

- [About this blog](#)
- [Visit our web site](#)
- [RSS](#)

Big data and development organizations: What happens when you move from theory to practice?

January 31, 2013 by [Milica Begovic Radojevic](#), [Giulio Quaggiotto](#) and [Benjamin Renoust](#)

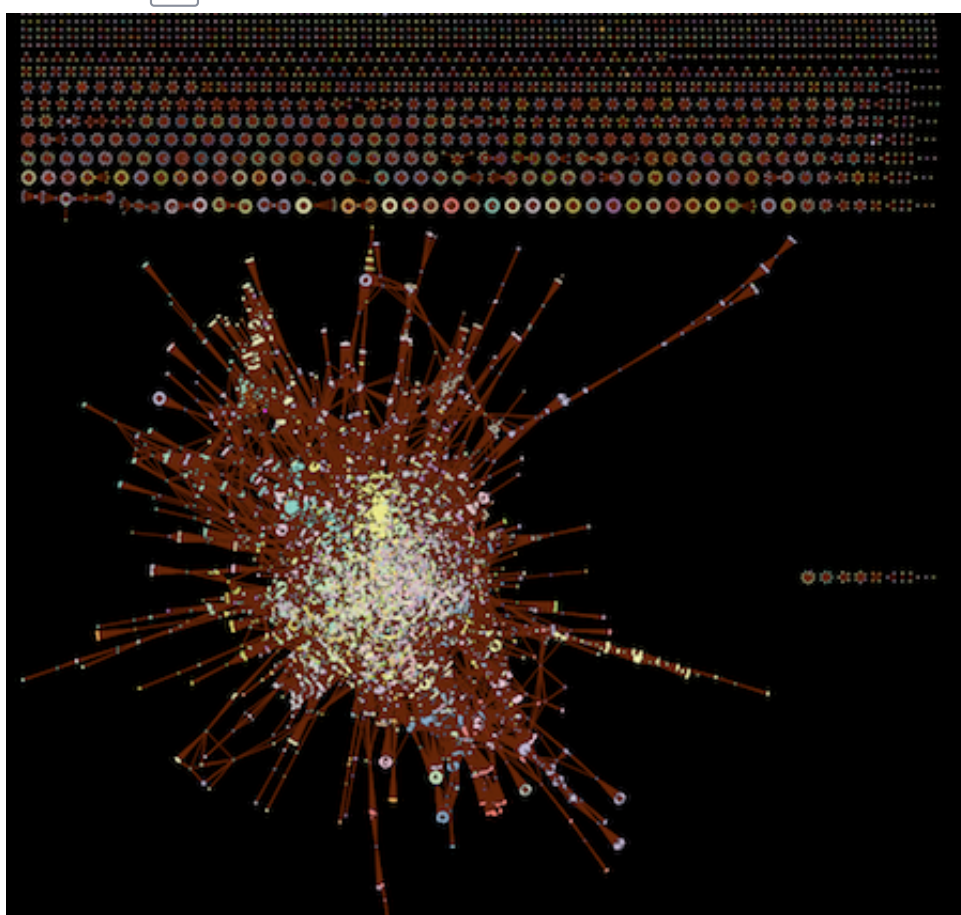
Filed under: [Development](#) [Development 2.0](#) [Guest posts](#)

J'aime

18

155

34



The overall interaction of suppliers of the major World Bank awarded projects through all sectors from 2007 to 2013 (see [data](#)) – each fully connected subnetwork (clique) represents a collaboration of all suppliers on at least one project

It all sounds very good in principle, but what happens when you explore big data for development in practice? It is this “learning by doing” philosophy that inspired our [big data for development challenge](#) with the World Bank and the UN Global Pulse.

Last week’s [Masters of Networks](#) provided the perfect opportunity for us to get hands-on and start delving into datasets.

First challenge: access to data. As Chris Kreutz [noted](#), the world of big data is still, by and

large, a world of closed data.

Luckily, the World Bank's excellent [open financial data site](#) (including the details of World Bank contract awards) came to the rescue. It was great to see the eyes of data geeks shining when they realized the wealth of data sets available (development organizations, please take note: follow the Bank's example!).

Second challenge: What questions? Since we were blessed with having a wealth of social network analysis gurus and economists in the room (including, among others, Guy Melancon, Bruno Pinaud, Marie-Luce Viaud, Raffaele Miniaci, Paolo Gurisatti, and Margherita Russo), we decided to query the data from a social network angle:

- Do certain networks of companies tend to win the majority of World Bank contracts in any given sector?
- What are the linkages of winning companies with other suppliers within a project?
- What is the pattern of knowledge transfer (for example, do projects foster south-south collaboration)?

