Concept Mapping: Renewing Learning and Thinking
Proc. of the Eighth Int. Conference on Concept Mapping
Medellín, Colombia 2018

DO CMAPS AFFECT HOW WE PERCEIVE PROBABILISTIC INFORMATION?

Natalia Derbentseva[1], Frank Safayeni[2], Peter Kwantes[1] & Shadi Ghajar[1]
[1]Defence Research and Development Canada, Canada
[2]University of Waterloo, Canada
Email: {natalia.derbentseva, peter.kwantes, shadi.ghajar}@drdc-rddc.gc.ca; fsafayeni@uwaterloo.ca

Abstract. Concept maps (Cmaps) have been effectively applied in many different fields, including intelligence analysis. One of the fundamental characteristics of intelligence information is the uncertainty that is associated with available evidence or derived assessments. While Cmaps have the necessary capability to represent uncertainty, it is not well understood how the graphical nature of Cmaps affects the interpretation and comprehension of that information. A study was conducted to investigate this issue. The same information was presented to participants in either a text format or as a Cmap; and participants were asked to estimate the likelihood of seven statements based on the provided information. For two of the seven questions participants' estimates were significantly different, with the Cmap group providing more conservative values than the text group. Although, the obtained results suggest that the Cmap format may affect how map readers interpret probabilistic information, however, more research is needed to better understand the exact impact of Cmap representation of information on perception of uncertainty.

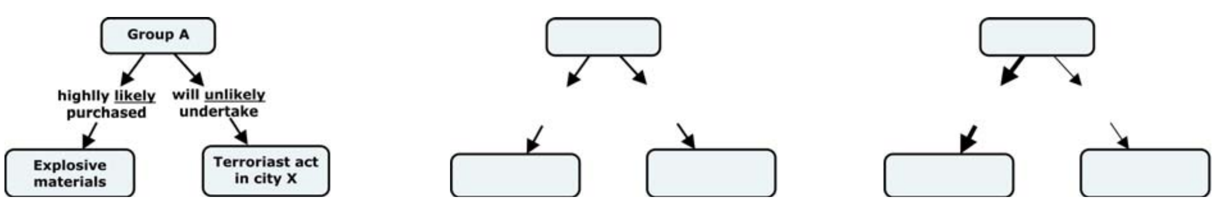
Keywords: concept mapping, communication of uncertainty, likelihood

1 Introduction

Concept maps (Cmaps, Novak & Govin, 1984) are an efficient graphical knowledge representation technique, which has gained a wide application in various fields including intelligence analysis (e.g., Huer & Pherson, 2015; Derbentseva & Mandel, 2011). The use of Cmaps in intelligence analysis offers many potential advantages, such as providing analysts with a mechanism to decompose, externalize and articulate the problem space, to support information integration, inference-making and to reduce cognitive demands (Hoffman & Shattuck, 2006). In addition, a Cmap is a visual product that can be used to communicate the results of analyst's thinking both within the analytic team and to the consumers of intelligence products.

One of the inherent properties of intelligence analysis is dealing with uncertainty both in the available information and in the conclusions drawn from that information. For example, intelligence analysts may use such probabilistic terms as "unlikely", "even chance", "likely" or "almost certain" to express different degrees of uncertainty in their products. Therefore, it is very important to understand how the tools that analysts use in their work, including Cmaps, affect the representation and communication of probabilistic information.

Through the requirement to label all relationships, Cmaps have the capability to represent a variety of different types of relationships in a single map, including those with different levels of uncertainty. Structure and text
b) Structure only, no text, same line styles c) Structure only, no text, different line styles
Figure 1 a) illustrates a simple example of two propositions that have different probabilities associated with them.



a) Structure and text b) Structure only, no text, same line styles c) Structure only, no text, different line styles

Figure 1: Three views of a Cmap with different probability relationships.

