EuroRV3: EuroVis Workshop on Reproducibility, Verification, and Validation in Visualization (2018)
K. Lawonn, N. Smit, L. Linsen, and R. Kosara (Editors)

Visual Analytics-enabled Bayesian Network Approach to Reasoning about Public Camera Data

E. Chuprikova[1], A. M. MacEachren[2], J. Cron [1] and L. Meng[1]

[1]Chair of Cartography, Department of Civil, Geo and Environmental Engineering, Technical University of Munich, Germany
[2] GeoVISTA Center, Department of Geography, The Pennsylvania State University, University Park, PA, USA

Abstract

The Visual Analytics (VA) approach has become an important tool for gaining insights on various data sets. Thus, significant research has been conducted to integrate statistical methods in the interactive environment of VA where data visualization provides support to analysts in understanding and exploring the data. However, much of the data explored with VA is inherently uncertain due to limits of our knowledge about a phenomenon, randomness and indeterminism, and vagueness. The Bayesian Network (BN) is a graphical model that provides techniques for reasoning under conditions of uncertainty in a consistent and mathematically rigorous manner. While several software tools for visualizing and editing BNs exist, they have an evident shortcoming when spatial data. In this study, we propose a Visual Analytics-enabled BN approach for reasoning under uncertainty. We describe the implementation procedure using an example of heterogeneous data that includes locations of security surveillance cameras installed in public places.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Line and curve generation

1. Introduction

Uncertainty is a natural outcome of scientific research and it is an inevitable characteristic of data handling processes; therefore, visualization of data uncertainty has been the focus of a substantial body of research. However, the research outcomes have been limited by a lack of a complementary focus on uncertainty in reasoning processes [Mac15]. The goal of analytical reasoning is to gain insight from data, and Visual Analytics (VA) facilitates this process and enables human reasoning about complex problems through visual interfaces and computational methods that can process large, messy, and heterogeneous data. The reasoning process with such data sets must typically cope with uncertain conditions. In the current research we adopt an uncertainty definition from Bayesian statistics, where it is associated with such factors as ignorance (due to limits of knowledge about a phenomena); randomness and indeterminism (as not all events are determined by causal relationships and there is always room for physical randomness); and vagueness (as statements we make are often vague) [KN10].

Over the last decades there has been a growing interest in Bayesian inference due to increased recognition that it is an effective method for supporting decision-making practices. Various methods for representing and reasoning with uncertainty have been developed, including the fuzzy set theory, formal logics [Blo06], probabilistic clustering [LLZ*17], and Bayesian Networks (BNs), also called belief networks and causal probabilistic networks [SPK98]. Several studies have indicated that BNs provide an effective approach to assess data uncertainty; therefore, the use and visualization of BNs has received explicit attention. The BNs graphically represent uncertain quantities and decisions that reveal the probabilistic dependencies among the variables and the related information flows [VC16]. The most relevant advantage of BNs is that they provide an effective mechanism to deal with uncertainties and information from different sources, such as expert judgment and observable patterns, and are able to consider conditional dependencies in data. The efficiency of Bayesian methods lies in providing uncertainty estimates about the given alternatives, and the fact that human reasoning can be integrated makes it a promising method for the implementation within VA.

In the current contribution, we present a VA approach for spatial, heterogeneous data analysis that uses integrated BNs to support reasoning about spatial patterns in the context of uncertain conditions. More specifically, we address the issue of decision-making under uncertainty when performing a classification task. When decision makers deal with classification problems, they often quantify the likelihood of a given event on the basis of their personal knowledge. This study proposes a technique that facilitates interaction between the analyst and the BN's probabilistic model through VA interface supporting expert input. The classification task is conducted using locations of surveillance cameras in public places in the city of Moscow. The provision of multiple kinds of surveillance

DOI: 10.2312/eurorv3.20181141

cameras may increase the chance that if anything happens in a location it will be possible to: (a) know that it happened, and (b) gain some information about what happened and perhaps who was responsible. The application would also allow an analyst to explore the likelihood of an undesirable/unsafe situation happening without it being noticed (or without it being noticed soon enough to do anything about it). Interactive visualization enables an analyst to participate in the data exploration and reason about the information provided by the data considering the state of uncertainty that is inevitably incorporated during the data exploration.

