



Project Acronym: **CATALYST**  
Project Full Title: **Collective Applied Intelligence and Analytics for Social Innovation**  
Grant Agreement: **6611188**  
Project Duration: **24 months (Oct. 2013 - Sept. 2015)**

**Masters of Networks 3 – Track 2 report  
Comparative analysis of interaction networks in online communities**

Date: **March 10-11, 2015**  
Location: **Rome, Italy**

Acknowledgement: **All participants of the Masters of Networks 3 event**



This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement n°6611188

# Masters of Networks 3: questions and methods

## The questions

*Frame questions from a domain point of view. That is, ignore network science for now: just ask what you would like to know!*

Are healthier communities those with a greater share of contributors content than managers content, or viceversa? [Noemi]

There could be a typical "user life-cycle": people sign up, they look around for a while, then they become more active as they get to learn the community they entered. Later still, they get bored and are less and less active, and finally go dead. Is this true? Is it true of ALL communities? And of ALL users? [Alberto]

Assuming there is a user cycle, is there a tipping point in that cycle in which there is maximum potential for conversion into an active user? An indicator or sorts? [Noemi]

In many social networks, connectivity – considered by some literature a measure of influence and power – is distributed in a very unequal way (power laws). Is this true of all online communities? Are there more egalitarian communities, and what can we say about them? [Alberto]

We've uploaded the gexf files to this google drive folder.

<https://drive.google.com/folderview?id=0B-sizRV1qoRXfk5vTINyVTAwZmVfY25PdFJpM3dxSFFwUVILVUpETko3QjhHamZfNEk5Ukk&usp=sharing>

## Notes

Q: Are healthier communities those with a greater share of contributors content than managers content, or viceversa? [Noemi]

We began by trying to find out if there's a correlation between moderators activity and community "health" conceptualized as users activity that's not directed back at moderators.

Independent var: the number of moderators' comments in time

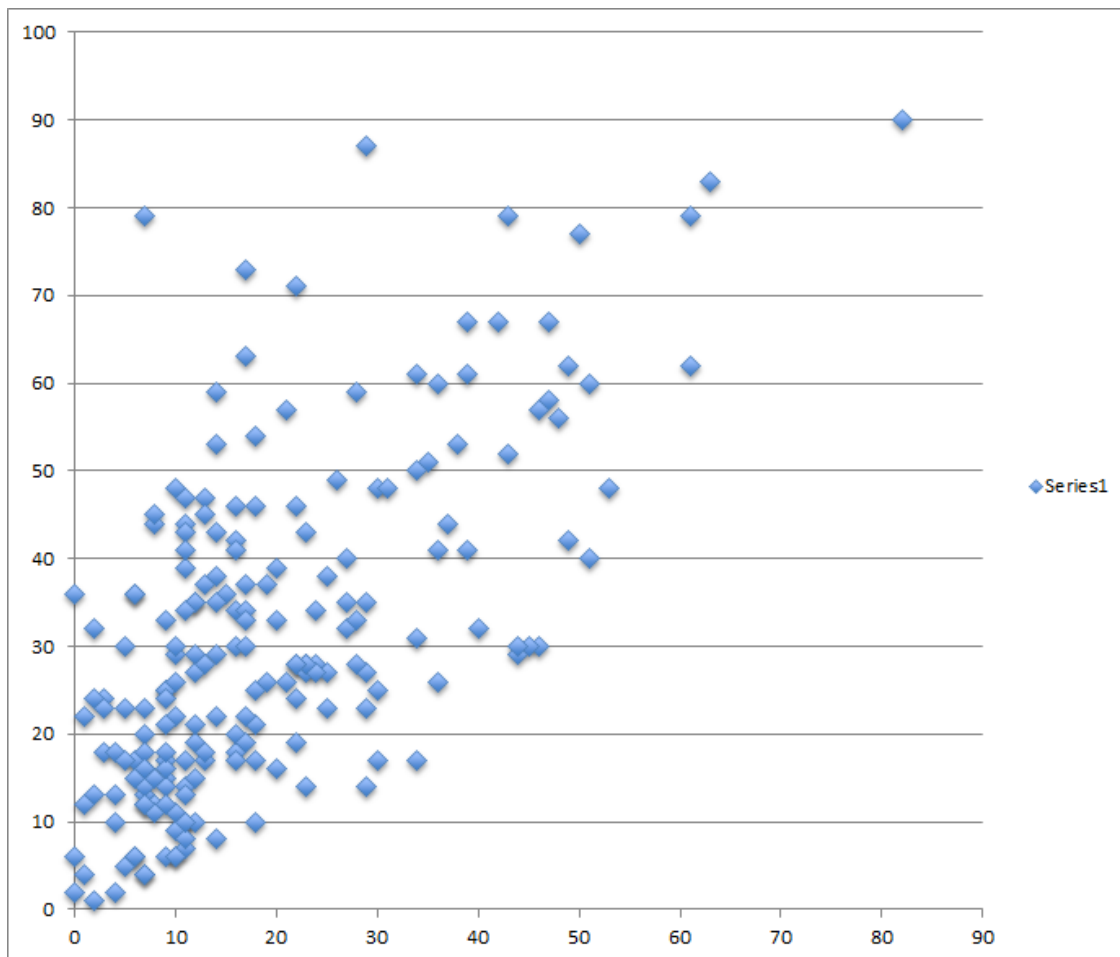
Dependent var: the no. of non-moderator edges (stacked comments)/ comments created in a unit of time.

