ALBERTO COTTICA

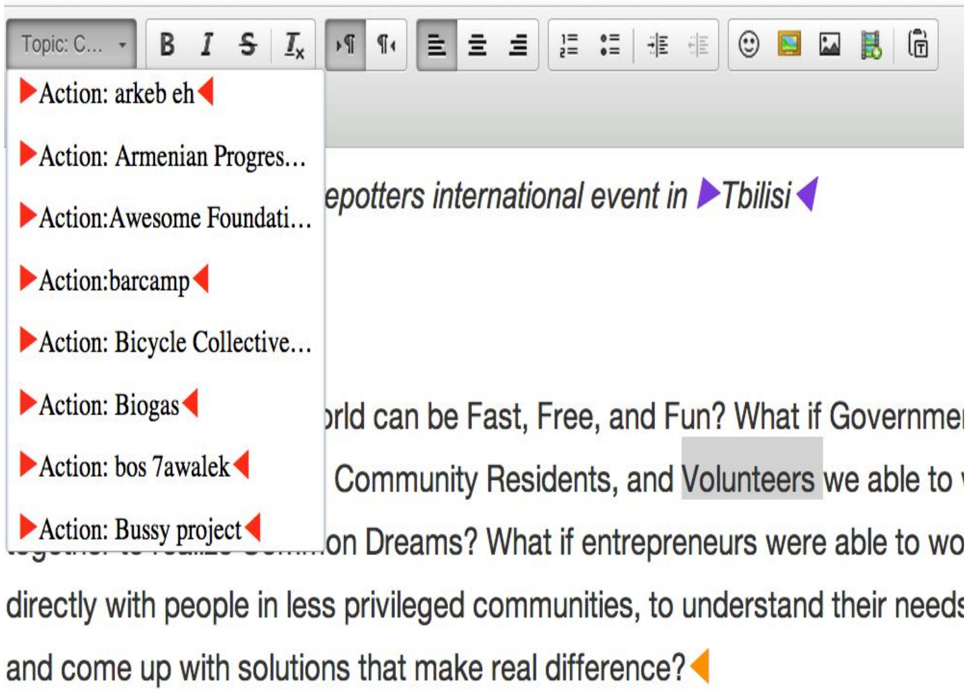
PEOPLE, WORDS, DATA: HARVESTING COLLECTIVE INTELLIGENCE

This is *La scuola di Atene:* 500 years ago, at the peak of the Renaissance and his own powers, Raphael paints it as a tribute to the wisdom and courage of humans seeking truth. The whole Top 40 chart of Greek philosophers is depicted: Pythagoras is right in the front left, studying a large tome. Epicurus also to the left, facing us, crowned with laurel. Socrates. Heraclitus. And in the center of the action: Plato and Aristotle. The two mighty philosophers are deep in discussion as they stride towards us. And this is only appropriate, because discussion, “dialogue” as Plato liked to call it is what powers their knowledge. Look around the fresco: debate is everywhere. Twenty-five centuries after Plato, and five after Raphael, this is still how science works.

Dialogue – conversation – is *not* just about accumulating information. It's a process that augments information, by setting it in a richer context. A well-run debate can feel like walking into a room with a fragment of map to find others that have its other pieces and ending up with a whole map. The whole is more, much more of the sum of its parts. And it never ends: questions beget answers and new questions, and it starts all over again, like a living thing. In fact, life itself works in a similar way, with genetic information being continuously traded amongst organism to give rise to always new, wonderful life forms.

This is why Edgeryders encourages truth-seeking, result-oriented conversation as a knowledge engine. Nobody is smarter than everybody, and the smarts is in the interaction – the conversation. But this raises a problem: conversations don't scale well. A hundred people cannot have a conversation, in the sense that they all keep a reasonably similar outlook on what is being discussed and what conclusions are being reached. They have to splinter into smaller groups. A great community convener, like John, or Noemi, can get a great many interesting people involved in discussing a problem. A lot of insights are generated and validated; but they are validated *locally*, by subsets of a few people. How to generalise? How can we tell which insights are solid, and which are the product of a branch of the conversation that simply misfired?
At Edgeryders, we do it in three moves.

The first move is online ethnography. Ethnography is a qualitative research technique that results in the description of a group that encodes the point of view of that group. Ethnographers prefer to work with no hypothesis to prove, simply letting their respondents take the study to whatever *they* think is important. So, the first thing we do when trying to make sense of an online conversation, is deploy one or more ethnographers.