In the next sections, we review relevant literature and introduce the VA approach realized by the integration of BNs into a VA interface for the analysis of video camera data from Moscow. The paper presents results of the integration of Bayesian statistical methods for reasoning about the potential to monitor behavior in places effectively through integration of camera data.

2. Related work

Recent reviews of literature focused on visualizing uncertainty of spatial data [MRH*05, SRK16] have shown a variety of approaches, including glyphs, isolines and isosurface, grid structures, 3D, and choropleth maps. However, the uncertainty visualization perspective (as evident in work to date) deals primarily with visually signifying ambiguous data rather than with reasoning under uncertainty [Mac15]. At the same time, well informed decisions with a spatial context depend essentially upon adequate data, knowledge of uncertainty, and the ability to reason about it.

As several studies have indicated, BNs provide an effective approach to access data uncertainty; therefore, the visualization of BNs has received substantial attention. [ZRNG99] reported on the utility of temporal order, color, size, proximity (closeness), and animation techniques for mapping cause-effect relationships in a BN model. [CSLG05] proposed using heat maps for visualizing the conditional probability tables. [CMS11] proposed a method to support expert analysis of BN models using a "thought bubble line" to connect nodes in a graph representation and their internal information at the side of the graph. [CE17] introduced two visualization techniques: inference diffs, for comparisons of effects of evidence using concentric pie and ring charts, and relevance filtering to guide the user to variables of interest in the model.

Despite the fact that the BNs are a commonly applied modeling technique in different applications, the visual exploration is limited to cause-effect relationships among the variables, and only scant attention has been paid to the development of visualization support for the spatial component of the data. Although people are typically poor at numerical reasoning about probability, human thought is sensitive to subtle patterns of qualitative Bayesian, probabilistic reasoning [OC09]. Therefore, the development of visual interfaces that allow users to iteratively deal with BN analysis in a spatial context might enhance the usability of this method and open new perspectives for application of Bayesian reasoning in GIScience and cartography.

Previous research has demonstrated the potential of developing uncertainty-aware VA. In one recent study, integration of uncertainty (confidence) visualization with computational methods in a VA application demonstrated how confidence in an estimate on multiple interacting factors (based on weather predictions) can be simultaneously visualized [KTB*18]. One of the first examples of a framework that merges Bayesian statistics and VA is proposed by [HLH15]. Their solution focuses on the human-computer interaction that helps experts synthesize information in the data, interact with the data, and guide automated, analytical procedures. Another application is described in [SDC14], where the authors presented an analytical framework called Abuse User Analytics (AuA) aiming to provide information about the behavior of online social network users. The AuA processes data users' discussions, and renders information about users' abusive activities. The analysis and visualization implemented within AuA utilize BNs to model the users' choices and monitor changes in their behavior in text-based communities. Hence, the VA approach might facilitate abstraction of numerical details from Bayesian statistics and represent the modeling through qualitative characteristics that promote human reasoning, and yet preserve the semantics underlying Bayesian inference, and BNs in particular. The implementation of uncertainty-aware VA using BNs that considers spatial data contexts, seems to offer potentially useful tools for spatial data analysis.

3. Materials and Methods

3.1. Data

Recent open data initiatives have transformed the availability of and ease of access to high-quality spatial data and have opened up new perspectives for analysis and data visualization. The current research leverages data acquired from the open portal of the municipality of Moscow. The primary input information includes data from video surveillance cameras in the city installed in: (a) mass gathering places, (b) common areas of residential complexes, and (c) public points of police assistance. The point data from the cameras was aggregated to a hexagonal grid, with a size selected to be as geographically precise as possible while containing one or more cameras from each category in most cells. The density of the cameras in each grid cell is described as high, medium and low. The qualitative description of the cameras' density is given in order to manipulate this information in BN. Besides, this description facilitates human reasoning and it can deal with human uncertain knowledge [Oss01]. The point grid, which is used for the final visualization, represents centers of the hexagons and provides an overview over the whole city area.

