Measuring Collaborative Sensemaking

Tom Duffy
University of Birmingham
txd983@bham.ac.uk

Chris Baber
University of Birmingham
c.baber@bham.ac.uk

Neville Stanton
University of Southampton
n.staton@soton.ac.uk

ABSTRACT

Problems of collaborative sensemaking are evident in major incident response where sharing salient information is key to the shared understanding of the situation. In this paper we propose that differences in sensemaking performance can be captured through quantitative methods derived from consideration of network structure and information diffusion as the group collaborates to achieve consensus in a problem-solving task. We present analysis from a large international study in which groups of people collaborate to solve an intelligence analysis problem. Our initial analysis suggests that 'edge' groups are able to collaborate more efficiently and perform better than those which have a hierarchical control structure.

Keywords

Collaborative Sensemaking, social networks, intelligence analysis, consensus.

INTRODUCTION

Collaboration and teamwork during major incidents and disasters are difficult due mostly to the unprecedented nature of the event and the high levels of complexity and uncertainty involved. In such events, a shared understanding of the situation can be especially important during the initial stages of the incident in order to inform strategic level of response. We can see this as a problem of *collaborative sensemaking*. Analysis of sensemaking capabilities is usually done from the standpoint of human information processing or organizational dynamics. While these approaches provide insight into the *nature* of sensemaking, they do not provide indication of how well a group of people engages in sensemaking (except, of course, through qualitative assessment of the different groups). In this paper we propose that differences in sensemaking performance can be captured through quantitative methods derived from consideration of network structure and information diffusion as the group collaborates to solve the problem.

Sensemaking is an ongoing process through which meaning is constructed and explanations developed in an effort to establish what is going on (Weick 1995) and addresses the approaches by which agents make sense of the unknown, their reasons for doing this, and the properties/attributes of the output of the process. As such, one can consider sensemaking as a precursor to any decisions regarding response to a major incident, because it focuses on the fundamental question of what is happening (McMaster and Baber, 2012). We propose that these processes of sensemaking occur prior to, but play a key role in informing, situation awareness.

Weick (1995) suggests that sensemaking involves achieving consensus of plausible meanings. Weick does not elaborate on generic processes by which this might occur, although his book contains examples of these processes in action. In much the same way, instances of collaborative sensemaking have been observed in combat support headquarters (McMaster et al 2012), safety-critical incidents (Wahlström, 2011) and multi-agency emergency response (McMaster & Baber, 2012). These studies concentrate on the social interactions and cognitive processes of team members working collaboratively on tasks which involve sensemaking. The 'results' reported in these studies, therefore, take the form of discussions of observed aspects of behavior. It can be difficult to make generalizations about the collaborative sensemaking process from such studies since they involve bespoke groups working in specific habitats. Thus, the approach taken to study the sensemaking capabilities of groups tends to be based on case studies and interviews providing qualitative analysis. Approaches have been taken to apply more quantitative methods in this field of research. For example, Houghton et al (2006) in their study of C2 in emergency services operations and Stanton et al (2012) in their analyses of command structures, use social network analysis (SNA) to identify and explain the workings of groups, however these studies do not explicitly consider how structure might affect collaborative sensemaking.