Where companies gravitate in relation to other suppliers may give us information about whether some win contracts as a function of their location in a network (the more intricate one is, the higher the exposure to new information).

Conversely, that exposure may have negative effects in case of a financial crisis spreading through: since it spreads really fast, the more intricate one company is, the more vulnerable to catching the 'flu' (single dependency on a highly exposed company being even worse).

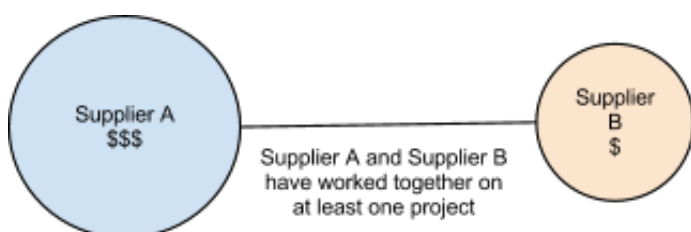
As an important note, these claims need to be refined – possibly using using Burt's Structural Holes theory for instance – with respect to the relative position of an actor in the network. In other words, if you catch the flu and are nested within a community with no other connection to the outside world, you'll recover when everyone does. Conversely, if you sit at the border of a "hole" (in Burt's sense), you'll recover faster – and hopefully help (one of) your community to recover.

But the overall point remains – knowing where some projects or firms are located in relation to others may give us important clues about risk management or factors that may lead some projects to under-perform.

Third challenge: What methodology? In our analysis we used information from one particular dataset ([major contract awards from 2007 to 2013](#)).

We created a network of all suppliers (**nodes** in the charts below) who won contracts in the given time period across all sectors. The **edges** (links connecting nodes) show that two companies are working together under one or more projects.

The brighter the edge, the more projects two related suppliers cooperate on. The size of a node reflects the total funding each supplier received (the larger the node, the larger the amount), and the colour of a node a country that the supplier comes from.



How we construct the network: we're linking two suppliers when they have worked

together on at least one project

So... what did we find?

Before we get there, a quick caveat: none of us have an in-depth knowledge of the World Bank's operation.

Yet, we think it is worthwhile sharing our findings (if we can even call them that) as an example of the challenges of querying and telling a narrative from open data sets from a social network analysis perspective.

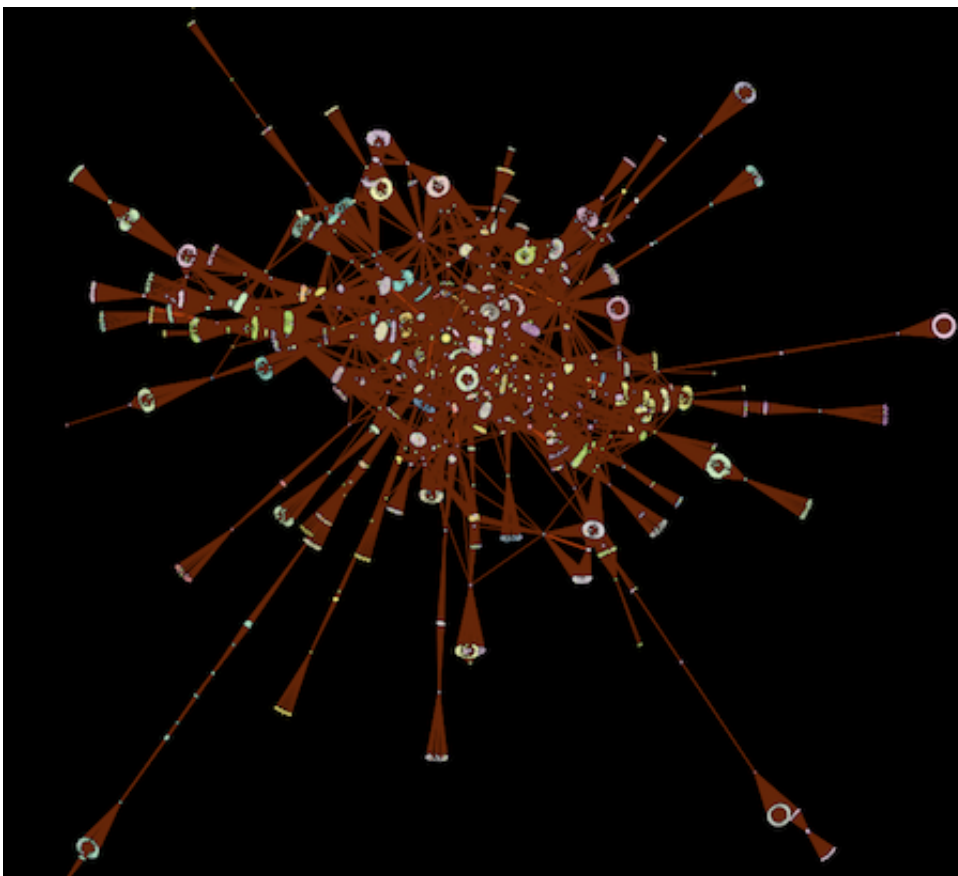
1. First, we wanted to see what the networks of suppliers within different sectors look like.