3.2. Visual Analytics Approach for Bayesian Network analysis

In this research, we apply BNs to represent probabilistic knowledge about spatially referenced data. The BN is a probabilistic graphical model that gives a concise representation of a joint probability distribution on a set of statistical variables [Pea85]. The BN model is built on Bayes theorem (see equation 1) and described by qualitative and quantitative components such as Directed Acyclic Graphs (DAGs) and Conditional Probability Tables (CPTs), respectively. The reasoning process is based on modeling uncertainty in the reasoning rules (through CPTs) and the uncertainty in data sources (through the priors) [SPK98].

$$P(X|Y) = \frac{P(Y|X) \cdot P(X)}{P(Y)} \quad (1)$$

The qualitative component is a graphical model structure of the dependencies among the variables. The structure is realized by means of a Directed Acyclic Graph (DAG). The DAG consists of nodes (random variables) and edges that connect these nodes (see Fig. 1). Nodes are the labeled circles. Edges define probabilistic relations among nodes. Here, we introduce a VA-supported application of BN to explore the geographic coverage provided by integration of three kinds of public security cameras, particularly for use in monitoring activity associated with potential mass gathering events. The three data sets are represented are: residential complex cameras (R), mass gathering cameras (M), and police station surveillance cameras (P). Each is depicted in the BN as a node. Such discrete nodes are also called parent nodes, as they do not have predecessors. Furthermore, the parent nodes R, M, P, and the "Mass event" node are described by prior probability distributions. For the nodes, R, M, and P the prior probability distributions are given based on the density of the locations within a given grid cell and characterized using qualitative values: high, medium, and low.

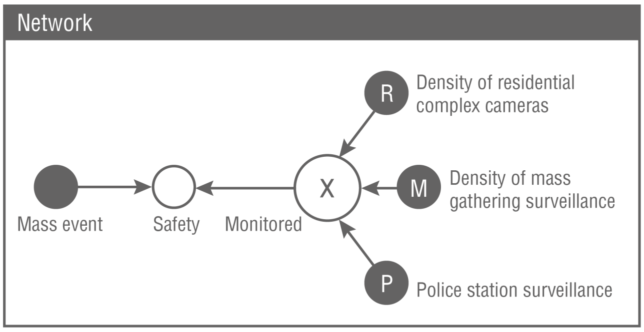

Figure 1: *Bayesian Network structure for reasoning on spatial heterogeneous public surveillance data.*

The quantitative component of the BN is determined by CPTs and (in the system presented here) can be elicited from an analyst via an interactive facility provided by the VA interface. The interaction process within the application is realized through the CPT panel, where the conditional probabilities can be adjusted in order to see how the distributions change. The Bayesian probability theory allows reasoning under uncertainty, thus, it provides a consistent mechanism to replace the uncertainty state with a numerical value in the interval [0, 1] representing degrees of truth, belief or plausibility [Ort10]. The aggregated sum of conditional probabilities equals to 1, thus by changing one value, other values are adjusted. The node X (Monitored) is a child node of R, M, and P; therefore, it is defined by a probability distribution over its outcomes (highly monitored, medium monitored, low monitored), conditional on the outcomes of its predecessors (nodes R, M, and P) (see Fig. 2). The size of the CPTs in the BN is exponential in the number of parent nodes. In the case given, three nodes can generate a table with 27 columns. Given that the node X includes a prior probability distribution with three outcomes (highly monitored, medium monitored, low monitored), the CPT for this node has three rows, therefore, there are 81 possible intersections.

Accordingly, the numerical values of subjective belief in the interval [0, 1] are introduced through the user-interface using a bar chart representation within panels "Monitored" (Fig. 2) and "Safety"(Fig. 3). In the panel "Monitored", a potential analyst can introduce values within the CPT on how the data given are related. For example, an analyst assumes that a high density of police and residential complex cameras and a medium density of mass gathering cameras result in the city zone being classified as high monitored with a given probability. Such an assumption might be made due to the fact that mass gathering cameras have a wider angle of observation, thus fewer cameras are needed to cover an area completely. The same procedure is valid for the node "Safety" as it is defined by outcomes of its predecessors: "Monitored" and "Mass Event" (see Fig. 3). The numerical parameters of the CPTs are visualized and the user can change them through the user interface to represent their subjective beliefs.

