

Theory and Development of Tool Frameworks for Interactive Knowledge and Data Visualization in Computational Social Science

Tom McDermott
Systems Engineering Research Center
Stevens Institute of Technology
Atlanta, Georgia
Email: tmcdermo@stevens.edu

Molly Nadolski
Georgia Tech Research Institute
Georgia Institute of Technology
Atlanta, Georgia
Email: molly.nadolski@gtri.gatech.edu

Abstract— We present a framework and sample case study linking knowledge visualization forms with visualization of human and social data analytics. This involves integration of qualitative methods to identify and connect conceptual models with computational social science approaches for data analytics. The framework addresses two challenges: explicitly linking conceptual knowledge visualization to data analytic tools and using that linkage to explore complexity in social situations. In social or organizational change management strategies, combining new data and knowledge is critical for decision making. When the situation is complex, data must be placed in a knowledge framework to build team learning and to create new mental models for strategic change. As situational complexity increases, the role of knowledge transfer in team social networks becomes more critical, and the ability to visualize knowledge (as opposed to information) becomes paramount to insight and effective decision making. We demonstrate a framework that is derived from theories in the systems thinking and complexity thinking domains, which is then linked to how leaders and managers visualize and communicate data, information, and knowledge. A case study based on Russia’s multi-domain influence in the country of Moldova is used to demonstrate explicit linkage between knowledge visualization forms and social data analytics.

Keywords-knowledge management; complex adaptive systems; data visualization; conceptual modeling; leadership.

I. INTRODUCTION

This research explores ways to explicitly link knowledge-driven conceptual models with data-driven techniques that create insight for decision making in complex systems and situations. Visualization of knowledge in a conceptual form is primarily a qualitative process completed in group facilitation activities. It often uses data and information visualization but the linking of conceptual relationships to the data analytic tools seldom leaves the “whiteboard.” With the advent of semantic data analysis and learning tools, there is more opportunity to formally link conceptual learning with data search strategies. This paper presents theory linking data, information, and knowledge visualization forms, then presents a case study exploring tools for combining visualization to demonstrate the possibilities.

The central thesis of this work is that as complexity increases, the role of knowledge transfer in social networks becomes more critical, and the ability to visualize knowledge (as opposed to information) becomes paramount to decision making and strategy. Complex situations are characterized by periods of human social learning followed by periods of stable execution. Visualization tools allow the mapping of large amounts of data to visual patterns that aid human information processing. Exploring data and designing visualization

approaches requires a modeling framework as shown in Figure 1. We follow a systems thinking framework that semantically links a qualitative description of the system architecture and measurable constructs to a specification of quantitative methods for computational modeling of those measures.

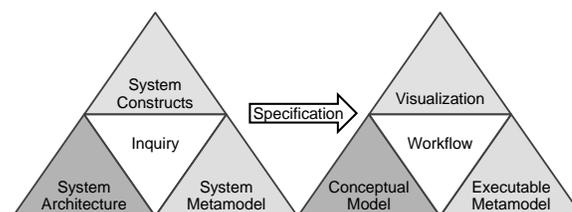


Figure 1. The bridge between qualitative analysis and social analytic model specification.

We define the qualitative aspects in Figure 1 as “System Metamodeling” using three fundamental abstraction approaches: system metamodels, system constructs, and system architecture models. These are determined in a participative and inquiry-based process. We describe quantitative aspects as “Executable Metamodeling” determined by a specification and design workflow using conceptual models, computation, and data visualization. It is useful to think about this as a tool framework. The tools support structuring the systems metamodel, creating the conceptual models, creating the executable metamodels, analyzing and visualizing the decision space, and managing the contained knowledge over time [2].

The system metamodel is described as the set of constructs and rules used to define semantic relationships across information sets, associated data sets, and methodologies or processes. The metamodel definition on the semantic side is an architectural description of the system using modeling views and stakeholder viewpoints. The executable metamodel is the dataset design and any associated computational models. This is further discussed in section III.

A research effort was conducted using the framework to explicitly deconstruct qualitative methods into a “system metamodel” in order to identify and specify “executable metamodels” using computational social science approaches.