Do **companies tend to be tightly linked among each other** (many companies working together on many projects) or is there **less connection between individual suppliers**?

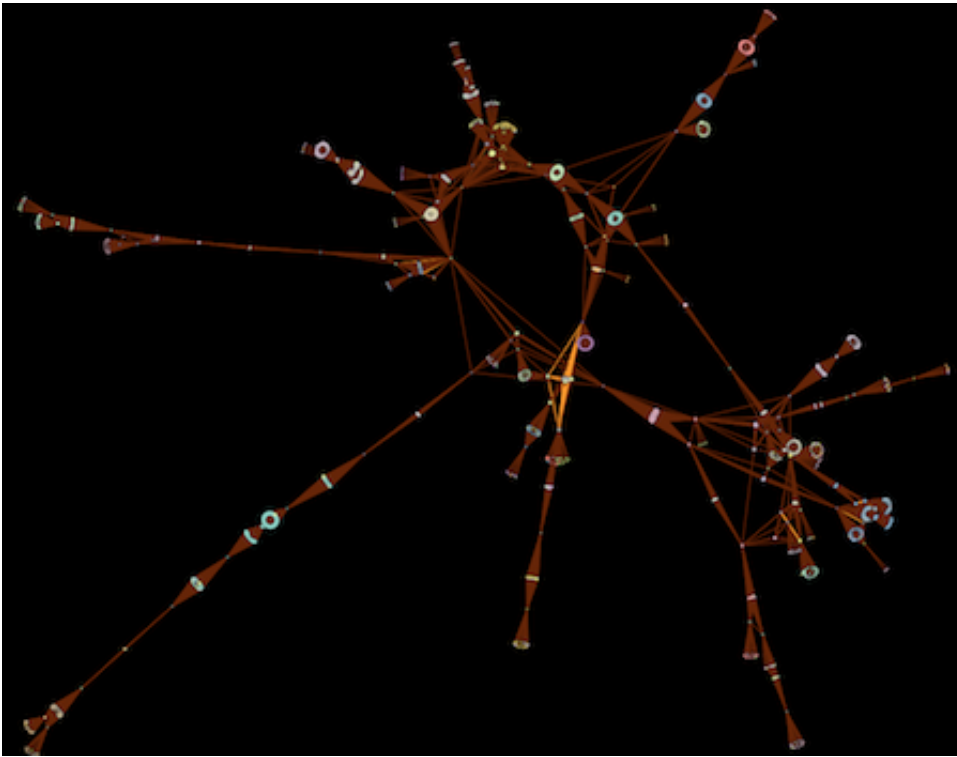
Graph 1 below shows companies working in the health sector, and they form a very dense network compared to education (Graph 2), where connections are more spread and less tangled.

The overall look of a network may allow a comparison between sectors to see whether there are **dominant companies who tend to win majority of contracts versus contracts being evenly spread**.

In this particular case, the topology of a network may indicate that in the education sector, there are less dominant companies or that projects tend to spread contracts more evenly among different suppliers.



*Graph 1 – World Bank supplier network for the **health** sector. It shows a very dense structure of interconnected communities*



*Graph 2 – The **education** sector supplier network presents a much sparser topology*

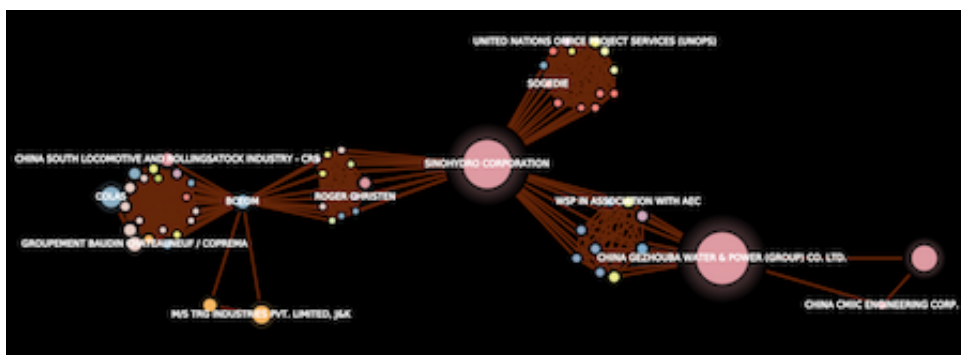
2. We followed up looking at **whether there are companies who tend to win a majority of contracts** in any given sector and whether we can learn anything interesting from the **linkages between them and other suppliers** within any given project.

