All for One or One for All?
Analyzing Collaboration Patterns in Research Environments

Mario CATALDI
Université Paris VIII
Studying influence among researchers

The evaluation of the work of a researcher and its impact on the research community has been deeply studied in literature through the definition of several measures (*h-index* and its variations).

Problem: they usually assume the **co-authorship to be a proportional collaboration** between the parts, missing out their relationships and the relative scientific influences.

*Great CV, but obtained by co-authoring papers with a great professor.*
Studying scientific relationships

Authors represent entities (nodes) in a graph and co-authored papers define, in some way a sort of relationship existing among the nodes.
1 – Productivity measure

In literature there is plenty of methods for evaluating the output of a researcher. Most of them consider their publication records as the basis for the scientific evaluation.

Paper authored (or co-authored) by a1

\[ O_{a_i} = \{o_{a_i,1}, o_{a_i,2}, \ldots, o_{a_i,n}\}, \]

Productivity as number of papers

\[ p_{a_i} = |O_{a_i}|, \]

Paper authored (or co-authored) by a1 and a2

\[ O_{a_i,a_j} = O_{a_i} \cap O_{a_j} = \{o_{(a_i,a_j),1}, o_{(a_i,a_j),2}, \ldots, o_{(a_i,a_j),m}\}, \]

Productivity as number of papers in common

\[ p_{a_i,a_j} = |O_{a_i,a_j}|. \]
2 – Estimating dependences

The weight of each edge estimates how much the collaboration between a pair of authors is dependent on the collaboration with a common co-author.

\[ w_{a_i,a_k} = 1 - \frac{p_{a_i,a_j,a_k} + p_{a_i,a_k,-a_j}}{p_{a_i,a_j,a_k} + p_{a_i,a_k,-a_j}} \]

\( p \) represents the productivity, based on same high level criteria, of multiple authors.
2 - Estimating dependences

At this point it is now possible to quantify how much the productivity of $A_1$ is dependent on the collaboration with $A_2$ by evaluating all the scientific collaborations of $A_1$ and calculating how much they have been dependent on the scientific contributions of $A_2$.

\[ d_{a_1}^{a_j} = \frac{p_{a_1,a_j}}{p_{a_1}} \times \frac{\sum_{a_k \in \text{Net}_{a_1,a_j}} w_{a_1,a_k}}{|\text{Net}_{a_1,a_j}|}, \]
3 - Dependence Curves

These d-index values can now be leveraged to graphically map the scientific dependences of an author, along his/her whole career, as a curve that plots the d-index values related to all her co-authors.
4 - Dependence Trajectory

Given the complete set of dependence values, for each year and relative to each co-author, we calculate the researcher’s dependence trajectory, by calculating the standard deviation, along the time, of each dependence value, for each co-author, from the optimal attended value of 0.
The web application

~2M scientific authors
~4M papers
Case study
Case study

Faloutsos vs Leskovec vs Philip S. Yu (most prolific author..almost 700 papers)

d-index is not strictly correlated to the number of papers (even if the productivity measure is based on it)
Case study

Number of papers and number of co-authors are not strictly correlated. In fact, a higher number of research outcomes (or co-authors) does not automatically imply a lower dependence coefficient.
Case study

Longer career -> higher number of papers -> lower dep. coeff?

Length of career does not imply a higher number of papers. Why?
Thank you!

For any info:
http://d-index.di.unito.it/