However, it is not clear how the graphical nature of Cmaps impacts the interpretation and comprehension of probabilistic information. For example, if all textual cues are removed from the example map in Figure 1 a), then the resulting graph in Figure 1 b) shows no indication that the two relationships have different strength. Could these visual cues (i.e., visual similarity) impact a map reader's perception of the likelihood information expressed in the linking phrases and reduce the perceived difference in their likelihood? In other words, it can be argued that if a low probability relationship is represented with a regular line connection; the regular line may visually imply a stronger relationship (and thus suggesting a higher probability) than the linking phrase indicates. Similarly, it can be argued that manipulation of the visual cues in a map, e.g., the thickness of the connecting lines such as shown in Figure 1 c), can help differentiating among the strength of the relationships.

While many aspects of Cmapping have been investigated (e.g., CMC 2004 – 2016); there are no empirical studies that examined whether the Cmap representation has an impact on communication and perception of uncertainty. To begin filling this gap a study was conducted to investigate how map readers interpret uncertainty represented in Cmaps. The purpose of the study was two-fold:

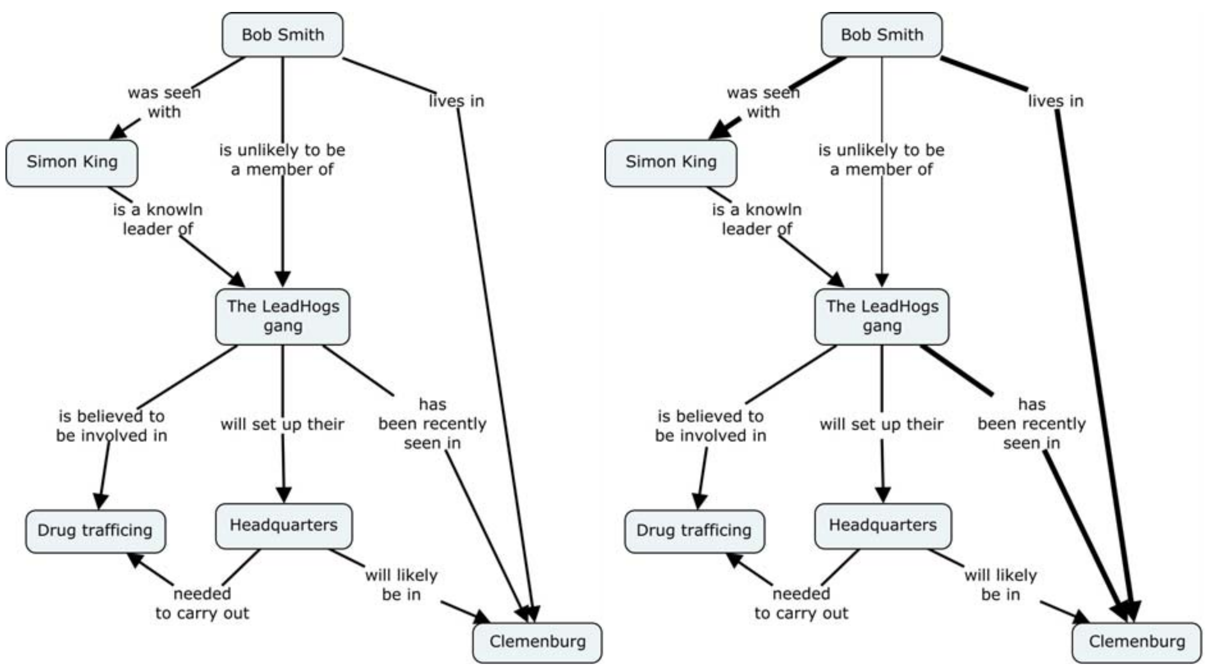
1. Examine whether the Cmap representation of information affects the reader's comprehension of uncertainty compared to the textual format;
2. Examine whether manipulation of visual cues in a Cmap has an effect on the reader's comprehension of uncertainty information.

2 Method

2.1 Stimuli

A short text paragraph and two Cmap representations of the text were used as stimuli; all are shown in Figure 2. The two Cmap representations of the paragraph were created by an experienced Cmapper and were designed to be as close to the paragraph as practically possible. The only difference between the two Cmap representations was the thickness of the connecting lines – Cmap1 had homogeneous line thickness, and Cmap 2 had varied line thickness with higher likelihood relationships represented with thicker lines than lower likelihood relationships.