The research domain was the study of Russian foreign and security policies. Here, the conventional focus on single-factor explanations has been challenged by the emerging cross-domain character of Russia’s statecraft and pursuit of so-called “new generation warfare” in gray zone conflicts. The domain is currently dominated by the application of a narrow set of research methods, such as comparative case studies, regression analysis, expert surveys, and interviews that confront systematic problems related to limited, out of sample, and disconnected data. The case study uses the metamodeling

framework to engage more holistic methods that encourage flow of knowledge across multiple domain experts and methods [3] [4].

There are a growing number of actor and event coding schemes and software tools that offer opportunity to automatically extract stakeholders, sentiments, and events related to foreign and security policy analyses. They parse a growing body of English and Russian language (and other languages) databases and media publications on daily and even real-time bases. They target actors and types of events, but struggle with visualization of complex and multi-domain relationships.

We used a combination of scenario analysis and systems thinking methods, such as narratives, taxonomies, and accompanying visual diagrams, to create the systems meta-model. The strategies and tactics within Russia gray zone operational and definitional domain and the key pressure points are identified by scenario development, which captures fundamental change in the present state. The result was a set of conceptual models, and we identified methods to bridge them to computational models via data mining to realize the executable meta-model. This is presented in section IV.

The use of the framework is recommended in periods of program complexity. Challenges of program complexity may have any or all of the following three dimensions: (1) the scale of the project and supporting enterprise, with the variety of organizational disciplines, processes, and tools that might be used to execute the program; (2) the uncertainties created by newness, originality, and innovation; and (3) the external context the program is surrounded by, including social and political factors and market dynamics [5].

The data visualization challenge is to support the combination of qualitative or heuristic decisions that must be made in conjunction with quantitative data driven decisions. These can be categorized as knowledge-driven versus data-driven decisions. How the network of decision-makers collaboratively use data, information, and shared knowledge is paramount.

Progress on visually linking qualitative knowledge with quantitative data is still lacking but can be enabled by machine learning tools that find and semantically link knowledge and data. In complex and uncertain situations, the flexibility of most data visualization tools to create the “story” that is needed to move forward falls short, and decision makers must explore data in more qualitative frameworks. The challenge is to situationally master the combination of tools and visualization forms that visualize data and information to collect, transfer and communicate shared knowledge. The goal of this research is to build linkages between qualitative and quantitative visualization methods to aid decision making. This is discussed further in section II.

II. CONTEXT: COMPLEXITY IN SOCIAL SYSTEMS

Although all human and social systems can be described generally as complex adaptive systems, they undergo periods of stability and disruption driven by internal and external drivers of change. It is important to understand the drivers of this situational complexity and choose appropriate analytical constructs to assess and implement change. Geraldi and Albrecht [5] provided a useful categorization of complexity

from the complex project management domain ascribing complexity across three dimensions: Complexity of Fact, Complexity of Faith, and Complexity of Interaction.

A. Dimensions of Complexity

Complexity of Fact is a measure of the number of entities and their interdependence as an issue of interdependent information. Given large complexity of fact, it is difficult to obtain and use information rapidly enough to support decision making. Data modelers search for available higher-level abstractions or simplifications to base their data for decisions. In this dimension, there exists a fundamental problem of abstraction – one needs to visualize the whole of the project and represent information in patterns that are embedded in the whole. Visualization frameworks that maintain the holistic perspective while allowing access to detailed information are necessary.

Complexity of Faith relates to program situational uncertainty, often associated with the newness of the problem being solved or methods used to solve it. Complexity of Faith implies a need for learning. In periods of high uncertainty, decisions rely more on shared knowledge than on availability of data and information. Decision makers and their visualization methods and tools in these periods must encourage facilitation, knowledge transfer, and more abstract conceptualization of decision alternatives.

Complexity of Interactions exist at interfaces between different systems and domains. These include people, disciplines, locations, external stakeholders, and social and political factors. In complex situations, it is important to frame program interactions in a larger enterprise architecture framework. Understanding interactions drives the need for facilitation and conceptualization methods and tools.