Graph 3 (below) shows a subset of suppliers who won contracts in 2007 in the transport sector. We can see that two companies are dominating the network (pink nodes), and that these are also densely connected to several sub networks of companies (showing that they supply services across several projects). We also see that companies from certain countries are dominant (pink or orange, and blue in the left part).

One question jumps out as we're looking at this network – **if any type of shock affects the largest companies (pink nodes), how would that impact projects they are supplying services to?**

How would that **impact other companies that collaborate with them?**

What factors lead to a single company winning most contracts relative to others working in this project (blue node on the left, the amounts show a difference between 1 and 2 orders of magnitude with the other suppliers in the group)? Contrast that with a sub-network to the upper part of the network, where there seems to be an even spread of contracts among the companies.



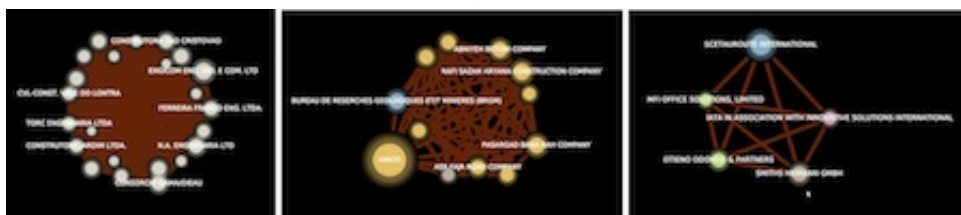
Graph 3 – The transportation sector in 2007, a connected component of suppliers over six different projects. Big differences between node size means some order of magnitude of difference in the overall amount of money raised by a supplier ([>>see larger image](#))

3. We then wondered about **patterns of knowledge transfer**. Do a majority of suppliers win contracts in geographical proximity to their country of origin or not? We found three different cases with a quick analysis (see below, Graph 4).

1. On the far left, we see a network of companies from Brazil providing services to a project in the same country.
2. In the middle Graph, most companies are located in Iran, with few from Europe (France and Germany).
3. In the last case, suppliers from many different countries are involved in the same project.

Given more time, we would find it interesting to analyze the data further to answer questions like:

- To what extent do World Bank contracts facilitate collaboration of suppliers from a close geographic region?
- How does the historical relationship between different countries change throughout time as reflected in their cooperation on implementation of the World Bank contracts?

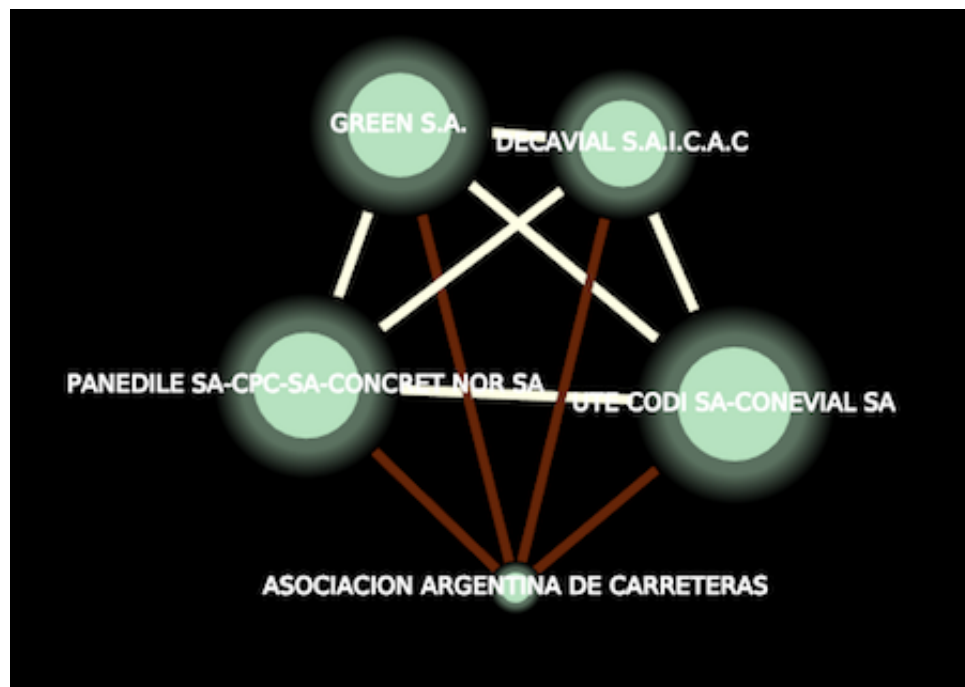


Examples of supplier interactions by country around one project: on the [left](#), all suppliers are from Brazil; in the [middle](#) most of the suppliers are from Iran and two other from Germany and France; on the [right](#) suppliers come diverse unrelated countries. (Links will take you to bigger versions of each image)