- a) Text paragraph: Bob Smith lives in Clemenburg and recently was seen with Simon King, who is a known leader of the LeadHogs gang. However, it was assessed that Bob Smith is unlikely to be a member of the LeadHogs gang. The LeadHogs gang has been recently seen in Clemenburg. It is believed that the LeadHogs gang is involved in drug trafficking. Current intelligence suggests that the LeadHogs gang will likely set up their headquarters in Clemenburg to carry out their dealings.



b) Cmap # 1, homogeneous line styles

c) Cmap #2, with varied line thickness

Figure 2. Stimuli: a) text paragraph; b) Cmap 1; and c) Cmap 2.

2.2 Procedure

The study included three random assignment conditions based on the format of information participants received – Paragraph, Cmap 1 and Cmap 2. After reading the study information and consenting to participate, participants were asked to estimate the likelihood of seven statements based on the information provided in the paragraph or one of the Cmaps using a scale from 0% (no chance at all) to 100% (certain) with 10% increments and a "Not Applicable" option. The seven questions posed to the participants are listed in the second column of Table 1. Four types of questions were included in the list: *facts* (questions 1-3); *high likelihood statements* (questions 4-5); *low likelihood statement* (question 6); and *inference* (question 7). The order of questions was randomized for each participant.

A total of 150 participants were recruited for this study using Amazon's Mechanical Turk online recruitment tool. Data was collected through an on-line computer task administered through a survey administration platform SurveyMonkey. To ensure that the participants paid attention to the task, an unambiguous task question was included at the end of the task. Data from eight participants who failed to answer this question correctly was excluded from the analysis. Majority of participants (85%) had no or very limited previous experience with Cmaps; 65% were male; 83% were 18 – 44 years old and 17% were over 45 years old.

3 Results

The observed median likelihood estimates along with their 95% confidence intervals for each of the seven questions are reported by information format in the three right most columns of Table 1. Kruskal Wallis test indicated that there was a significant difference in groups' estimates for question #4 (*High likelihood statement*: "*The likelihood that The LeadHogs gang is involved in drug trafficking*") $\chi^2(2, N = 142) = 9.85$, $p = .007$ and question #7 (*Inference*: "*The likelihood that Bob Smith will participate in drug trafficking in Clemenberg*") $\chi^2(2, N = 142) = 7.162$, $p = .028$.

Follow-up Mann-Whitney U pairwise comparisons were performed controlling the False Discovery Rate (FDR) at $\alpha = .05$ using the Benjamini-Hochberg procedure. These comparisons indicated that mean ranks for both question #4 and #7 were significantly lower in the Cmap 1 group than in the Paragraph group.

#	**Question: What is the likelihood that**	**Corresponding statement in the stimuli**	***Type of statement***	***Expected estimate***	***Observed median estimates,% [95% Confidence interval]***		
					Paragr. (N=49)	*Cmap 1 (N =45)*	*Cmap 2 (N=48)*
1	Bob Smith lives in Clemenberg	Bob Smith **lives** in Clemenburg	*Fact*	*100%*	100 [100, 100]	100 [100, 100]	100 [100, 100]
2	Bob Smith knows Simon King	Bob Smith **was seen** with Simon King	*Fact*	*100%*	90 [80, 90]	90 [80, 100]	90 [80, 90]
3	Simon King is a leader of the LeadHogs gang	Simon King **is a known** leader of the LeadHogs gang	*Fact*	*100%*	100 [100, 100]	100 [90, 100]	100 [90, 100]
4	The LeadHogs gang is involved in drug trafficking	It **is believed** that the LeadHogs gang is involved in drug trafficking	*High likelihood statement*	*75%*	**80 [80, 90]**	**70 [70, 80]**	80 [60, 80]
5	The LeadHogs gang will set up their headquarters in Clemenberg	the LeadHogs gang **will likely** set up their headquarters in Clemenburg	*High likelihood statement*	*75%*	80 (80 – 80)	70 (70 – 80)	80 (70 – 80)
6	Bob Smith is a member of the LeadHogs gang	Bob Smith **is unlikely** to be a member of the LeadHogs gang	*Low likelihood statement*	*20%*	20 [10, 30]	20 [10, 20]	20 [10, 20]
7	Bob Smith will participate in drug trafficking in Clemenberg	***inference***	*Low likelihood inference*	*~20%*	**30 (30 – 50)**	**20 (10 – 30)**	30 (20 – 40)

Table 1: Likelihood estimation questions presented to the participants.