In the field of disaster management we are usually working in exceptional circumstances and new organizations/ groups/ teams of people are forming around problems. These ad hoc teams are often only in existence for the duration of that particular problem. Emergency responders tend to be highly trained individuals working in teams, the purpose of training being to maintain the functionality of a community of practice (Wenger et al., 2002). When dealing with unknowns these groups work as networks of exploration (Baber et al, 2008). Such networks 'do sensemaking'. They start as loosely defined groups with little consensus and differing ideas to what the problem domain is. Over time they find a common understanding of the problem, a shared language for talking about it and an agreed way of responding. McMaster & Baber (2011) suggest that, in addition to emergency responders following Standard Operating Procedures, there is a need for them "to practice generic skills related to information sharing and collaborative sense-making". Additionally it is reasonable to suggest that there is a level of confidence you can assign to any output from a process where people have reached consensus. That output is not the end of the process, since the process is one of continuous refinement of a plausible working hypothesis. Finally, it is important to achieve consensus on this 'plausible working hypothesis' *before* moving on to activities related to situation awareness (which are involved in gathering evidence in support of this hypothesis). We are not claiming that sensemaking is *always* performed as a separate, precursor activity to situation awareness, and are aware of many situations (particularly in simple or routine incidents) in which sensemaking and situation awareness are inextricably linked. We are, however, proposing that complex incidents will require sensemaking (particularly in order for all parties to agree on the nature of the incident) and that this will feed into the ongoing processes associated with situation awareness of the responders.

In conclusion, when we talk about collaborative sensemaking, the emphasis is on the ability of a group to form, organize and reach consensus in the face of a novel problem. The studies on this to date have relied on case studies and interviews in an effort produce qualitative analysis of the collaborative sensemaking process. Our aim is to capture quantitative data that will give insights into the workings of sensemaking groups.

ELICIT

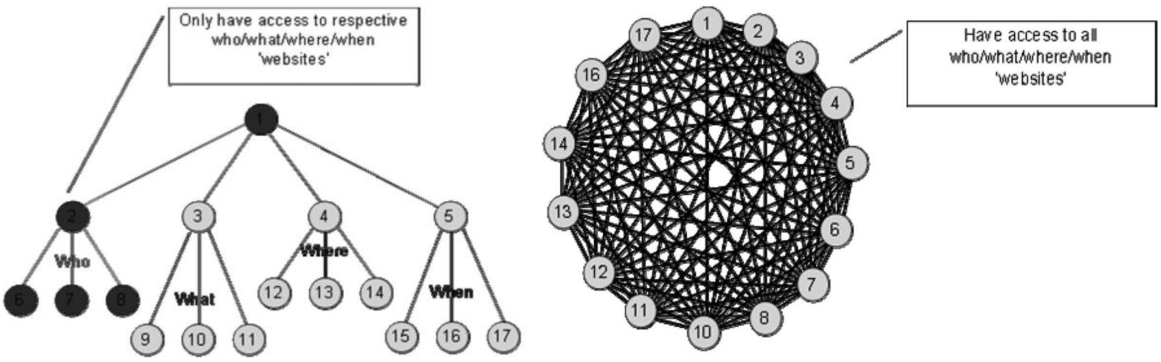

Figure 1 - Possible Connections in Hierarchical and Edge groups in the studies

The Experimental Laboratory for Investigating Collaboration Information-sharing and Trust (ELICIT)* is an experimental platform for running studies into the capabilities of different command and control structures. The platform provides a software environment for the observation and participation in a command and control task. The task in question involved groups of participants, organized in a pre-ordained command structures, sharing information snippets (factoids) in an attempt to collaboratively find the Who, What, Where and When of an anticipated terrorist attack. No one person has all the factoids and in order for a successful conclusion to be reached, players must share with each other.

The study from which we have data was an investigation into the differences in extremes of the NATO Approach Space, i.e., a hierarchical command and control structure and an edge organization (in which everyone can communicate with everyone in the network). These are illustrated by figure 1. We envisaged that the edge networks would represent a network of exploration to be compared against a more traditional command structure. The study had a total of 544 participants from five countries; Portugal, Singapore, Canada, United Kingdom and United States (Stanton et al, 2012).

