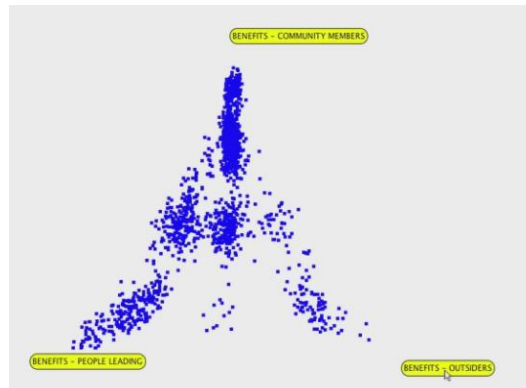


I have been listening to the [videos](#) that Cognitive Edge have made available on the workings of the SenseMaker Suite. They are very useful for getting an idea of how it works.

One thing that interested me was the distribution of stories in the triads (three cornered scales). They were as shown in the image below (1 dot = 1 story). What I see are seven clusters of stories: one at each corner, one in the middle of each side, and one in the centre.



What the respondents were doing, it seems, was *classifying* the story as a combination of one or more of the three types of benefits.

1. Only this one is true (one of the three points)
2. These two are true (one of the middles of each side)
3. All three are true (centre of triad)

They were not really rating the stories on a scale. In that respect the rating scale may be more sophisticated than is really necessary. If so, the survey instrument could have simply listed the three options, with yes/no buttons next to them. For example as drafted below:

1. Please tell us how you would classify your story, by choosing one or more of these descriptions below

	Yes	No
This story is about xxxx	<input type="radio"/>	<input type="radio"/>
This story is about yyyy	<input type="radio"/>	<input type="radio"/>
This story is about zzzz	<input type="radio"/>	<input type="radio"/>

The classificatory approach is used later on in the SenseMaker Kenya application, in the section titled “Multi-choice questions” where respondents are asked to pick up to three categories that their story relates to. In the sample data set that was shared with me most people (51%) made full use of this option, 19% chose two categories and 1% managed to get away with using four or five categories.

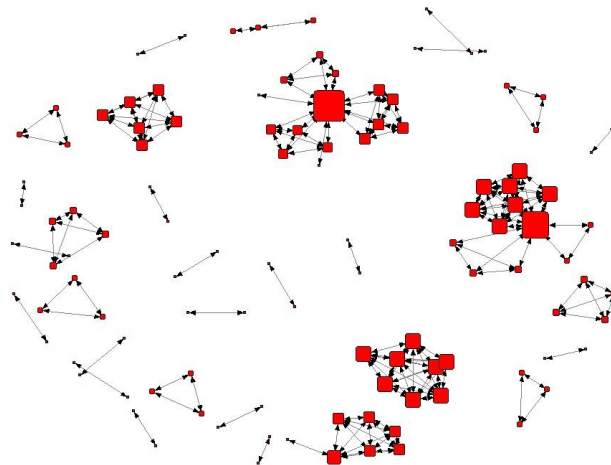
The advantage of using a longer list of non-exclusive categories like this is that you can capture a much richer picture of the many types of stories, as seen by the respondents. With the three categories in the triad shown above there were seven possible combinations of categories; with 12 categories in the multiple choice

question the number of category combinations is much higher (I am not even sure how to calculate them, but they escalate very rapidly). Respondents can self-organise their stories into a much more differentiated set of groupings. Self-organising seems to be at the heart of the idea of SenseMaker idea.

The question then is how to explore those groupings. There seem to be two possible ways of doing this. The first is simply to search for all stories that were categorised with two or more attributes of prior interest, say, FREEDOM and CREATIVITY. This can be done easily enough with Excel or Access. It should be relatively easy to design an interface where the respondents themselves could do this kind of exploration.

The second approach involves finding stories that shared two, three or more sets of categories of any kind. Because these can be large in number some kind of automated process is needed. One way of doing this is by using social network analysis software. The sample data that was shared with me was converted into a “two mode” matrix showing story headlines x categories, with cell values saying which categories applied to which stories. This data was then imported into [Ucinet](#) and converted into a “one mode” matrix, showing which stories were connected to which stories by sharing the same category, pair of categories or triple of categories. This data was then visualised, as a network of stories, using the associated NetDraw package. In the diagram below a filter has been applied to show only those stories who share three categories each.