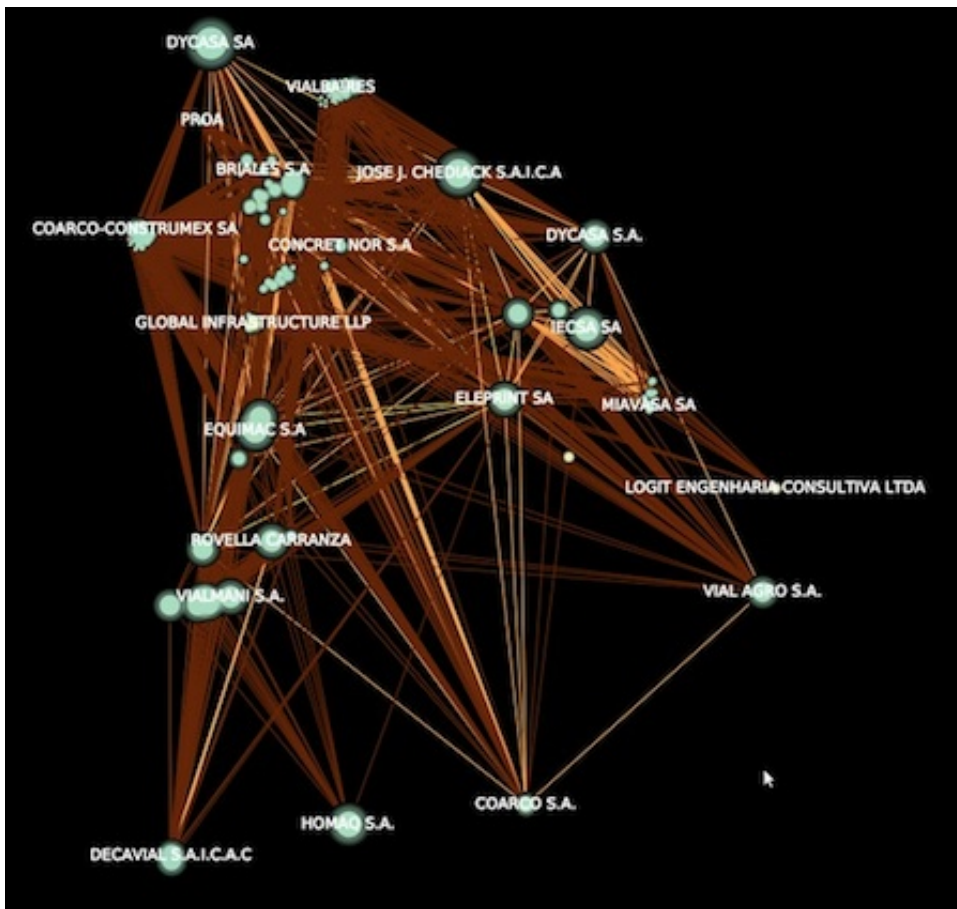
4. Lastly, we wondered if there are **networks of companies that tend to apply jointly for certain projects, what are the patterns of collaboration between companies who cooperate on many projects, and what drives this?**

In the Graph on the left, we found a case of four companies collaborating on two projects (white edges connecting the dots) with an additional company (smallest node connected to other via red edges) joining in subsequently.

When we look at the Graph on the right that represents a larger network of projects and suppliers, we see cases where these companies collaborate on many other projects. We'd need to do more analysis to understand the complexity of these relationships across many projects, and to understand better what drives companies to collaborate in any one sector.



Graph 5 – Interaction between companies: four of those five suppliers are involved in two projects (transports sector in 2007)



Graph 6 – Interaction between companies: extract from a dense area of the overall (years/sectors) supplier network, we can observe here many companies working on many different projects together

We have still quite a lot of work to do...