* http://www.dodccrp.org/html4/elicit.html

PERFORMANCE CRITERIA

For each factoid-set in the ELICIT study there is a 'correct' answer. At regular points in the experiment, each member of the group submits a guess, informed by the factoids that they have obtained by that point in the experiment. For the purposes of our analysis, the 'correct' answer represents a ground truth against which these guesses can be assessed. This relationship between ground truth and guesses can be considered as a signal detection problem:

- Hit – Participant gives term that appears in the trial answer
- False Alarm – Participant gives term that is in the domain but not in the trial answer
- Miss – Hit not reported
- Correct Rejection – Words left out that are in the domain and aren't part of the given solution.

$$\left({No.Hits}/{No.HitsOfPossibleHits} \right) \times \left({FalseAlarms}/{AllKeyTerms} \right) \quad ^{\dagger}$$

It could be argued that creating a performance measure for sensemaking is inherently problematic. For this paper, the provision of a 'ground truth' allows us to make some comparison. In real incidents the definition of ground truth is thwart with problems but we suggest that it could be considered in terms of the consensus surrounding the incident responders definition of the situation. In other words, the movement towards consensus can be considered analogous to the construction of a ground truth that informs response. This is not the place to begin the debate on whether 'ground truth' reflects some objective reality of the situation or whether it is always socially constructed (either during or after the event). However, this assumption does offer us a valuable means of exploring the ways in which teams of people collaborate to make sense in a poorly defined problem space in which they need to gather pieces of information and then decide the most appropriate way to combine these into a meaningful answer. The analysis reported in this paper considers the performance of groups as they work towards their maximal performance and the ways in which factoids are passed between individuals in the network. In this instance, Density is an indication of the proportion of *actual* connections used by the network relative to the number of connections which are available in that network.

RESULTS

Table 1 shows the median performance scores of all groups at the end of each experiment. As one might expect, the edge networks tend to score higher on our measure of performance than the hierarchical networks (i.e., all in all groups, except for the two groups labeled 'Canada'). In terms of the networks that were formed during the performance of the task (i.e., counting the number of connections between individuals in the groups as the result of information exchanges), it is possible to derive a number of metrics applying Social Network Analysis statistics. For the purposes of this paper, we shall simply report the density of the networks. Figure shows that the density of the networks in the edge groups tends to be *lower* than those in the Hierarchical groups. Not only was density in the edge groups lower than in the hierarchical groups, but also the quantity of sharing was also generally lower in the edge groups. While most of the hierarchical groups demonstrate sharing in the region of 200-500 items, those in the Edge groups tend to have much lower totals of sharing.

	Performance (%)	
Group Name	**Edge**	**Hier.**
Boston	62.25	7.24
Canada	25.1	44.74
Cranfield	63.39	6.35
NPS	76.66	31.05
Singapore	70.1	37.6
Southampton	64.39	31.05
Westpoint	76.66	28.67

Table 1 - Median performance score of each group.

The results in table 1 imply that Edge groups tend to achieve better overall performance. Thus, the Edge groups seem to be more efficient than the Hierarchical groups. However, figure 2 suggests that the Edge groups (in which any member is able to connect with any other member) have less information sharing and make less use of available connections than those in the hierarchical network.