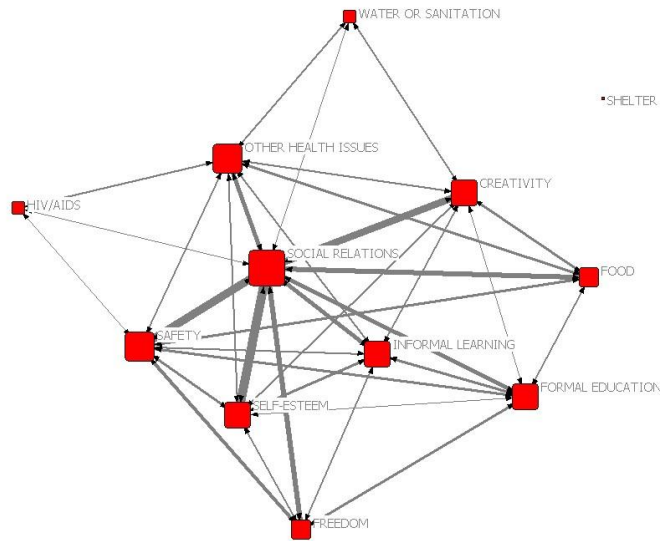
In the diagram there are number of story “cliques” (all connected to each other by the use of the same categories). My assumption is that the more categories a set of stories share, the more likely they are about the same topic overall. This assumption needs to be explored by looking at the contents of the linked stories.



PS: Because some stories were given 4 or 5 categories they stand out as “connectors” of two more cliques of stories. This muddies the waters a bit, and perhaps they should have been removed from the data set. The other complication is that the data set I examined seems to need more cleaning, there were a number of identically titled stories.

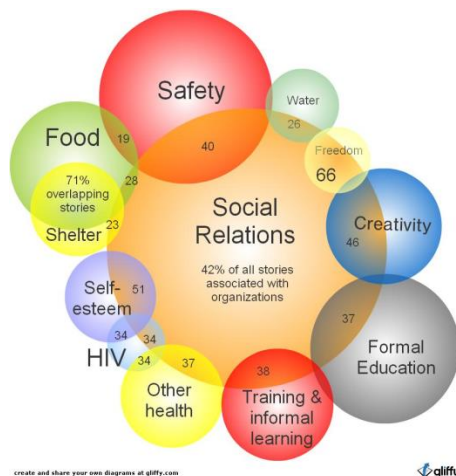
Returning to the two-mode matrix of data, this can also be converted into a second type of one made matrix. This is a categories x categories matrix, where cells describe the number of stories that categories by the each combination of categories.

With 12 categories there are many possible combinations of two, three or more categories. However, using Netdraw they can be visualised and then filtered to focus on the strongest relationships. The diagram below shows the strongest relationships between the categories (i.e. having an above average number of shared stories).



Key: Thicker lines = more shared stories, bigger nodes = more connected to other categories (alt: more stories in that category). N= 295 stories

This provides an alternative view to that provided in the recent [Huffington Post blog](#) (below). The use of a Venn diagram in that article clearly highlights the nature of the links with the Social Relations category, but at the expense of providing information about relationships between the other categories. These relationships are relevant. The diagram above, based on a sub-sample, suggests that the stories about HIV/AIDS are a very separate group to the stories about formal education. Likewise those about FREEDOM versus WATER OR SANITATION.



Possible future directions...

- Repeat the above analysis with a larger data and cleaner dataset. And follow up with an exploration of the actual contents of individual cliques of stories
- Carry out the same kind of analysis using respondent attributes, rather than story attributes as above.

1. IG: *From my admittedly quick read, your document is about how to do it not about why to do it. I still miss that. Is it about making the findings more 'rigorous' or analysing it differently to understand X? The latter comes up right at the end in the network/linkage analysis example but I am still left wondering. I can see that relationships/theme clusters are important in order to understand if eg working with 'food' initiatives has a stronger link to social relations than working with Hiv/AIDS initiatives. Then they give entry points for stimulating positive patterns, for example. Is that what you are trying to say? Making this more explicit in your document as part of your blog might make it easier for people to see the pertinence.*

- RD: My understanding is that self-indexing is seen as important because it enables some emergent structures within the respondents' data to be made visible, before the researcher starts to apply their own interpretative framework to structure the data. The cliques of similar stories are exactly that kind of emergent structure. They would be worth investigating.

2. IG: *Important - triads are NOT a yes/no question, although I would argue that depending on how their use is facilitated, they are treated as such by people in which case they get downgraded to an MCQ. Which is your point, I think? They are scales. Imagine each triad in a box with bottom left corner in the bottom left corner of the box. Then the location of the X or the bubble (if done on the web) is a scale from 0 to 100 (though I would argue that this is a level of granularity that is finer than what the person doing the index is considering). If they were treated entirely as ratings, then the clusters would be much tighter and less diffuse. Dave? I know you have thought about this.*

- RD: The triads are supposed to be seen and used as scales, but it seems from what little I have seen, that they are in fact used as a categorising device. That in itself is not a problem, but as I say, a simpler approach to categorising is available, and is used elsewhere in the survey instrument. And the use of multiple non-exclusive categories can enable a lot of self-organised structure to emerge.
-