Figure 2: *Conditional Probability Table for the node X "Monitored". The CPT can be elicited from an analyst via an interactive facility provided by the VA interface.*

The inference outcome is registered in the data attribute and it includes a categorical value "noticed/unnoticed" and a numerical value, which is the value of the conditional distribution. In this research, we chose proportional circles that characterize the likelihood of different city zones being monitored (Fig. 4). The proportional circles represent the data set discretely, showing patterns where the cameras' coverage is dense, rather than representing an interpolated value over the city district. The size and color of circles directly signify the probability value at a given location and may be updated when new evidence is set (for instance, evidence

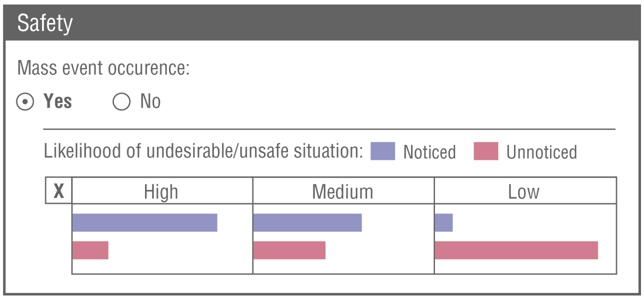

Figure 3: *Conditional Probability Table for the node "Safety".*

of a mass gathering event or updated conditional distribution with CPT for "Monitored"). Apart from interaction via the CPT panel, the user can browse the map and obtain details about each point.

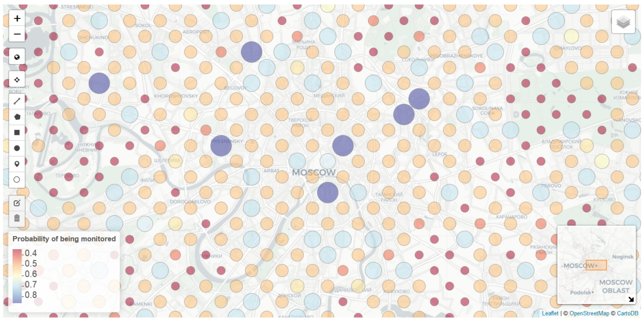

Figure 4: *Map panel. The size and color of the proportional circles characterize the likelihood of different city zones being monitored given the probability value in the interval [0, 1]. The Map panel also includes zoom, selection, and layer controls, map legend, and an overview map for a rapid navigation.*

To produce an extensible, generally applicable system, the VA-BN application has been implemented using interoperable methods in an open system setting. Specifically, the application prototype is developed based on the three-tier architecture including Database Management System (DBMS), namely PostgreSQL, Servlet Engine (GeoServer and R) and Client side components (R Shiny). The implemented application utilizes R packages bnlearn, that provides BN structure, parameter learning and inference [Scu09], and gRain that implements propagation algorithm in BN [Høj13].

4. Discussion

Targeting a classification task focused on heterogeneous spatial data under states of uncertainty, we combine a Visual Analytics (VA) approach and Bayesian statistics that address four objectives: (i) evaluate the feasibility of using a probabilistic graphical model, namely a Bayesian Network (BN), to represent conditional dependencies utilizing heterogeneous spatial data; (ii) provide visualization support where results of the inference can be observed in a spatial context; (iii) support users when introducing subjective beliefs for characterizing conditional probabilities; (iv) exploit advances

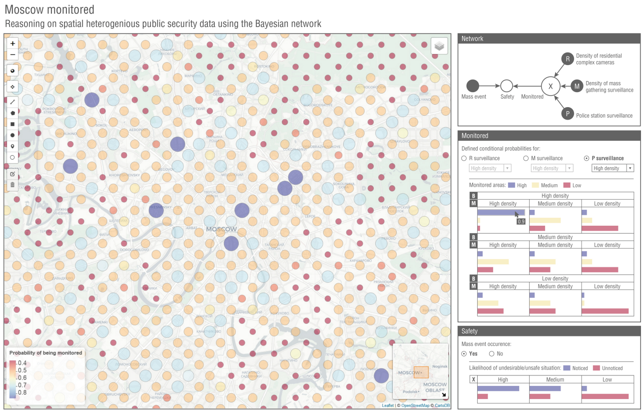

Figure 5: *Uncertainty-Aware Visual Analytics Interface. The interface is developed using data of public surveillance cameras installed in mass gathering places, common areas of residential complexes, and points of police assistance.*

in VA development by the integration of spatial data and computational capacity. We illustrate the capabilities of our proposed uncertainty-aware VA application by presenting a case study using surveillance camera data from three sources in Moscow: mass gathering places, residential complexes, and points of police assistance.

