International Journal of Technology Management* 66 (2/3) (January): 243–265. doi:10.1504/IJTM.2014.064606. http://www.inderscienceonline.com/doi/abs/10.1504/IJTM.2014.064606.

Thomas, Llewellyn D. W., and Erkko Autio. 2012. "Modeling the Ecosystem: A Meta-Synthesis of Ecosystem and Related Literatures." In *DRUID 2012, June 19-21, CBS, Copenhagen, Denmark.*, 28.

Venkat Ramaswamy. 2009. "Leading the Transformation to Co-creation of Value." *Strategy & Leadership* 37 (2) (March): 32–37. doi:10.1108/10878570910941208. http://www.emeraldinsight.com/doi/abs/10.1108/10878570910941208.

Weick, Karl E., Kathleen M. Sutcliffe, and David Obstfeld. 2005. "Organizing and the Process of Sensemaking." *Organization Science* 16 (4) (August): 409–421.

doi:10.1287/orsc.1050.0133.
Wong, Pak Chung, and J. Thomas. 2004. "Visual Analytics." *IEEE Computer Graphics and Applications* 24 (5) (October): 20 – 21. doi:10.1109/MCG.2004.39. http://ieeexplore.ieee.org/ielx5/38/29442/01333623.pdf?tp=&arnumber=1333623&isnumber=29442.