-normalizing for time: we took the total as the shortest timeline of a community starting from its current date and going backwards

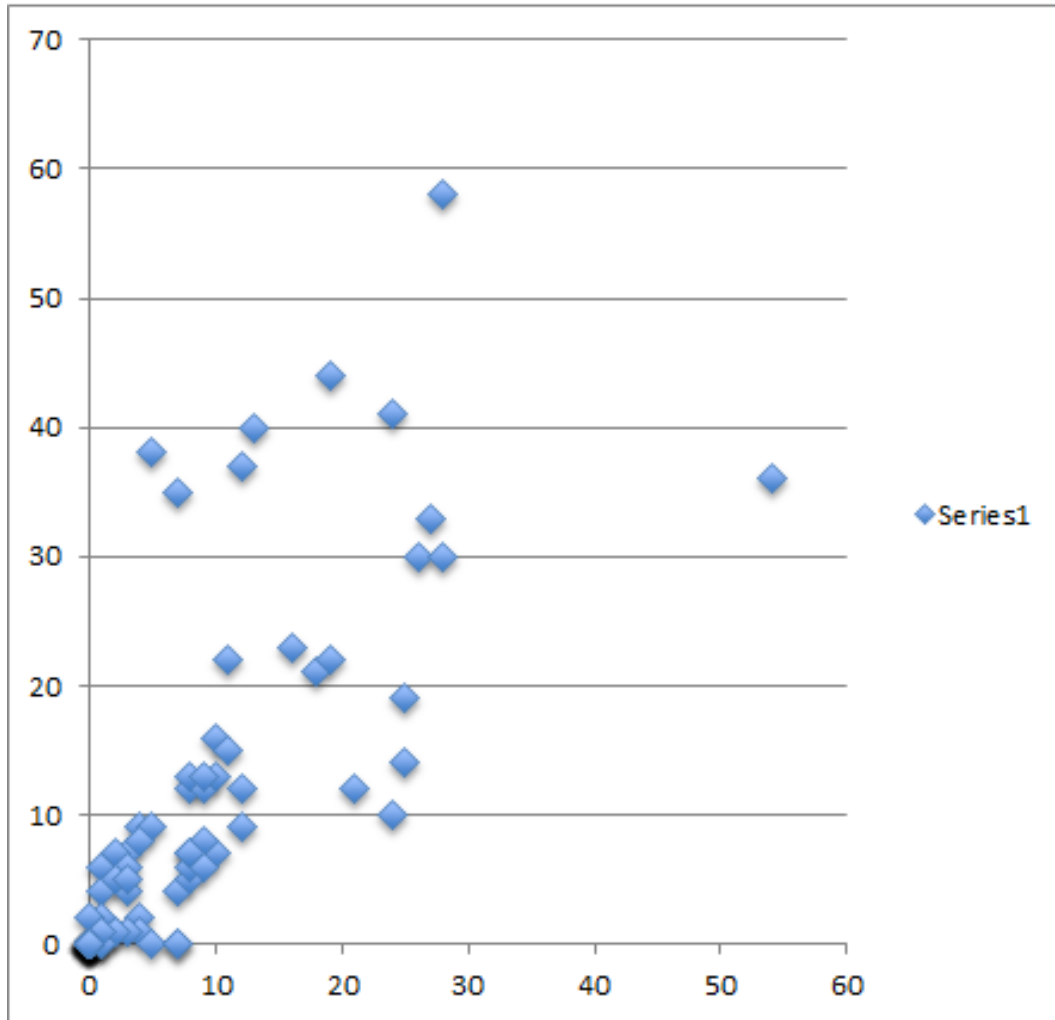
-time unit: a week - should be enough time to give a member time to respond to a comment, but not too long to be redundant