By exploiting the potential of VA development, the prototype uncertainty-aware application integrates data, computational capacity, and visualization facility within an application that supports human-computer interaction processes. The developed application is realized through support of connections among spatial data, BN modeling, visualization, and the users in order to examine city areas monitored by public camera facilities under the subjective judgment of the spatial data availability, quality, and relevance. The method introduced is capable of obtaining probabilistic predictions based on user input and the BN model structure; therefore it combines both human and Bayesian reasoning. The user interaction involves user-defined quantitative input and an interactive map. The analysis is driven directly by inputs formulated by the analysts through the user interface. The output of the analysis is visualized using proportional circles to signify the probability of public safety under consideration of a mass event occurrence. The results of the analysis of heterogeneous public security data aim to complement the existing city models with insights for designing effective city planning strategies and supporting decision-making in environmental modeling.

In future research, we plan to conduct a usability study with a group of analysts to evaluate the feasibility of applying the proposed method in different scenarios. Another issue to be tackled is scalability of the interface design when more data is fused into the BN as the conditional probabilities will grow exponentially.

Acknowledgements

This work is supported by the International Graduate School of Science and Engineering (IGSSE), Technical University of Munich.

References

[Blo06] BLOCH I.: Spatial Reasoning under Imprecision Using Fuzzy Set Theory, Formal Logics and Mathematical Morphology. *International Journal of Approximate Reasoning 41*, 2 (2006), 77 – 95. Advances in Fuzzy Sets and Rough Sets. URL: http://www.sciencedirect.com/science/article/pii/S0888613X05000502, doi:https://doi.org/10.1016/j.ijar.2005.06.011. 1

[CCKL09] CORREA C. D., CHAN Y. H., KWAN-LIU M.: A Framework for Uncertainty-Aware Visual Analytics. *VAST 09 - IEEE Symposium on Visual Analytics Science and Technology, Proceedings* (2009), 51–58. doi:10.1109/VAST.2009.5332611.

[CE17] CHAMPION C., ELKAN C.: Visualizing the Consequences of Evidence in Bayesian Networks. *CoRR abs/1707.00791* (2017). URL: http://arxiv.org/abs/1707.00791, arXiv:1707.00791. 2

[CMS11] COSSALTER M., MENGSHOEL O. J., SELKER T.: Visualizing and Understanding Large-Scale Bayesian Networks. *AAAI-11 Workshop on Scalable Integration of Analytics and Visualization 23*, August (2011), 12–21. URL: http://works.bepress.com/ole{_}mengshoel/23/. 2

[CSLG05] CHIANG C.-H., SHAUGHNESSY P., LIVINGSTON G., GRINSTEIN G.: Visualizing Graphical Probabilistic Models. *Technical Report, UML CS* (2005). 2

[HLH15] HOUSE L., LEMAN S., HAN C.: Bayesian Visual Analytics: BaVA. *Statistical Analysis and Data Mining 8*, 1 (2015), 1–13. doi:10.1002/sam.11253. 2

[Høj13] HØJSGAARD S.: Bayesian Networks in R with the gRain package. *Aalborg University 46*, 2012 (2013), 1–17. URL: http://cran.r-project.org/web/packages/gRain/vignettes/gRain-intro.pdf. 4

[KKEM10] KEIM D., KOHLHAMMER J., ELLIS G., MANSMANN F.: Mastering the Information Age: Solving Problems with Visual Analytics, Eurographics Association, 2010.