To manage complexity, one must develop strategies and tools that support 1) exchange and visualization of information across social networks, 2) exchange and visualization of knowledge between decision agents, and 3) evolutionary planning that includes cycles of learning. The presence of all three of these strategies imply that project learning be data-driven, so visualization methods that support access to both information and knowledge must be used. These methods are discussed on literature related to systems thinking and the complexity sciences.

B. Modeling Complexity using Enterprise Architecture

Sociotechnical systems analysis is a specific methodology that supports modeling of multiple factors across all layers of a complex situation, enterprise or societal construct using sets of tools derived from system science and system modeling. The methods recognize that factors arise from the interaction of many and diverse enterprises that can be defined by their entities, relationships, established processes, pursued strategies, and emergent phenomena. The sociotechnical systems analysis attempts to capture the combined conceptual, data, and analytical modeling artifacts necessary to completely describe the problem [6] [7].

With respect to social situations, the method produces a set of artifacts that describe the system context and boundaries, system entities and relationships, primary construct variables, potential causal variables, and phenomena of interest. This is the system meta-model. The process is conducted such that

insight can be fed into dynamic computer models. The sociotechnical systems analysis produces artifacts that communicate the abstractions and aggregation of behaviors across different scales, helping to explicitly document both the assumed and modeled variables. At the core are entities and their relationships, which can be organized into associated databases and warehouses. The entity-relationship model can be created, modified, and refined over periods of short- and long-term study. Standardized coding of the data entities then makes relevant data elements accessible to researchers and analysts.

The conceptual model representations produced by the sociotechnical systems analysis serve as a bridge between the qualitative aspects of the problem (system meta-model) and the quantitative analysis approach (executable meta-model). This is the purpose of the framework shown previously in Figure 1. In the long-term, we expect that advances in machine learning and semantic graphs can bring the system meta-model (data constructs and conceptual models) and executable meta-model (collected data and algorithms) into the same visualization toolsets. The bridge between the two meta-models is a conceptual model that uses semantic relationships to specify the analytical models. This process of deconstructing complexity is tightly linked to mindsets and tools from the domain of systems thinking.

C. Modeling Complexity of Gray Zone Warfare

Russia’s use of “Gray-Zone” warfare is an example of intentional complexity as used to hide or ambiguate purposeful actions leading to objectives that would prepare or position a state for possible armed conflict. Figure 2 depicts the set of factors used in Gray Zone conflict.

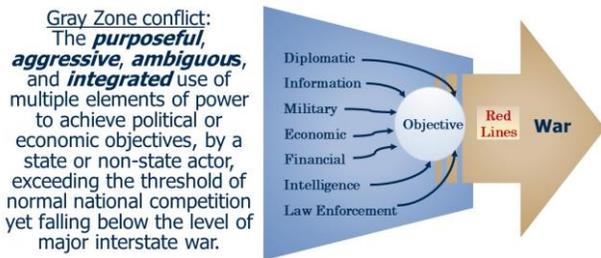


Figure 2. Multiple Objectives used in Gray Zone conflict.

All three types of complexity are exhibited in Gray Zone conflict. Complexity of Fact is exhibited in the almost infinite number of actions and related events that a nation state might pursue. For a data analyst, finding the right abstractions to represent as evidence of such conflict is a continual challenge, as well as ensuring those abstractions are relevant to the current time or predictive of continuing events. The relationship between different objectives - such as diplomatic exchanges, troop movements, and economic sanctions - must be considered holistically. Care must be taken against presenting individual constructs, such as diplomatic exchanges, as a single indicator of state objectives. A conceptual model that shows the relationship between all objectives is important but often ignored by data analysts and even policy makers.

Likewise, Complexity of Faith and Complexity of Interaction are reflected in the pursuit of multiple

simultaneous objectives in dynamic combinations. In Gray Zone conflict this is intended to create uncertainty in the adversary and the relationship between an action and the adversary’s response – dynamic interaction – is the intent of the strategy. Again, a conceptual model in important before data analysis is relied upon. In section IV a conceptual modeling tool called a “Systemigram” is used to visualize knowledge about Russia’s strategies in the country of Moldova.