[Spreadsheet with data & timelines for communities](#)

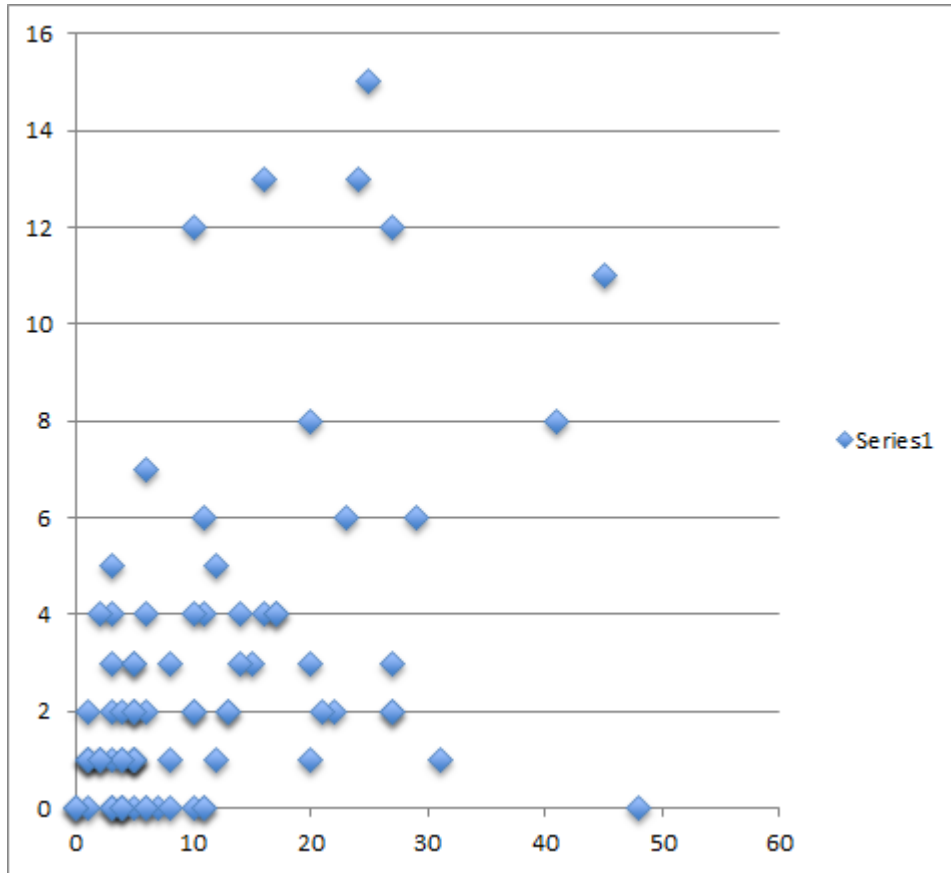
Results:

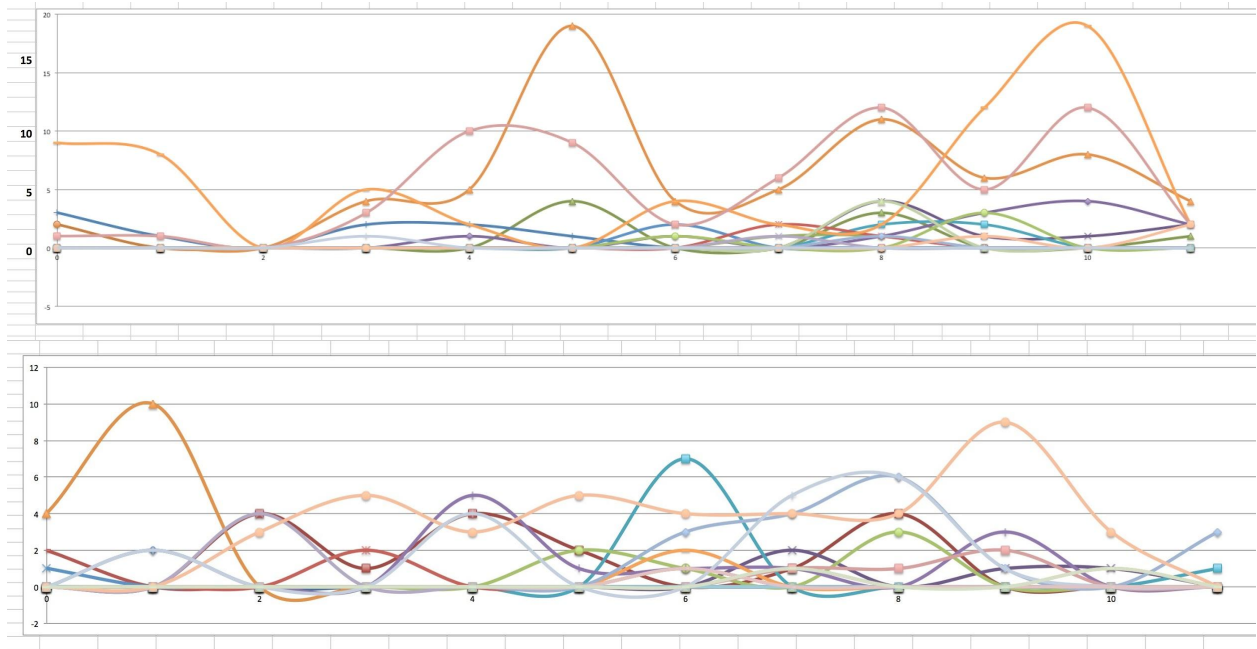


*Edgeryders. Correlation between moderator comments and non-moderator comments using a unit of 3 days. When regression was run, the relationship proved significant at a coefficient level of  $p=.0000$  (?). The strength of the relationship is quite low, meaning that there may be other explanatory factors for “community health” as defined by our dependent variable.*

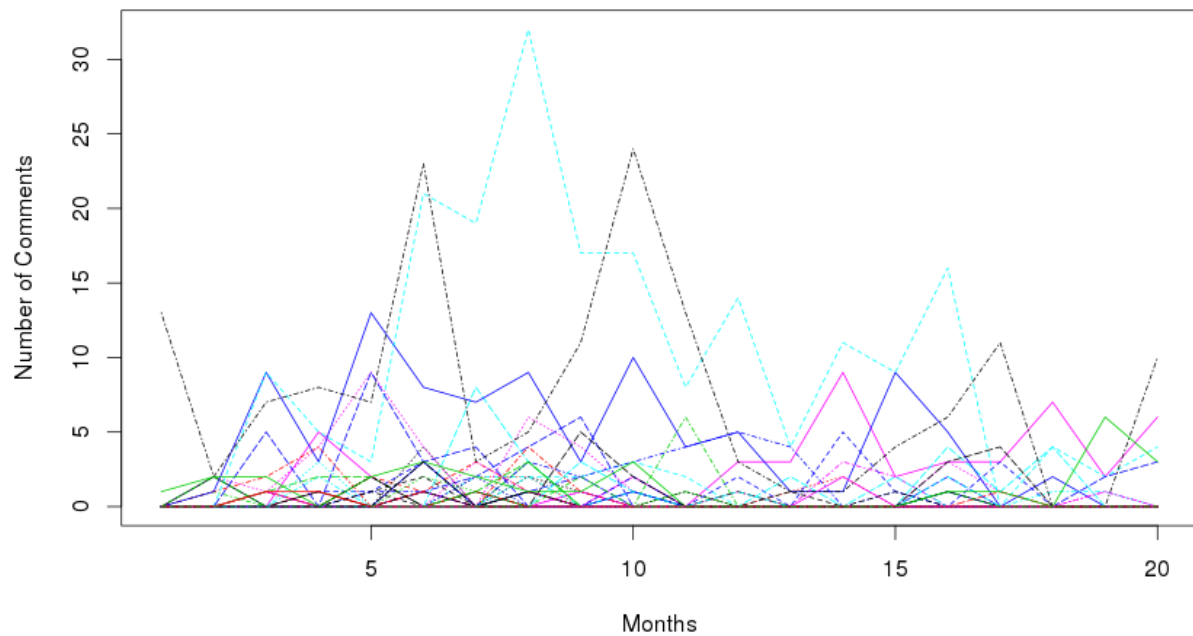


*Matera. Correlation between moderator comments and non-moderator comments using a unit of a week.*





Activity of users registered in Feb-March 2013 up to March 2015 (throughout their first 20 months)



About only 12% of all users posted in the first 2 weeks of activity (right after signing up).