† *AllKeyTerms* represent all remaining terms after stop/noise word removal

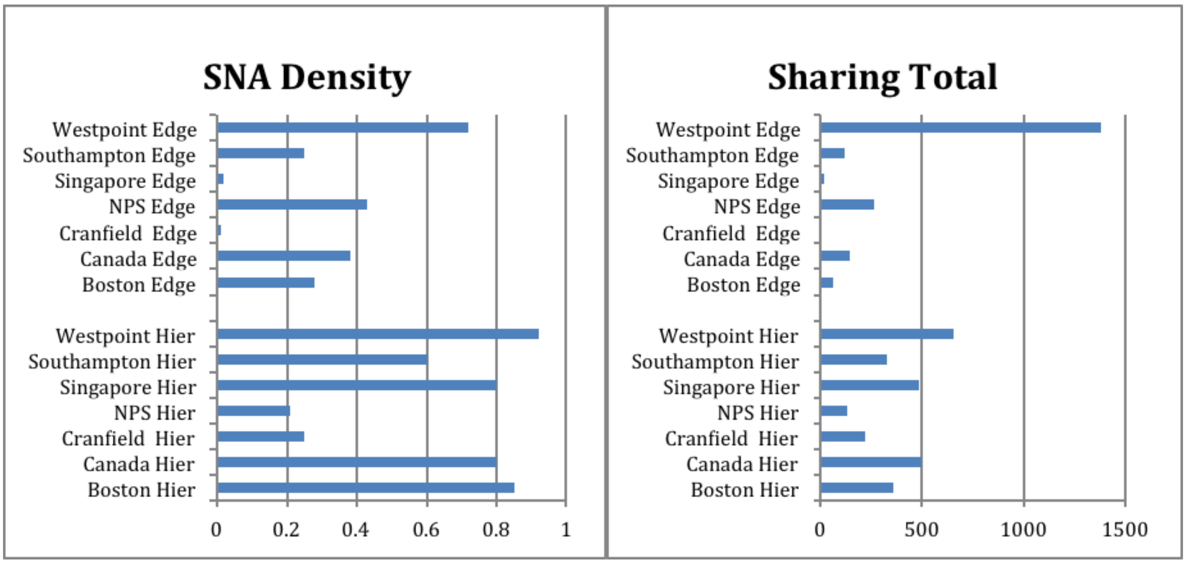

Figure 2 - Contrasting groups in terms of SNA Density and Total Factoids Shared

A line of best fit was added to the results sets (see Figure) to smooth the data points and view a general behavior of the group. A majority of the groups' results displayed a natural tendency towards a 'S' shaped curve and as such a Bass diffusion curve seemed to the most appropriate. This was given by the formula:

$$N(t) = N_{(t-1)} + p(m - N_{(t-1)}) + q\left(\frac{N_{(t-1)}}{m}\right)(m - N_{(t-1)})$$

Where 'p', 'q' and 'm' are coefficients of the curve. In traditional Bass diffusion models 'p' and 'q' are the coefficients of innovation and imitation respectively. 'm' is the ultimate final market potential. In the case of our results 'm' is actually the final median performance score of the group. Values of p and q were adjusted to find the lowest possible Mean Fitting Error (MFE) of the curve. The MFE was calculated from the root mean square of the sum of all the data point errors between the raw data and its respective point on the curve.

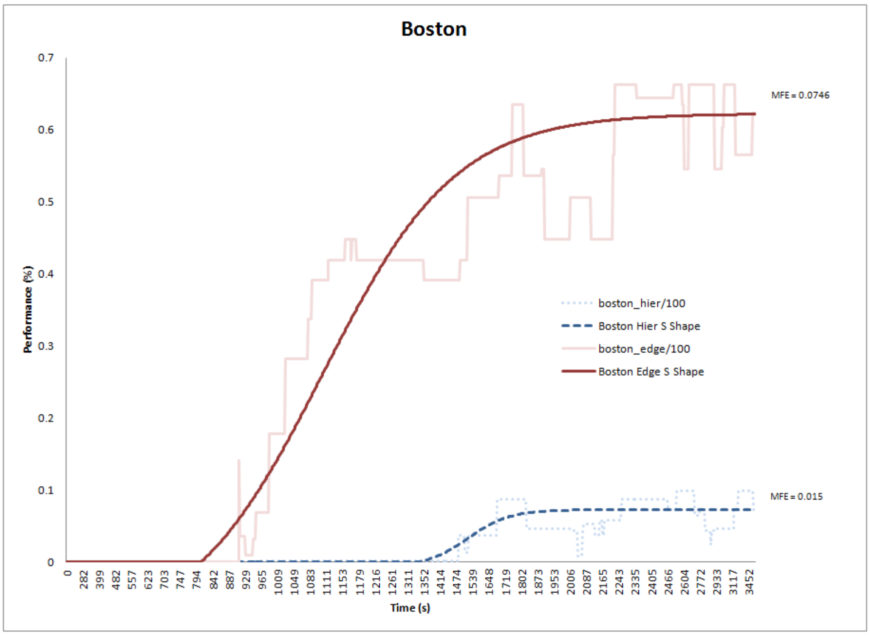

Figure 3 - Median Performance for Boston groups with fitted Bass 'S' shape curve

Broadly, the results seen in figure 3 are mimicked by the other groups' results. The Edge groups have curves which reach their asymptote earlier than Hierarchical groups and these asymptotes tend to represent much higher performance in the Edge groups.