III. SYSTEMS AND COMPLEXITY THINKING IN DATA ANALYTICS DESIGN

McDermott and Freeman describe three systems thinking characteristics and competencies that are essential to managing complexity. These are *sensemaking*, *adaptive and computational thinking*, and a *design mindset*. Each of these competencies use data-driven activities that encourage visualizing the relationships between project uncertainties and the underlying project activities [1].

Sensemaking is a collaborative process that involves collection of knowledge, visually describing or modeling a problem or solution in the wider context and learning by doing. *Adaptive and computational thinking* is the ability to situationally adjust a team’s thinking and related activities by employing analytics and simulation methods that make sense of large amounts of data (or to understand when data is lacking). Evolution of strategy and planning is a process of iterative design. Building a *design mindset* moves the entrained thinking of the team away from continued use of available data, methods and tools to a participatory team process of understanding and selecting new data, methods and tools. In complex situations, one should strategically design and redesign data analytic measures in response to new insights and understanding of the situation. Computational thinking then relates existing data constructs (generally performance measures) to new constructs selected to guide future evaluation and prediction. The challenge in data visualization is to begin automating these relationships.

Systems thinking encourages modeling or mapping the program and external context together in order to visualize the internal and external interactions that drive project execution. Conceptual models are developed by experts who have the capacity to develop objectively multiple views of a system and its context based on background research and discussions at expert meetings. Conceptual modeling as a visualization strategy is thus a core aspect of any data analytic activity that involves complexity and uncertainty.

The foundation of sensemaking is visualization. Visual modeling is used to frame the program and situation within the enterprise system that is addressing it.

A. System & Complexity Thinking in Data Analytics Design

Boulton, Allen, and Bowman describe the social networks in a design strategy as a “learning multi-agent model” – networks of individual agents who act according to their experience and their beliefs, but ideally aligned around common goals. Inducing change in complex systems requires self-organization around new shared beliefs [8]. A core concern of human social analytics is discovering, measuring, and informing those beliefs. However, creation of new shared

beliefs is a knowledge transfer activity across individual agents.

Snowden [9] in his work on complexity and the “Cynefin” framework recognized knowledge exchange in the form of a “learning multi-agent model.” Snowden’s work suggests that a data artifact may relate to a team’s knowledge, but it is the continual flow of new knowledge and artifacts that must be managed in complex situations. Knowledge exchange requires both content and a context. The content can only be exchanged if the context can be shared (language, education, experience, culture, etc.). Visualization tools that address both changing content and changing context are essential.

Data visualization tools and strategies are part of that knowledge exchange but can work against it – they represent the situation only in the context of the tool and existing data, and force information exchange over knowledge transfer. Visualizing a future path of execution, in complexity and uncertainty, requires representation of information in a metamodel framework that tells a story.

This gives us frameworks to evaluate two types of visualization, those that focus on data and information transfer and associated execution, and those that focus knowledge transfer and associated strategy development. Both have application in complex situations. Having a basis for and learning when to apply each type is the key to design of successful human/social analytics. In section IV we demonstrate how these can come together in a data analytic analysis.

Knowledge visualization will often use text-based content in a form that emphasizes relationships or patterns. While information visualization is typically used to explore large amounts of abstract data, knowledge visualization is more used to aid in communication of abstract knowledge. The visual models provide the conceptual language for shared context that is required for knowledge flow. Knowledge visualization tools tend to support the sensemaking process, helping the observer to fill in additional insights based on patterns in the underlying information and data.

The purpose is to properly conceptualize, structure, relate, and validate the relationships in the complex situations from factual data and information to the higher levels of abstraction or aggregation needed to relate meaning and knowledge. This is key to the abstraction of “Fact” to “Faith.” Narrative conceptual modeling forms are used to express emergence, relating to evolving situations in the domain or enterprise. In our research we have settled on a conceptual modeling tool called a “systemigram” which blends narrative and diagram. The next section provides an example.

IV. CASE STUDY: INTERACTIVE CONCEPTUAL AND DATA ANALYTIC MODELING AND VISUALIZATION

In the case study, we used a combination of scenario analysis and systems thinking methods, such as creating narratives, taxonomies, and accompanying visual diagrams, to build data analytic models that searched for evidence of Russian Gray Zone conflict in the country of Moldova. The Russian objectives, strategies and tactics are identified by scenario development. This leads to a set of conceptual

models, and we identified methods to bridge them to computational models via data mining.