So, one big take from this exercise is that, if we really want to get down to the granularity that would help improve development operations, we still need to drill down to answer more questions, for example:

- What characterizes an environment in which certain projects tend to under-deliver?
- Are there patterns in terms of what partners are involved in implementation (NGO vs. Gov't vs. private sector); geographic location; development sector; conflict or post conflict setting; size of contracts issued under a project?
- Is there a good match between a project's location and development priorities for that geographic region?
- What is the pattern between contracts awarded to companies from any given country and that country's relationship with the World Bank (e.g. amount of loans and programs it receives)?
- How can we account for the role of intermediary firms (e.g. consulting companies who prepare documentation that a lead supplier submits) who may play an instrumental role in bidding for projects, but are rarely captured in the official process?
- Are there links between companies who tend to win most contracts and their employees (e.g. do they tend to employ ex-development workers, or former government officials)?

And with the new questions, came suggestions on additional datasets that could be useful in providing answers, such as:

- Partners for each project
- Audit reports

- Project and staff evaluations
- Data on all companies/bidders in the contracting process (not just those who win in order to understanding variables that lead to companies winning or losing contracts)
- Cities in addition to countries and regions that project is implemented in
- Roles (and specific knowledge) of companies involved.

Last but not least, from a network science and data mining perspective, a few **questions** emerged for future analysis:

- What model(s) would best answer our questions? (single entity networks such as the supplier example studied here, or multi-level/multiple entities networks)
- How to combine network topology with data attributes/variables so we can cluster the network and then what would represent a cluster?
- What meaning can we attach to traditional network measures (such as Centrality)?
- What other variables may we consider representing in our networks in order to better understand what drives the interaction of specific supplier network?
- Considering the multitude of parameters for each project, would interactive visualization provide a better tool for understanding and analyzing patterns related to companies' behavior?

Considering that this analysis looked at the level of supplier network, we may want to transpose Ronald Burt's [theory on structural holes in organizations](#) that considers topology of a company's internal network and actors connecting communities to better understand the company's behavior and the roles of its actors.

At the end, Masters of Networks proved very valuable for the World Bank, UNDP, and the UN Global Pulse initiative on using big data to improve operational effectiveness of development organizations.

We are at the beginning of understanding better the potential buried in the heaps of operational data of development organizations.

Still, there is a lot we have to do but in having more iterations in refining the questions we're asking and more analysis to ensure that we add a layer of context to the data.

We will take that next dive in [Vienna on February 23rd with the Open Knowledge Foundation](#), that will bring us one step closer to getting more out of big data for development.

P.S. Special thanks to all colleagues who participated in the work of this group, and who provided insight and advice that fed into this write-up.

Tags: [Benjamin Renoust](#) [big data](#) [Giulio Quaggiotto](#) [Milica Begovic](#) [Radojevic](#) [open data](#) [social network analysis](#) [World Bank](#)

10 comments



Leave a message...

Best ▾ Community

Share



[prasannalaldas](#) · 9 months ago

Very, very intriguing work (even for somebody with a limited understanding of SNA!), and

a really promising start to the 'big data for development challenge'. Thanks so much for the shout-out to World Bank Finances - this is just the sort of thing you hope people would do with data after you put it out there. Happily some of the additional datasets you ask for are already out there and some others may be on the way (thanks for the 'demand!'). It will also be interesting to hear from procurement specialists on the ground about how these preliminary findings may or may not be useful to them.

1 ^ | v Reply Share ›



Cari · 8 months ago

This is a beautifully articulated article. Thanks for taking the time to walk us through your process in language that anyone can understand. I am not someone who gravitates to delving into data and I know very little about SNA, but, like Prasannalaldas, I am intrigued. What I appreciate now after reading your post is what a layered process this "getting down to the granularity" is, i.e., questions, data, more questions, more data. Truly is a dive. Thanks for working out loud on this.

^ | v Reply Share ›



Guest · 8 months ago

Alexander, you make a very good point that the links between companies provide only a vague association that could mean many different things, from companies working together to competitors just being "active" in the market, etc. The World Bank's dataset provides supplier information only on the main supplier, not subcontractors or sub-subcontractors, so the limitations are substantial at this point. But the potential for what this type of visualization can have is apparent. As you imply, the important aspect of these visual "experiments" is that they provide crucial insights for the World Bank about how data could be used to build all sorts of models. It also informs the institution of the level of granularity needed for specific types of data, so that a true narrative can be articulated about what it is that these models may be showing us.



^ | v Reply Share ›



Alexander Korolyov · 9 months ago

A truly fascinating visualization! It's very beautiful.

I think there might be an issue related to what links between companies might actually mean.