## Curated Notes (Noemi)

Posted here: <https://edgeryders.eu/en/blog/what-i-learned-at-masters-of-networks>

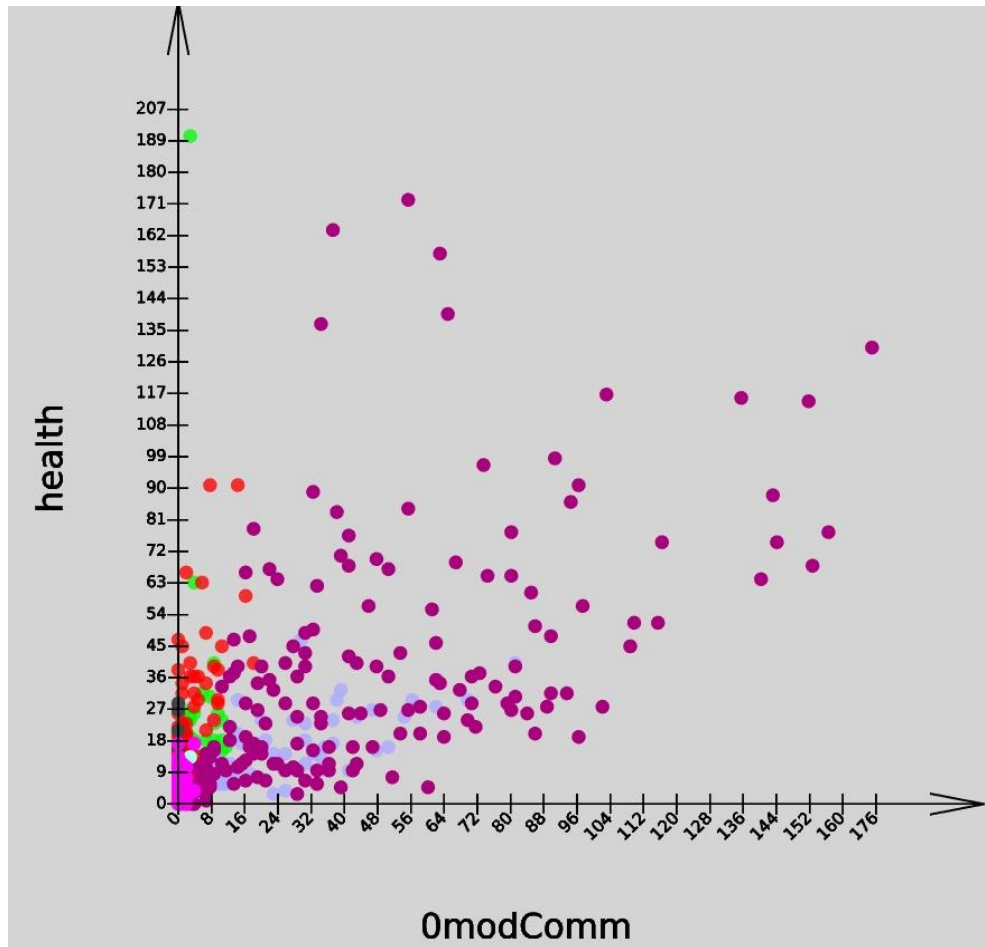
### “Know Thy Community”: data driven community building

The [Masters of Networks](#) event I attended this week in Rome re-asserted (in case you still have doubts!) how empowering it is to have full, collective access to the data you and your community are producing.

Gathering a series of serious collective intelligence researchers, data scientists and community managers, the hackathon focused on **making sense of the different ways in which people engage in online conversations** (mailing lists, forums, social environments like twitter, or other independent platforms). These conversations are scalable and encodable, so for every comment posted, a relationship is created in the database between the sender and the person to which the comment was addressed. When you extract the database and put it in a network analysis software, you are provided with a graph visualization of the network and key metrics hinting at several characteristics, among which the number of unique relationships formed (edges), number of comments (degrees), no. of connections a user would need to make to reach any other user (average distance).