In order to visualize the problem space of gray zone warfare in Moldova, the conceptual model of the system was produced using the systemigram tool that describes the many actors, interactions, processes, resources, and feedback loops present in the Russia-Moldova interaction. Systemigrams are a qualitative model of system behavior that is useful for human reasoning about the dynamics of complex systems [10]. A systemigram serves to illustrate the relationships and flows that exist within a system in its specific context. Figure 3 shows the full systemigram diagram.

The term systemigram is derived from the phrase and portmanteau “systemic diagram.” Consisting of both narrative and diagram, systemigrams are a type of conceptual modeling tool closely related to concept maps – or “maps of learning.” These maps explore complex situations through the perspective of those embedded in the system or those who are affected by a specific challenge occurring within the system [10].

The systemigram models are developed by experts who have the capacity to develop objectively multiple views of a system based on background research and discussions at expert meetings. To frame such a model, a context for analysis was provided, then tested against the context of Russia’s perspective of gray zone warfare in Moldova. Central questions of interest are developed to derive system boundaries and ontology structure to support informing the computational model development.

However, they are not quantitative models that are able to make predictions. In order to enable the tools of inferential statistics and machine learning to bear on these systemigrams, they must be transformed into quantitative models. This required further addressing and defining ontological and semantic challenges to effectively model the semantics, and identifying the transformations that exist and occur between qualitative and quantitative approaches. In this phase, the team used exploration of the Global Database of Events Language and Tone (GDELT) events and established coding, as defined by the broader research on Russia-EU aggression, to find evidence that either confirm or deny Western perceptions of gray-zone operations and tactics, established by real event chains. This creates a model that links a series of events to sets of tactics, which together, form a perceived strategy.

GDELT includes events reported in the global media coded with many pieces of extracted data. The primary information used in this project is the Date/Time, Location, Actors, and Cameo Event Code [11]. The Actors are the entities (people, government agencies, corporations) that engage in behaviors and the CAMEO codes are a systematic representation of the behaviors on a scale from 1, Make Public Statements, to 20, Unconventional Mass Violence. These codes are extracted from the news articles to enable researchers to study the dynamics of the global system of large-scale human behavior. These machine learning techniques will be specifically adapted to take as input the occurrences and frequencies of coded diplomatic events as related to gray zone conflict.

TABLE 1. EVOLUTION OF SELECTED MOLDOVA SYSTEM CONCEPTS OVER THE TIME BETWEEN 12/2016-5/2018

Month	0	1	2	3	5	6	7	8	9	13
201612	25	6	44	13	2	50	43	2515	5437	54
201701	49	58	37	58	6	231	115	7856	9861	182
201702	24	57	25	18	2	317	161	6909	10616	122
201703	44	49	35	37	1	244	173	6558	11034	176
201704	19	20	10	31	4	177	112	5318	10403	172
201705	28	29	11	24		170	82	5900	7729	122
201707	44	72	39	80		171	202	6452	11060	676
201708	44	63	23	53		113	97	3303	6540	346
201709	51	60	15	26		221	120	3504	6625	142
201710	18	34	6	11	1	173	82	2577	4048	34
201711	13	34	6	26	3	165	123	3620	7635	76
201712	25	12	22	7	2	199	106	4956	8776	8
201801	12	16	10	10	4	183	119	3888	5548	64
201802	15	15	20	16		146	152	4850	6646	42
201803	10	4	17	11		296	186	6193	10409	50
201804	10	8	19	8	1	193	120	4513	8186	80

Concept 5 and concept 13 displayed the most obvious correlation of concept to events. Both of these concepts are shown in the full systemigram in Figure 3 as node (Russian Peacekeepers) link (deployed to alleviate) node (Frozen Conflict). These are periods of Russian troop withdrawals from Moldova. In 07/2017 and 12/2017 Russian troops were withdrawn from Syria which was picked up as a local maximum (676) and minimum (9) of the GDELT queries associated with concept 13. Concept 3, node (Russian Foreign Policy) link (Strategy attempts to) node (Create stability and pursue multivectorism) was also found to correlate with signals from the other two concepts and thus provide strategic analysts and planners with confirmatory signals of Russian objectives. We also see that some events with large coverage and a persistent presence such as node (Western States: United States) link (Stokes conflict to disrupt foreign policy agenda) node (Russian Foreign Policy Challenges and Threats), and this node to link (Informs strategy and policy) to node (Russia) follow a long term cyclical trend rising and falling together without significant spikes or crashes in concepts 8 and 9. These events form the background environment which is important to shaping the understanding of politics and decision-makers, and quantified systemigrams can help us understand these entities for foresight.