Based on this dataset, two companies working on the same project might indicate one of

the following different things:

- A project is quite large and covers many different activities. For example, one of the projects you covered (Kenian Northern Corridor Transport Improvement Project) awarded contracts over the period of 5 years in areas like licensing of drivers, as well as in road construction. Companies that won these two contracts most likely did not collaborate, but there can be some degree of collaboration in some other cases.
- The companies are competitors rather than collaborators; they won different lots of the same tender (happens all the time in e.g. road construction).
- The companies formed a consortium. Although I am not sure this most interesting case is reflected in the dataset - does it include information on subcontractors?

^ | v Reply Share ›



Benjamin Renoust · 9 months ago

Just a precision for those who wonder: we've generated the visualizations with Tulip (tulip.labri.fr)

^ | v Reply Share ›



Steve Ardire · 9 months ago

Terrific SNA/ONA use case and like your candid disclosure of still quite a lot of work to do with very good questions for future analysis

^ | v Reply Share ›

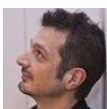


Millie Begovic R. → Steve Ardire · 9 months ago

Steve, thanks! it really is one of the first steps where this type of work will hopefully show the richness that we have in data coming out of projects implemented globally. More data we have, the more value we can get out of it. In this particular case, there are already few data-sets identified as 'would really like to have' available, so the next step may enrich the data.

Millie

^ | v Reply Share ›



Alberto Cottica · 9 months ago

Delighted to read this! I knew you guys would hit it off (well, technically I attached a high probability to the event that Masters of Networks would kickstart conversations like this).

My two cents: I think there is still a certain amount of grit in the analysis. This comes from the fact that the network analyst group has a mathematical focus and looks at networks as interesting mathematic objects with certain properties; Millie and Giulio are driven by burning questions about very real-life policy issues that they are trying to address by looking at those data. More specifically, it seems to me that:

1. some of the questions ending the post could be addressed simply by staying close to

the definition of edge employed to build this particular network. This, for example, is the case of the meaning of traditional network measures: consider closely the nature of your edge and you will have the answer to that.

2. some of the questions halfway through the post simply don't look like network questions (yet?) to me. "What characterizes an environment in which certain projects tend to under-deliver?" you ask. Under-delivery is a function of many things, and there is no particular reason to think that, a priori, you will be able to extract the signal of network configuration from the loud noise of productivity in that particular country or region. Networks are good at mapping phenomena that are directly traceable to relationships. You should be able to detect gatekeeping, or dominant positions, stuff like that: but nobody ever said that you could use them to answer any question!

^ | v Reply Share ›



Benjamin Renoust → Alberto Cottica · 9 months ago

Thanks a lot Alberto! See how productive was MoN :)

Here we've just scratched the surface, I think the grit you feel there is might more come from the very limited amount of time we had to collaborate with Millie. She has been amazing in understanding what some specific network properties could bring to the overall reasoning, translating what I could see in more understandable language and she avoided the trap of believing that networks are magic wands which can answer every questions. I may have observed some rigor in the methodology but we should absolutely not extrapolate too much jumping into conclusions too fast which is easier when you get such representation.

You're right on the first point you raised, about the way we should interpret traditional network metrics, but this question comes along with the choice of model. If I use the betweenness centrality (BC) for instance (sorry guys for the jargon), I might extract the actors connecting communities, but Millie only can answer if it is correct. After all the BC just gives an indication on where the shortest paths go through. If I use the first model presented by Guy at MoN (fully bipartite), the BC would have put forward the projects only. But using the bipartite we have different tools that can observe the nature of the interactions between suppliers for example...

So many interesting things to do, we're just scratching the surface here...

^ | v Reply Share ›



Millie Begovic R. → Alberto Cottica · 9 months ago

Alberto :)

Yes, this is a perfect example of how minimal rules lead to very interesting emergent phenomena- linking us up with this fantastic group of network and data gurus has been a fantastic experience to date. Just one comment on your second point.

When it comes to studying environment in which some projects tend to deliver, I agree with you in part that this still is a very exploratory phase. We would like to understand if there are any patterns that characterize those projects and that may

19/10/13 Big data and development organizations: What happens when you move from theory to practice? | Voices from Eurasia - We help build better lives. understand if there are any patterns that characterize these projects and that may have a network twist to them, e.g. is there is particular network of partners that supplies services to them? But this is the one question we didnt even get to properly and that we hope to address in the next iteration, so lets see what comes out of it.

^ | v Reply Share ›

ALSO ON VOICES FROM EURASIA

WHAT'S THIS?

The winner (could) take it all... – MIT ClimateCoLab crowdsourcing ...

2 comments • a month ago



Millie — Leisa, thanks so much for the congratulations and the nice comment! :) We are super excited about Boston, ...

