A deepen for Questions and Answers closing the seminar

Paths and flows for centrality measures in networks

Daniela Bubboloni

11th March 2022

D^2 Seminars - Florence Center for Data Science

Q. 1. *Is there a package for implementing those computations?*
How much does it cost the computation of those centrality measures? Do you have a log-linear algorithm or do you need an approximation algorithm?

The computational problem was not at the core of our research because our main purpose was to make clear some definitions missing in the literature and prove that when a single vertex is considered two different concepts overlaps while this does not happen when more vertices are considered. On the other hand, now that two new group centrality measures Λ and Φ are available, it is certainly important to set the problem of their computations.

In UCINET something is available because there, for a vertex x, we have the computation of

$$\Lambda_2^N(x) = \sum_{\substack{y,z\in V\setminus\{x\}\\ y\neq z,\lambda_{yz}^N>0}} \frac{\lambda_{yz}^N(x)}{\lambda_{yz}^N}$$

By the result that we have established in the paper that measure is equal to

$$\Phi_2^N(x) = \sum_{\substack{y,z\in V\setminus\{x\}\\ y\neq z,\varphi_{yz}^N>0}} \frac{\varphi_{yz}^N(x)}{\varphi_{yz}^N}$$

and indeed in UCINET it is said that for computing $\Lambda_2^N(x)$ they compute $\Phi_2^N(x)$. Things become more complicate dealing with groups X of vertices. Borgatti and Everett propose to consider

$$\Phi_2^N(X) = \sum_{\substack{y,z\in V\setminus X\\ y\neq z,\varphi_{yz}^N>0}} \frac{\varphi_{yz}^N(X)}{\varphi_{yz}^N}.$$

Its computation should be possible by UCINET since it involves only differences of flows. Now the only difference between the value on X of our full flow vitality measure

$$\Phi^N(X) := \sum_{\substack{y,z\in V\\ y\neq z,\varphi_{yz}^N>0}} \frac{\varphi_{yz}^N(X)}{\varphi_{yz}^N}$$

and $\Phi_2^N(X)$ is the index set on which the sum is done. Of course this cannot change the computation in a significative way.

On the other hand, by our results, for X of size at least 2, dealing with paths or flows matters. More precisely, we cannot compute

$$\Lambda^N(X) := \sum_{\substack{y,z\in V\\ y\neq z,\lambda_{yz}^N>0}} \frac{\lambda_{yz}^N(X)}{\lambda_{yz}^N}$$

through the "easy" $\Phi^N(X)$. In principle, in order to compute $\Lambda^N(X)$ you need, for every $y \neq z$ in V, to find all the possible flows from y to z which realize a maximum flow, extract from each of them the associated sequence $\gamma = (\gamma_1, \dots, \gamma_{\varphi_{yz}^N})$ of arc-disjoint paths (in our paper we explain how to get that using the Flow Decomposition Theorem), annotate the number of indices j such that γ_j passes through some vertex of X and finally take the minimum of those numbers. This seems a very hard task. In fact, in Trimponias et al. (2017) it is proved that given a network $N = (V, A, c)$ such that $c(A) \subseteq \{0, 1\}$ and three distinct vertices $x, y, z \in V$, it is NP hard proving whether there is a path from y to z passing through x.

We are not expert on those computational aspects and surely possible contributions in this direction are welcome for some joint research.

Carlo De Bernardi, within our project on Trade Networks, is now working on some specific package for our definitions, but there we are mainly focused on Φ.

Q. 2. *How is the world trade network constructed and what is the difference between capacity and flows in that network?*

A Trade Network $N_t = (V, A, c)$ can be constructed looking, for a certain year t, to the trading activity of a set V of countries, not necessarily all the world countries. For the project with De Bernardi, Gori and Ricchiuti we are thinking, for instance, to the EU countries[1]. In our construction we have an arc $a = (i, j)$ from every country i in V to every other country j in V and we consider $Exports_{i,j}$ defined as the amount of exports of i towards j, during the year t. We assigns the capacity of (i, j) as

$$c(i, j) := Exports_{i,j}.$$

Having the network N_t in hands, there are many possible group centrality measures that can be used in order to detect the groups of countries that play the main role. Among them we bet on our group centrality measure Φ as a very suitable one.