V. CONCLUSIONS

This paper presented a framework to bridge qualitative and quantitative models and associated visualization of data with knowledge. Few efforts have been made to translate emerging research in these approaches into user-friendly and interactive visual tools that can aid in both hypotheses' generation, strategic forecasting, and scenario assessment. To create a holistic approach and toolset, this research draws on theories and frameworks from multiple disciplines including systems thinking, complexity thinking, and complex project management. Qualitative methods promote the development of conceptual models to aid in structuring and understanding

real world systems and problems, in this context, applied to Russia's application and perspective of gray zone conflict in Moldova. However, qualitative models are difficult to directly measure or validate. Thus, building off the conceptual models developed using the systemigram tool and identifying methods to bridge them to computational models via data mining provided insight into the system to inform further research and development of more quantitative models. The techniques presented here can be adapted to near-real-time detection of hot-spots and anomalies or used as a basis for post-hoc quantitative analysis.

The conversion of qualitative to quantitative models requires the construction of detailed mappings into coding spaces. These mappings are useful because they enable database systems to calculate measurements for each concept in the soft system model. This approach allows segmentation of the data by arbitrary filters on the actors involved and time spans covered, thus allowing researchers to compare these effects on various subsets of the corpus and draw conclusions about how these effects differ across regions, times, and combinatorial pairings of country- level characteristics for more foresightful strategy development.

REFERENCES

- [1] T. McDermott and D. Freeman, "Systems Thinking in the Systems Engineering Process, New Methods and Tools," in M. Frank, S. Kordova, H. Shaked (Eds.), *Systems Thinking: Foundation, Uses and Challenges*, Nova 2016.
- [2] J. Ernst, "What is metamodeling, and what is it good for". Retrieved from <http://infogrid.org/trac/wiki/Reference/WhatIsMetaModeling>. November 2015.
- [3] M. Nadolski and J. Fairbanks, "Complex Systems Analysis of Hybrid Warfare," 17th Annual Conference of Systems Engineering Research (CSER), Washington, DC, 2019.
- [4] Nadolski, Molly. *A Sociotechnical Modeling Framework for Complex Systems Analysis of Gray Zone Warfare. Conflicts in the Gray Zone: A Challenge to Adapt*. Budapest, Hungary, 2017.
- [5] J. G. Galdi and G. Adlbrecht, "On faith, fact and interaction in projects," *Project Management Journal*, 38 (1), pp. 32-43, 2007.
- [6] W. B. Rouse & D. Bodner, Multi-level modeling of complex socio-technical systems – phase 1, A013 - final technical report, SERC-2013-TR-020-2, Systems Engineering Research Center.
- [7] W. B. Rouse & M. Pennock, Multi-level modeling of socio-technical systems a013 - final technical report, SERC-2013-TR-020-3, Systems Engineering Research Center.
- [8] Boulton, J., Allen, P., & Bowman, C. 2015. *Embracing complexity: strategic perspectives for an age of turbulence*. Oxford University Press.
- [9] Snowden, D. 2002. Complex acts of knowing: paradox and descriptive self-awareness. *Journal of Knowledge Management*, 6 (2), 100-111. May 2002.
- [10] Sauser, B., & Boardman, J. (2013). *Systemic thinking: Building maps for worlds of systems*. Chichester: Wiley-Blackwell.
- [11] Conflict and Mediation Event Observations Event and Actor Codebook [Scholarly project]. (2012, March). GDELT Project. Retrieved from <http://data.gdelproject.org/documentation/CAMEO.Manual.1.1b3.pdf>