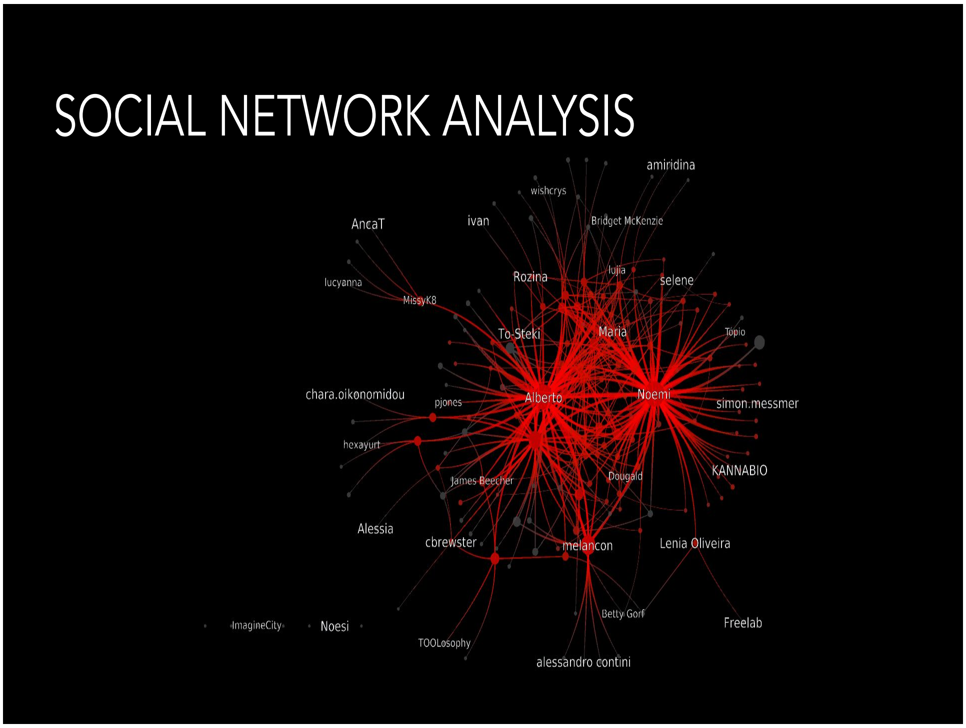


Ethnographers aggregate knowledge by interviewing people, and annotating the transcripts of their interviews. Annotations consist of a quotation from what the respondent has said, associated to one or more keywords. As she goes through the testimonies, the ethnographer builds and maintains an ontology of concepts and facts relevant to the issue being studied, as seen by respondents as a group. This is process is known as "coding". Nowadays this is done with computers, and the results entered in a kind of database. So, the researcher can call back all the quotations that have been coded with a keyword, say "greentech", and quickly get a grasp of what different people are saying about greentech. Do they agree? Is their controversy? What is important? What other concepts is greentech connected to? Etcetera.

In Edgeryders, we have an advantage: conversation happens on our platform, so ethnographers can start to code it right away – no need for interviewing and transcribing. We can be much faster and cheaper this way.

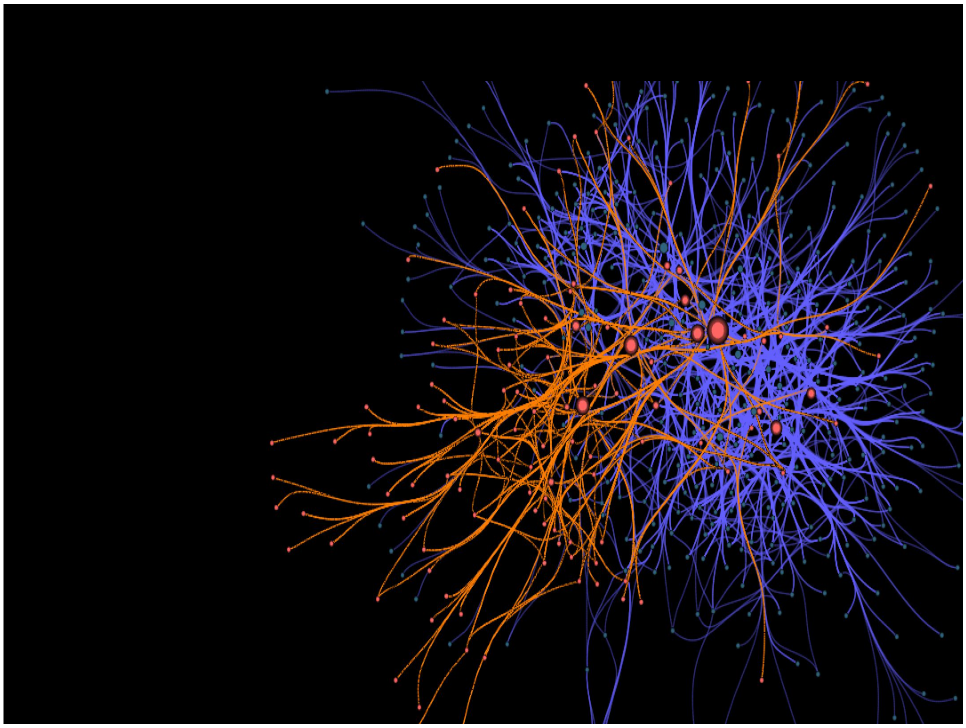
The Edgeryders platform features its own open source module to