Are quotas the solution for bringing more women into politics?

2 comments • a month ago



Barbora Galvankova — ear Mrs Popovska, thank you for taking time for reviewing the article and informing me about the ...

We want to spread the word about universal design in Ukraine: Are you ...

8 comments • 3 months ago



Kristina Bagramian — Totally agree! We should create environment that is as accessible as possible. Independence ...

Academia and agriculture: Engaging students for a better future

2 comments • a month ago



Damir Kurtagic — Hi Justyna, greetings to you too! It was truly greatto hear that you embarked on this Permaculture ...

Subscribe Add Disqus to your site

Who's blogging?

[Milica Begovic Radojevic](#)

Milica (a.k.a. Millie) is working on [social innovation for development](#) in [UNDP Europe and Central Asia](#), exploring micronarratives, data for development, including for conflict prevention, and many other extraordinary topics. [@ElaMi5](#)
[More posts »](#)

[Giulio Quaggiotto](#)

Giulio is practice leader for UNDP's Knowledge and Innovation ~~needs~~ team in Europe and Central Asia, and a fervent advocate of [Development 2.0](#). [@gquaggiotto](#)
[More posts »](#)

[Benjamin Renoust](#)

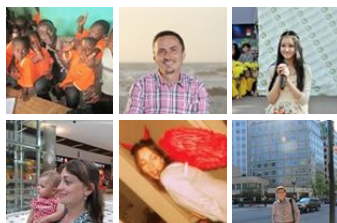
Benjamin is a Ph.D. candidate at the University of Bordeaux ([LABRI](#), [CNRS](#) & [Inria](#)) and the French National Institute for Audiovisual ([Ina](#)). Benjamin has special interests in social network analytics with emphasis on visual analytics of networks, ranging from classical social networks to news document networks.

[More posts »](#)

Stay connected with us



5 136 personnes aiment UNDP in Europe and Central Asia.



Module social Facebook

Follow @undpeurasia

Categories

- [Anticorruption](#)
- [Central Asia](#)
- [Climate change](#)
- [Development](#)
- [Development 2.0](#)
- [Disaster response](#)
- [Environment](#)
- [Gender equality](#)
- [Governance](#)
- [Guest posts](#)
- [Health](#)
- [HIV](#)
- [Human rights](#)
- [Migration and remittances](#)
- [Peace and security](#)
- [Poverty](#)
- [Roma](#)
- [Social inclusion](#)
- [Social innovation](#)

Recent Comments

- Alex Oprunenco on [The development story behind the MDGs: Seeing the forest beyond the trees](#)
- Azizkhon Bakhadirov on [Paternity leave: A little *chilla* with my brand new daughter](#)
- Olena on [Smart citizens versus smart cities: Bridging the gap in Odessa](#)
- George Hodge on [The development story behind the MDGs: Seeing the forest beyond the trees](#)
- Karin Plokker on [Paternity leave: A little *chilla* with my brand new daughter](#)

Blogroll

- [Caucasus Research Resource Centres Armenia](#)
- [Center for Eastern Studies](#)
- [Cyprus Community Media Centre](#)
- [East of Center](#)
- [Eastern approaches](#)
- [EBRD on Topic](#)
- [From poverty to power](#)
- [Genderstan](#)
- [OECD insights](#)
- [Praktikum in Muminabad, Tajikistan](#)
- [Social Science in the Caucasus](#)
- [Social transitions](#)
- [The Armenian Economist](#)

- [The Peace Exchange](#)
- [Экономическое обозрение](#)



[Latest tweets](#)

- Have u already seen the [#MDG](#) report from [#Serbia](#)? U can find it on our [@Scribd](#) profile together w/ other publications <http://t.co/37XuwGTcJA> →, 1 hour ago
- What's programme based budgeting? Is it anyhow beneficial for citizens? Check the case from [#Moldova](#) <http://t.co/9YpuHy3viY> →, 2 hours ago
- RT [@UNDPPhr](#): [@UNDP Eurasia](#) The 100% [#renewable](#) [#energy](#) on islands conference kicked off by a great presentation of El Hierro island →, 2 hours ago
- How best to challenge stereotypes of [#domesticviolence](#)? Can a talk show help? <http://t.co/ISiTvNGXow> [#VAW](#) cc [@EndDV](#) [@NOMOREorg](#) →, 3 hours ago
- RT [@UNDP SEESAC](#): Transparency of [#ArmsExports](#) in [#Balkans](#): FYR [#Macedonia](#) publishes its Annual Report on [#Arms](#) Exports <http://t.co/mfvIBQnfiw> [@controlarms](#) →, 4 hours ago