[KN10] KORB K. B., NICHOLSON A. E.: Chapter 1: Bayesian Reasoning. *Bayesian Artificial Intelligence* (2010), 3–28. doi:10.1016/j.cognition.2011.06.008. 1

[KTB*18] KUMPF A., TOST B., BAUMGART M., RIEMER M., WESTERMANN R., RAUTENHAUS M.: Visualizing Confidence in Cluster-Based Ensemble Weather Forecast Analyses. *IEEE Transactions on Visualization and Computer Graphics 24*, 1 (2018), 109–119. doi:10.1109/TVCG.2017.2745178. 2

[LLZ*17] LIN X., LI H., ZHANG Y., GAO L., ZHAO L., DENG M.: A Probabilistic Embedding Clustering Method for Urban Structure Detection. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-2/W7* (2017), 1263–1268. URL: https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLII-2-W7/1263/2017/, doi:10.5194/isprs-archives-XLII-2-W7-1263-2017. 1

[Mac15] MACEACHREN A. M.: Visual Analytics and Uncertainty: Its Not About the Data. In *EuroVis Workshop on Visual Analytics (EuroVA)* (2015), Bertini E., Roberts J. C., (Eds.), The Eurographics Association. doi:10.2312/eurova.20151104. 1, 2

[MRH*05] MACEACHREN A. M., ROBINSON A., HOPPER S., GARDNER S., MURRAY R., GAHEGAN M., HETZLER E.: Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know. *Information Science 32*, 3 (2005), 139–160. URL: http://www.tandfonline.com/doi/abs/10.1559/1523040054738936{%}5Cnpapers3://publication/doi/10.1559/1523040054738936, doi:10.1559/1523040054738936. 2

[OC09] OAKSFORD M., CHATER N.: Precis of Bayesian Rationality: The Probabilistic Approach to Human Reasoning. *Behavioral and Brain Sciences 32*, 1 (2009), 69–84. doi:10.1017/S0140525X09000284. 2

[Ort10] ORTEGA P. A.: Logic, Reasoning under Uncertainty and Causality. *Neural Information Processing Systems math.ST* (2010), 13. URL: http://eprints.pascal-network.org/archive/00007001/, arXiv:1008.2780. 3

[Oss01] OSSEIRAN A.: Qualitative Bayesian Networks. *Information Sciences 131*, 1 (2001), 87 – 106. URL: http://www.sciencedirect.com/science/article/pii/S0020025500000232, doi:https://doi.org/10.1016/S0020-0255(00)00023-2. 2

[Pea85] PEARL J.: Bayesian Networks: A Model of Self-Activated Memory for Evidential Reasoning. In *Proceedings of Cognitive Science Society (CSS-7)* (1985). 2

[Scu09] SCUTARI M.: Learning Bayesian Networks with the bnlearn R Package. URL: http://arxiv.org/abs/0908.3817, arXiv:0908.3817, doi:10.18637/jss.v035.i03. 4

[SDC14] SQUICCIARINI A. C., DUPONT J., CHEN R.: Online Abusive Users Analytics Through Visualization. In *Proceedings of the 23rd International Conference on World Wide Web* (New York, NY, USA, 2014), WWW '14 Companion, ACM, pp. 155–158. URL: http://doi.acm.org/10.1145/2567948.2577019, doi:10.1145/2567948.2577019. 2

[SPK98] STASSOPOULOU A., PETROU M., KITTLER J.: Application of a Bayesian Network in a GIS-based Decision Making System. *International Journal of Geographical Information Science 12*, 1 (1998), 23–46. doi:10.1080/136588198241996. 1, 2

[SRK16] SMITH MASON J., RETCHLESS D., KLIPPEL A.: Domains of Uncertainty Visualization Research: a Visual Summary Approach. *Cartography and Geographic Information Science 0406*, May (2016), 1–14. URL: http://www.tandfonline.com/doi/full/10.1080/15230406.2016.1154804, doi:10.1080/15230406.2016.1154804. 2

[VC16] VILLA V., COZZANI V.: Application of Bayesian Networks to Quantitative Assessment of Safety Barriers' Performance in the Prevention of Major Accidents. *Chemical Engineering Transactions 53* (2016), 151–156. doi:10.3303/CET1653026. 1

[ZRNG99] ZAPATA-RIVERA J. D., NEUFELD E., GREER J. E.: Visualization of Bayesian Belief Networks. In *IEEE Visualization 1999 Late Breaking Hot Topics Proceedings* (1999), Press, pp. 85–88. 2