do ethnographic coding of the content therein. We call it OpenEthnographer, because we have a vision of ethnography as a collaborative discipline.

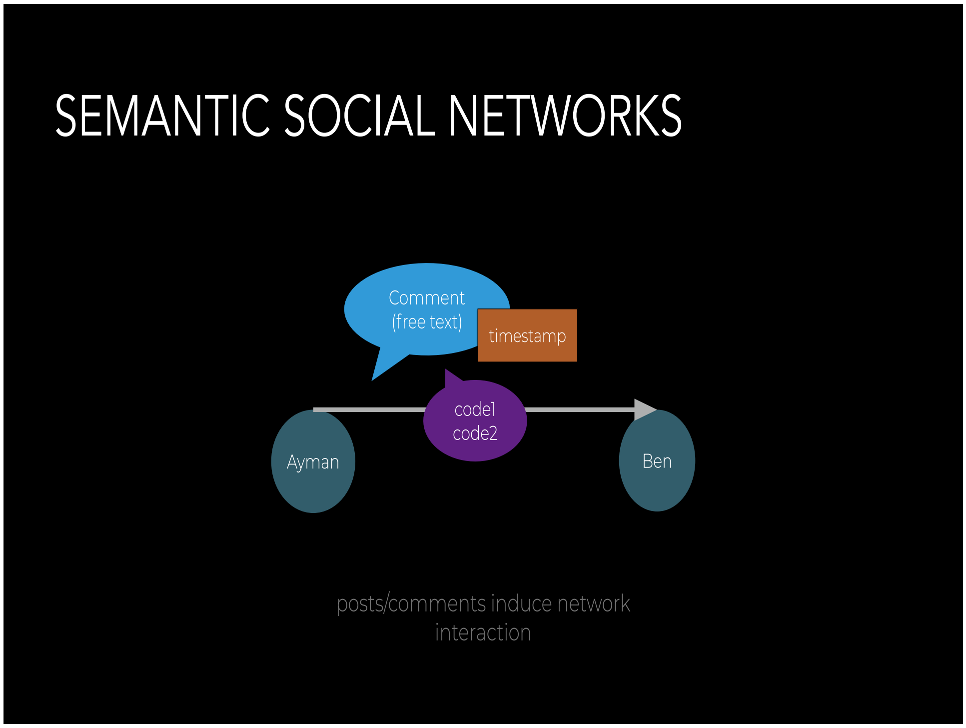


The second move is social network analysis. The Edgeryders conversation is encoded in a database – it has to be, for the platform to work at all. So, we can represent the discussion as a social network, where nodes represent people and arcs represent interactions – comments, essentially. At that point the pattern of connectivity in our conversation is described by a graph, and graphs are well understood mathematical objects. What's more, there is a rich literature in social science that associates measurable graph metrics to social phenomena like brokerage and authoritativeness. In general, the graph can help us tentatively assign reliability scores to individuals. In this one, for example, there is a densely connected core of about 40 people in the center, who can be assumed to have at least been exposed to what other people have to say. Their quotations carry more weight, because they are validating and correcting each other. Then there is a periphery of people that have only one or two connections, and even a few completely isolated nodes to the southwest. What these people have to say might be very wise and relevant, but it lacks in-group validation.