We had over ten comparable datasets to play with, and dove directly into our burning questions, my own being essentially **how to measure the effect (if any) of community managers work on community health?** This is harder than it seems because health could mean such different things: it could be a very high rate of user account creation and content generation, it could be dense conversations between members, or it could be many smaller groups connected by very efficient brokers, so that anyone is at two steps away from each other and resources if needed. For some, it simply means a lively environment where people are busy connecting to each other, so we operationalized it as users' activity (comments) that's not directed at moderators.

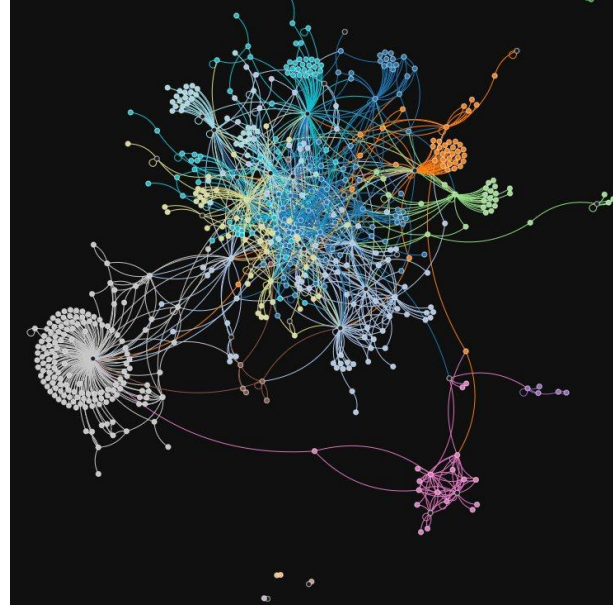
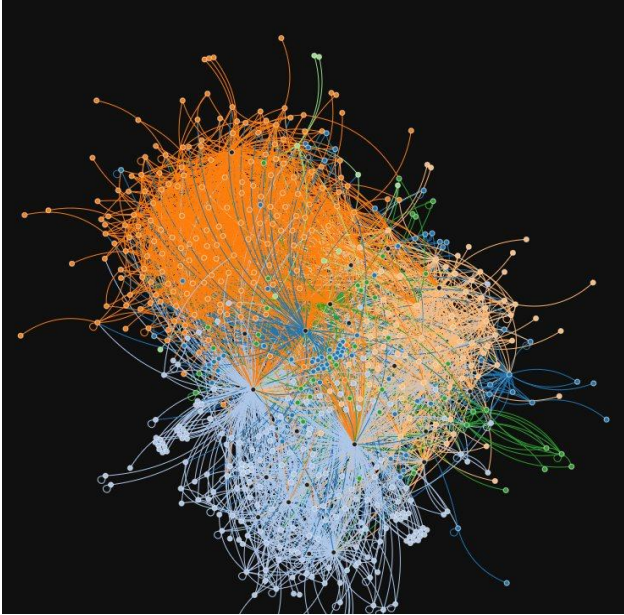
To community builders investing energy nurturing conversation after conversation, a correlation with the amount of community contributions would mean that they're doing something right.



The above is a scatterplot of how this relationship unfolds in Edgeryders (global network of change makers), Matera 2019 (local community of citizens), Innovatori PA (a community of practitioners in the NGO sector) and other mailing list communities. Each dot is the number of comments by community (Y axis) corresponding to a number of comments by moderator (X axis) in a week during the last two years. The violet dots are Edgeryders, and it is also where the relationship was most significant.

Context: These are mostly lasting communities, but their moderation policies are very different, and so is the response: you would see high activity when moderators are little visible, or low, but consistent activity allowing the community to keep decent member ranks. In Edgeryders, a small group go out of their way to make everyone feel welcome and encourage collaboration. Not surprisingly, a dozen moderators account for half of all relationships in the network, which means they know everyone and are part of most conversations. The different shapes of networks offer some clues as to how connections are made. Check out these beautiful visualizations computed with [Edgesense](#), an open web application developed by Wikitalia (dots are people, edges are relationships).