DISCUSSION

From this initial analysis, we propose two main conclusions. First, it is possible to derive quantitative measures of performance that allow us to compare and contrast teams of people engaged in collaborative sensemaking. These measures related to three broad areas: information sharing, network structure and utility of information. While the latter measure can be defined in terms of ground truth and guesses for this study, we accept that this might be more difficult in real incidents. However, in terms of information sharing and network structure, we can easily count messages passed between individuals in networks and can analyze the structures of these networks as they grow. Second, it is possible to apply notions of information diffusion to compare and contrast the performance of groups engaged in sensemaking. What is interesting from our analysis is that the Edge groups exhibit more efficient information diffusion behaviors; they are more likely to share factoids which have a bearing on the solution and are quicker to share the solution with their peers in the network. Indeed, we would suggest that the behavior of the Edge groups are directed at establishing consensus on the 'sense' for the problem (the results imply that the Edge groups exhibit minimal sharing of irrelevant factoids and maximal sharing of solutions that use relevant factoids). In contrast, the behavior of the Hierarchical groups are directed at establishing distribution of the factoids (with maximal sharing of *all* factoids without emphasis on 'correct' solutions). While this looks as if we are proposing a 'winner' for this experiment, the real message we wish readers to take away is that Hierarchical structures may well be more efficient when information needs to be disseminated to all team members (or broadcast to defined subsets of team members) and when there is no constraint on the utility of this information, but Edge structures are more efficient when the information needs to be shared on the basis of specific constraints. The classic sensemaking problem is that you don't know which people to involve in the problem until you know what's going on but, paradoxically, you don't know what is going on until you've got people involved. So, the Edge network provides a 'network of exploration' which is a kind of rapid prototyping approach involving many people. We suspect that, once the network of exploration has made sense of the problem, the next stage in response would be to initiate Standard Operating Procedures and these are most appropriately managed through more traditional Hierarchical command networks.

REFERENCES

1. Baber, C., Stanton, N., Houghton, R. & Cassia, M. (2008) Hierarchical command, communities of practice, networks of exploration: using simple models to explore NEC command structures, *Proceedings of NECTISE Conference*, Leeds
2. Houghton, R.J., Baber, C., McMaster, R., Stanton, N.A., Salmon, P., Stewart, R. & Walker, G. (2006) *Ergonomics, 49*, 1204-1225
3. Klein, G., Moon, B. and Hoffman, R.F. (2006). Making sense of sensemaking 2: a macrocognitive model. *IEEE Intelligent Systems*, *21*, 88-92
4. McMaster, R., Baber, C. and Bond, R. (2012) Sensemaking in a combat support headquarters, *Journal of battlefield technology*, 15, 1, 27-34.
5. McMaster, R. & Baber, C. (2012) Multi-agency operations: Cooperation during flooding, *Applied Ergonomics 43*, 38-47
6. Stanton, N.A., Walker, G.H., & Sorensen, L.J. (2012): It's a small world after all: contrasting hierarchical and edge networks in a simulated intelligence analysis task, *Ergonomics*, *55*, 265-281
7. Wahlström, M., Salovaara, A., Salo, L., & Oulasvirta, A. (2011) Resolving safety-critical incidents in a rally control center. *Human Computer Interaction*, *26, 9*-37
8. Weick, K.E. (1995) *Sensemaking in Organizations*, London: Sage.
9. Wenger, E., McDermott, R. and Snyder, W. (2002) *Cultivating communities of practice: a guide to managing knowledge,* Harvard, MA: Harvard Business School