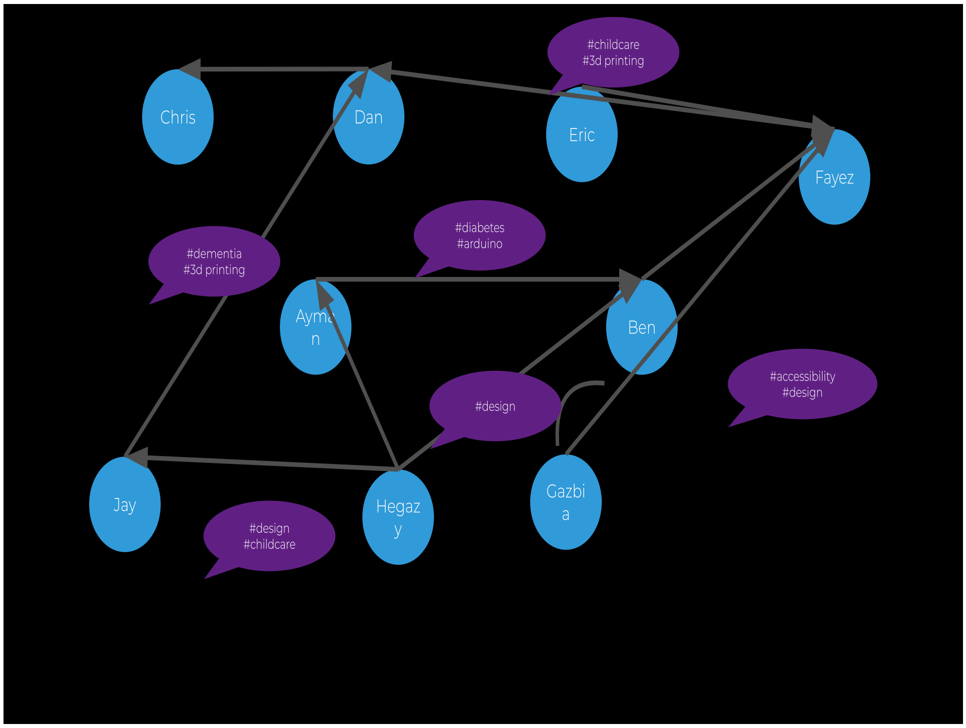
We developed software to do real-time social network analysis of the conversation on the Edgeryders platform. We call it Edgesense.



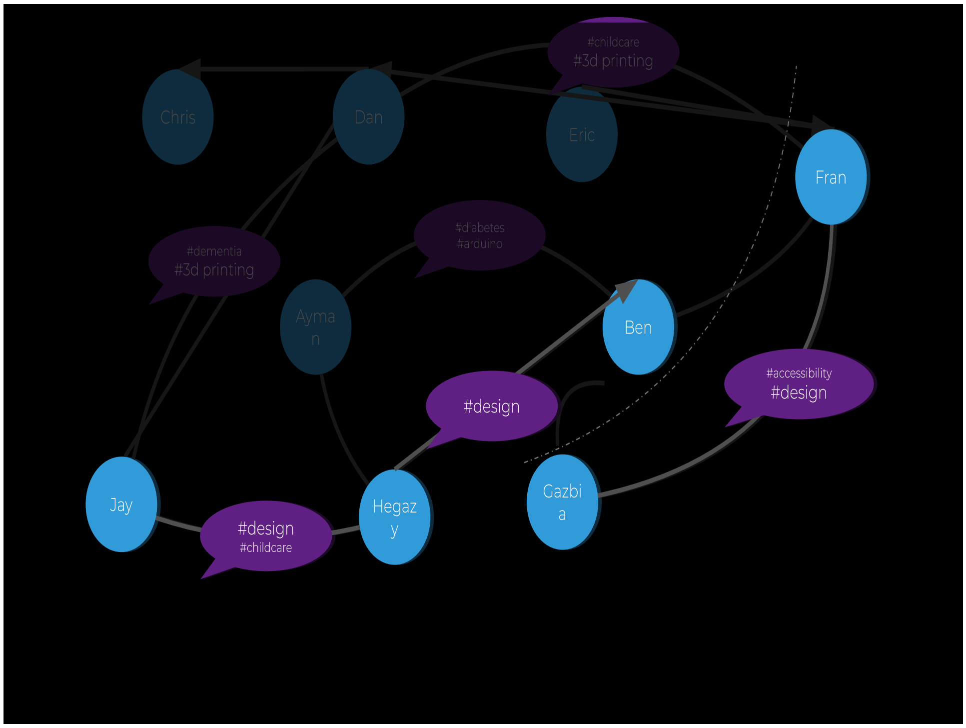
Thinking about collective intelligence in terms of networks has been fruitful for us. It generates intuition, and lets us do quick, simple checks on the general state of the community. For example, this visualisation comes from a project done two years ago. Like in the previous slide, dots represent people, and arcs represent comments. In orange, you see the new project; in blue, the Edgeryders conversation *outside* of the project. As you can see, the new conversation is well connected to the pre-existing one, through individual that participate in both; but is still maintains structural cohesion – you see that because the orange edges are grouped to the west of the centre. This is what we would hope to see: there is a healthy balance of specialists, that presumably bring their specific knowledge and interest; and generalists, who carry memory of previous debates and can connect the newcomers to the old-timers relevant to the new project.