4 Discussion

The goal of this study was to examine whether there is a difference in how people interpret probabilistic information presented in a Cmap format compared to text. Participants in the study were asked to estimate the likelihood of four different types of information presented in the stimuli:

- Information presented as *facts* (questions 1-3);
- Information presented as *high likelihood statements* (questions 4-5);
- Information presented as *low likelihood statement* (question 6); and
- Derive an *inference* (low likelihood) from information in the stimuli (question 7).

Based on the argument set forth in the introduction, we hypothesized that there will be no difference among the three format groups in participants' likelihood estimations of the *factual* information (questions 1-3, **H1a**) and the *high likelihood statements* (questions 4-5, **H1b**). However, we expected Cmap 1 group to provide a higher estimation for the *low likelihood statement* (question 6, **H2a**) and the *inference* statement (question 7, **H2b**) than the Paragraph group. Furthermore, we expected the Cmap 2 condition to be closer to the Paragraph group in all of their estimates (**H3**).

As expected, we did not observe any significant differences among the three representation groups in their estimations of the *factual* information (questions 1-3), supporting H1a. However, a significant difference was observed between the Paragraph and Cmap 1 conditions on one of the *high likelihood statements* – question 4 – but not on question 5; therefore rejecting H1b. No statistically significant difference was observed among the groups for the *low likelihood statement* estimates (question 6), therefore rejecting H2a. The likelihood estimate of the *inference statement* (question 7) given by the Cmap 1 group was significantly lower than that of the Paragraph group, thus rejecting H2b, because we expected the difference to be in the opposite direction. However, the *inference statement* was partially dependent on the *low likelihood statement* (question 6), for which there was no significant difference among the groups. Therefore, there is no obvious logical explanation of the different estimation of the *inference statement* (question 7) provided by the Cmap 1 and Paragraph groups. There were no significant differences between the Cmap 2 and Paragraph groups, and between the Cmap 2 and Cmap 1 groups, suggesting that the estimates provided by the Cmap 2 group were in between the Cmap 1 and Paragraph conditions, thus supporting H3.

On the two questions where significant differences were observed between the Cmap1 and Paragraph groups, the Cmap1 group provided more conservative (i.e., lower) likelihood estimates. However, it is premature to draw any definitive conclusions from these results, especially given that these differences were not observed consistently across all of the questions. Our results suggest that the Cmap format may affect how map readers interpret probabilistic information compared to its textual representation; however more research is needed to better understand the exact nature of this relationship.

5 Acknowledgements

This work was conducted under the Joint Intelligence Collection & Analysis Capability project (JFD4 05da), Defence Research and Development Canada, and was supported by the University of Waterloo, Canada.

References

CMC (2004 – 2016). Proceedings of the biennial international Concept Mapping Conferences. CMC, available at http://cmc.ihmc.us/cmc-proceedings/

Derbentseva, N. and D.R. Mandel (2011). Using Concept Maps to Improve the Practice and Organization of Intelligence in Canada. In B. Moon, et al. eds. *Applied Concept Mapping: Capturing, Analyzing, and Organizing Knowledge.* New York: Taylor & Francis Group: CRC Press, p. 109-130.

Heuer, R.J. & R.H. Pherson (2015). *Structured analytic techniques for intelligence analysis*. 2nd Ed: Sage CQ Press.

Hoffman, R. and L.G. Shattuck, Col (Ret) (2006). Should we rethink how we do OPORDs. *Military Review*, March – April issue, p. 100-107.

Novak, J. D., & Gowin, D. B. (1984). *Learning How to Learn*. New York, NY: Cambridge University Press.