The mathematical concept of flow seems to be very expressive in this context. However, in order to answer to your second question, observe that when we use the word flow we are dealing with a mathematical object which is a function $f : A \to \mathbb{N} \cup \{0\}$ (see the seminar for its formal complete definition) which does not overlap with the economic concept expressed by the same word. In particular when f is a flow, we have $f(a) \leq c(a)$ so that the value of the flow f on the arc a is not necessarily equal to the capacity of a. Note also that the number φ^N_{yz} appearing in our centrality measures stands for the maximum of the numbers $v(f) = \sum_{a \in A^+_y} f(a) - \sum_{a \in A^-_y} f(a)$, for f varying among flows from y to z.

Q. 3. *How one decide which measure is better?*

It is very interesting to decide which measure better fits a certain situation. The word "central" has various different meanings such as power, prestige, authority, best betweenness position and many other are conceivable. In the literaure there does not exist a satisfactory axiomatization of centrality and, presumably, an axiomatization capable to reach every nuance of the meaning of the word "central" is just impossible and insensible[2].

Moreover, it is unfortunately missing a mathematical analysis of the properties that make each centrality measure different from another one. The bad consequence of that is the frequent use of chaotic lists of centrality measures, chosen with the main motivation of their calculability and an a-posteriori use of the information reached by them. We think that there are two reasonable complementary ideas for affording these critical aspects.

The first idea one is looking for properties, isolating some relevant characteristics that a centrality measure could reasonably have, and then use them to discriminate one centrality measure from another one. For instance, we think that monotonicity could be considered as one of those relevant properties. It appears certainly unpleasant that a centrality level of a set X of vertices could be larger than the centrality level of some set Y of vertices containing X. The formal approach we used seems very promising for dealing with the properties of centrality measures, in particular, with those properties invoked by Sabidussi since 1996 as the main desirable. A great advantage for our two centrality measures comes from the fact that flows have a long consolidated mathematical history that could pave the road for a wide and deep investigation of many interesting properties for Φ and Λ.

The second idea is the application of a centrality measure μ one is interested in on some special networks, both concrete and artificial in principle, in order to concretely discover in which types of networks μ behaves better than other measures. The nature of μ could be, for instance, be recognized by discovering the shape of the networks on which it "fails" its diversification, in the sense that $\mu(x)$ gives the same value for all the vertices x. Within this spirit we are planning the implementation of Φ (and Λ if computation allows) on Trade Networks trying to understand if Φ behaves better than other classic centrality measures.

Q. 4. *Are there some connections between flows and graphical models?*

I do not see a clear connection between our research and graphical models. As a first issue, recall that for working with flows we need capacitated directed networks. Graphical models seems to be usually expressed by undirected graphs. One can try to associate with an undirected graph the capacitated directed network with two arcs in the two possible directions for every edge in the graph and think the capacities to be 1 over those arcs and 0 elsewhere. But

[1] In Clò et al. 2021 are considered the years t belonging to the period 2005-2012 and the World Trade Network which is construcetd as an undirected network.

[2] Somebody has recently tried to give an axiomatization of certain types of centrality measures such as the vitality measures or the distance based centrality measures (see the papers by Oskar Skibski). That is surely an interesting idea.

the point is, what does a flow from a vertex y to a vertex z represent for the original graphical model? Is it interesting and meaningful? There are also cyclic directed graphical models and for them there is no need to create a network, but the problem of the interpretations of flows and of the corresponding centrality measures remains. However flows are so natural within a network that I think that it could be worthy trying to use them in that context.

References

S. Clò, E. Marvasi, G. Ricchiuti *The internalization of state-owned enterprises: an analysis of cross-border M&As*, DISEI, Working paper No. 6, 2021.

G. Sabidussi, *The centrality index of a graph*, Psychometrika **31** (1966), 581-603.

G. Trimponias, Y. Xiao, H. Xu, X. Wu and Y. Geng, *Node-constrained traffic engineering: Theory and applications*, IEEE/ACM Transactions on Networking **27** (2017), 1344-1358.

D. Bubboloni, M. Gori, *Paths and flows for centrality measures in networks*, Networks (2022), https://doi.org/10.1002/net.22088.