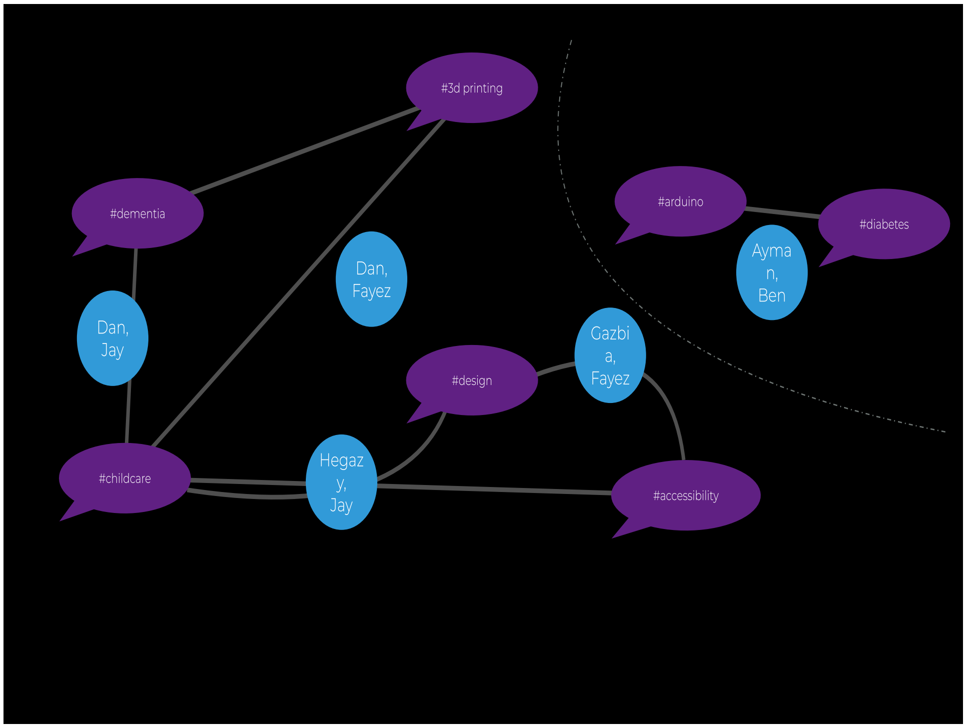
The third move is to pull in together ethnographic data and social network data to build something we call semantic social networks. The “atom” of a SSN is an interaction like this, where Ayman addresses a comment to Ben. In your social networking platform, the comment is made of some text; it has a source, Ayman, and a target, Ben; and it has a timestamp. After the ethnographer has worked her magic, it also has one or more ethnographic codes, so we know what this is about.



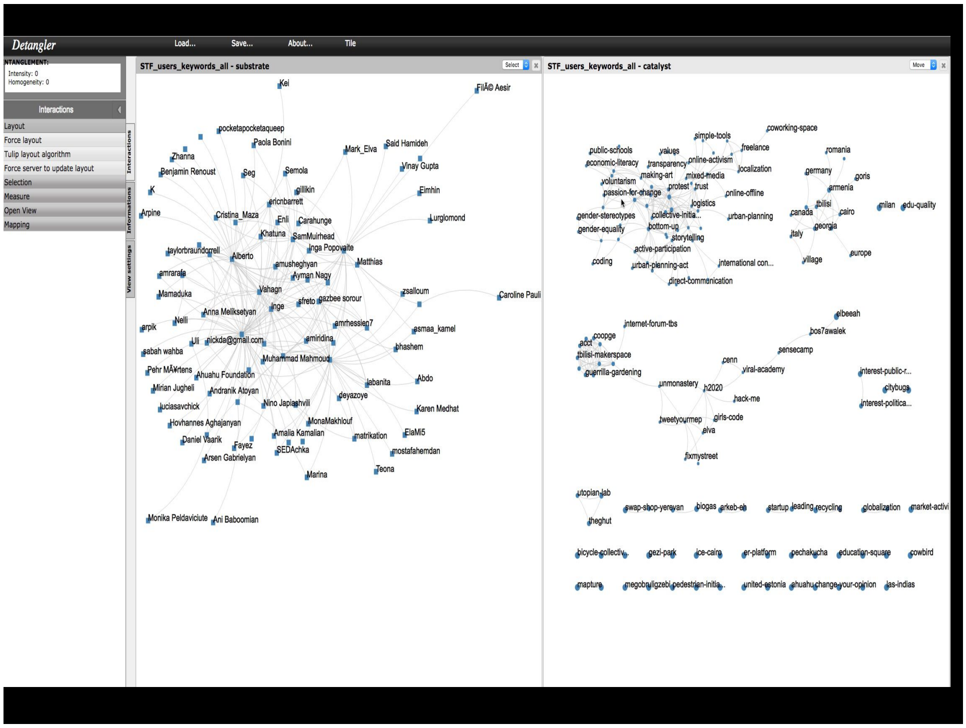
When you generalise this approach to a whole conversation, you get a fairly complex graph. You can do many things with it. Consider, for example, this toy network of a conversation about care.