[Innovatori PA](#)



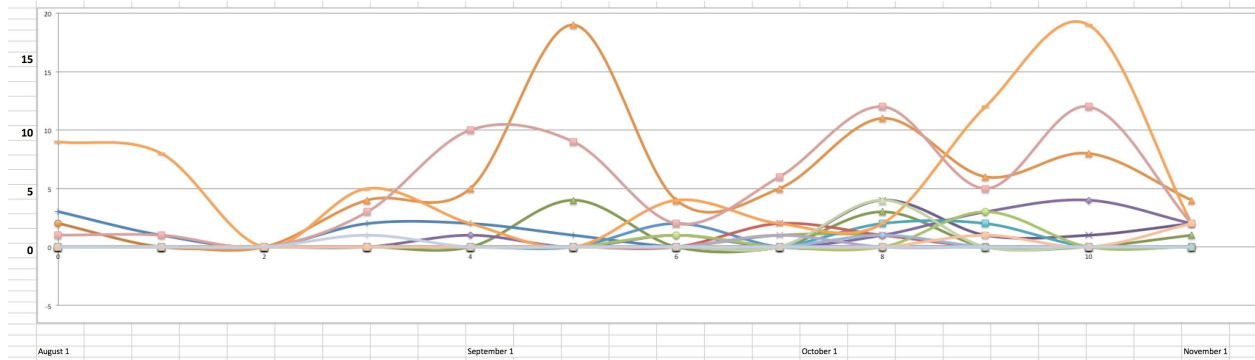
[ImaginationForPeople - correspondants Mailing List.](#)

Seeing this and hacking away more data (God bless the tons of it), we can decide where to go next and what new connections to help create, what groups could benefit of being brought closer together etc.

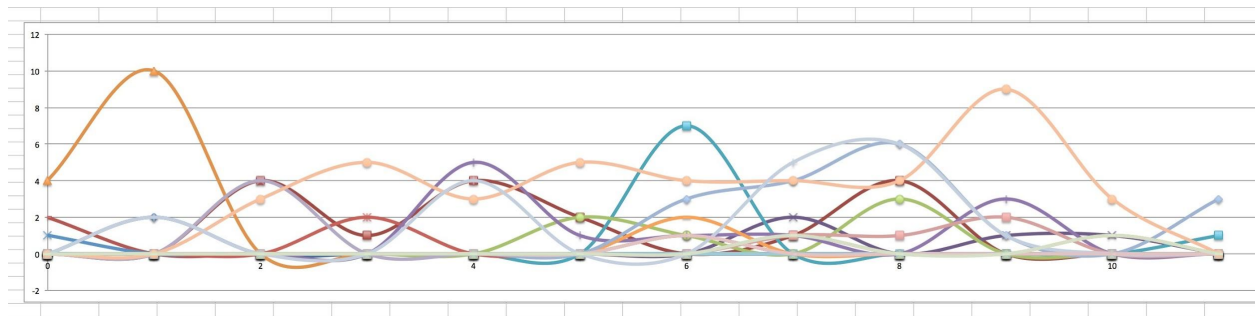
The second question we asked was **“Can we see user life cycle patterns that could inform engagement practices?”** Looking at the Edgeryders database which provided the most substantial data, there was no observable pattern in comments posted by users in their first months following registration. This confirms experience: some people start off as really active contributors and then disappear; others come in, look around, lay low, and then surface when a

topic perceived as meaningful emerges. The relevance of the content is clearly an important variable, and it is what most informs our style of engagement - which is stronger when associated with major events or community projects.

Observing Edgeryders registered in August/Sept 2013 throughout 3 months, and same for the year after, we see *some* common behavior (look at the curves below), mainly because the period is synced to the Living on the Edge events #3 and #4, which see more evenly distributed platform activity.



2013



2014

So much more is to be dug from the data, hacked and learned, and luckily we have in Ben Renoust, Guy Melancon, Luca Mearelli and others here true help (thanks all for providing the visualizations above!). Human behavior being highly complex and unpredictable, we are still a long way of learning what is shaping any person's ability to contribute. What's pretty clear to me, having worked closely with Edgeryders and confirmed more of own intuitions, is that online community management is crucial and pays off when working for social change, especially when forging relationships between global citizens or facilitating conditions for highly capable knowledge workers to work remotely from each other and in their own terms.