A simple one is to filter it by ethnographic code. In our toy example, I have pulled out the network of design. I now know that Jay, Hegazy, Gazbia, Ben and Fran are interested in design. I notice that they are not really talking to each other; Jay, Hegazy and Ben form a group, Gazbia and Fran another. I know that the discussion on design in the context of care risks being incoherent, or balkanised. If the study is ongoing, I can even let Gazbia and Fran know that there is a group around Hegazy which shares their interest for design, maybe they will get in touch, share notes, and change both the content of what is being said and the structure of the graph.



You can also reverse the idea, and build a network of ethnographic codes. Now the arcs represent co-occurrences: two codes are connected when they appear together in one comment. In this toy network, it turns out, I have two disconnected clusters of concepts.



We are still inventing the math for SSNA, but already at an intuitive level we see that the method is very powerful. I can use it, for example, to identify the emergent groups of specialists gathering around certain ethnographic codes.

For example, here we have a SSNA from one of our previous studies. You decide you want to look into the subset of this conversation which is about education. you find and select the relevant codes, and am referred to the people that have been using those codes. As you can see, people in this group are part of a connected graph, but only some are directly talking to each other about education issues. They have education as a common interest, but they are not all comparing notes around it yet. You can also decide to ask the data what ELSE the people talking about education are also talking about. And here it is: this is like doing association patterns, but with a group. This works also if the people in question do not know they are part of the group.

We use a metrics called entanglement to measure each code’s

contribution to the cohesiveness of a group of codes. This is very new, it first appeared in a PhD thesis less than 3 years ago. As I said, the method is new. At each project we improve and iterate.

It is important to lift our gaze from the data themselves and thing about what exactly we are trying to do with them.

We are not trying to do sentiment analysis. We are not, because we model people differently. Sentiment analysis was developed in the context of commercial communication: you need to get good at pressing the behavioural buttons that trigger purchase decisions. It models people as desire machines – use a certain shade of blue on your website and increase infinitesimally the probability that the person will buy an insurance policy. We ask people to be smart, committed, hard-working: citizen experts, not “consumers”, “beneficiaries”, “target groups” etc. And we model them as such. This generates high quality, nuanced contributions. We need fairly sophisticated analytical instruments to do justice to that quality.

An obvious application of this method is as a risk assessment support tool. Situated knowledge, validated by debate, can monitor the situation of projects on the ground in real time, or

close enough. When operating in unstable contexts, that can be especially valuable.

Second, we could not do this on Facebook. This kind of data analysis requires technical and legal means to access and process the data – in other words, open data. The Edgeryders platform has both. We consider ourselves part of the open data community, and some of us, myself included, are open data activists in our spare time.

And third. We like data. But at the end of the day, conversation graphs are just shorthand. The real thing behind the data is collective intelligence, people in conversation; and that is incredibly rich and more interesting than any shorthand, like all emergent phenomena.

In our community, we are committed to put data analysis at the service of this ever-changing, ongoing dialogue. We are mapmakers, not landscapers, and we do not confuse the map with the territory.

If we work together, yes, we will make graphs with you. And then we will encourage you to use them to find the most interesting people, and engage with them and what they have to say. Our data describe people, and people are *really there.* You can engage them, look them up, challenge them, hire them, become their friends.

Hans Rosling got it just right: we don't look at the data. We look

through the data, to better see the people behind them.