DIGITAL SYSTEMS FOR HUMANS
GRADUATE SCHOOL AND RESEARCH

UNIVERSITÉ CÔTE D'AZUR

SNIF
Scientific Networks and IDEX Funding

PI: Patrick Musso (Gredeg)

speakers: Małgorzata Sulkowska (EUR DS4H/Inria,I3S)
Alberto Corsini (EUR DS4H/Gredeg)

Gredeg Groupe de recherche en Droit, Economie et Gestion

i3S sophia antipolis

informatiques mathématiques Inria

skema BUSINESS SCHOOL

Scientific Networks and IDEX Funding

- Time frame: 2018-2021 (3 years + 1 year)
- Partners: Gredeg, I3S, Inria, Skema

- Goal: estimate the impact of public funding on researchers' *collaboration* and *productivity*

Scientific Networks and IDEX Funding

- Time frame: 2018-2021 (3 years + 1 year)
- Partners: Gredeg, I3S, Inria, Skema

- Goal: estimate the impact of public funding on researchers' *collaboration* and *productivity*

- Methodology:
  - I3S and Inria
    - ★ complex network theory approach
  - Gredeg and Skema
    - ★ econometric approach
    - ★ qualitative approach

Scientific Networks and IDEX Funding

- Time frame: 2018-2021 (3 years + 1 year)
- Partners: Gredeg, I3S, Inria, Skema

- Goal: estimate the impact of public funding on researchers' *collaboration* and *productivity*

- Methodology:
  - I3S and Inria
    - ★ complex network theory approach
  - Gredeg and Skema
    - ★ econometric approach
    - ★ qualitative approach

Outline of the talk

1. **Datasets.**
   Scopus database, French and European grants databases, Theses.fr repository, European Patent Office.
2. **Complex network theory approach.**
   Preferential attachment hypergraph models, node trajectory analysis.
3. **Econometric approach.**
   Difference-in-differences, multivariate regression analysis, word2vec algorithm.
4. **Conclusions.**

Outline of the talk

1. **Datasets.**
   Scopus database, French and European grants databases,
   Theses.fr repository, European Patent Office.
2. **Complex network theory approach.**
   Preferential attachment hypergraph models, node trajectory analysis.
3. **Econometric approach.**
   Difference-in-differences, multivariate regression analysis,
   word2vec algorithm.
4. **Conclusions.**

Datasets

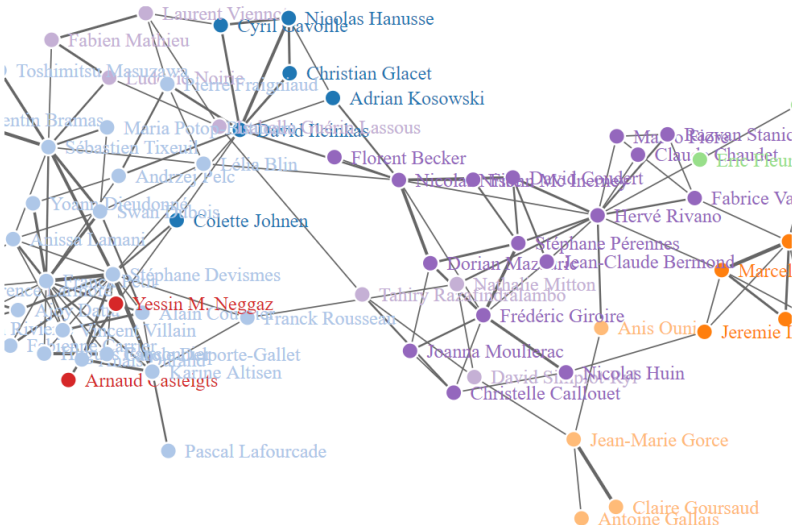
1. **Scopus** (Elsevier's abstract and citation database)
   $\sim 2.2 \cdot 10^6$ researchers, $\sim 3.9 \cdot 10^6$ articles (1990-2018)
2. **ANR** grants database (French National Research Agency)
   $\sim$ 13 000 grants (2006-2015)
3. **Horizon2020, FP7** grants databases (European Commission)
   $\sim$ 50 000 grants (2008-2020)
4. **Theses.fr** repository
   $\sim$ 170 000 PhD dissertations (2000-2014)
5. **European Patent Office**
   $\sim$ 135 000 French patent applications (2002-2015)

Outline of the talk

1. **Datasets.**
   Scopus database, French and European grants databases, Theses.fr repository, European Patent Office.
2. **Complex network theory approach.**
   Preferential attachment hypergraph models, node trajectory analysis.
3. **Econometric approach.**
   Difference-in-differences, multivariate regression analysis, word2vec algorithm.
4. **Conclusions.**

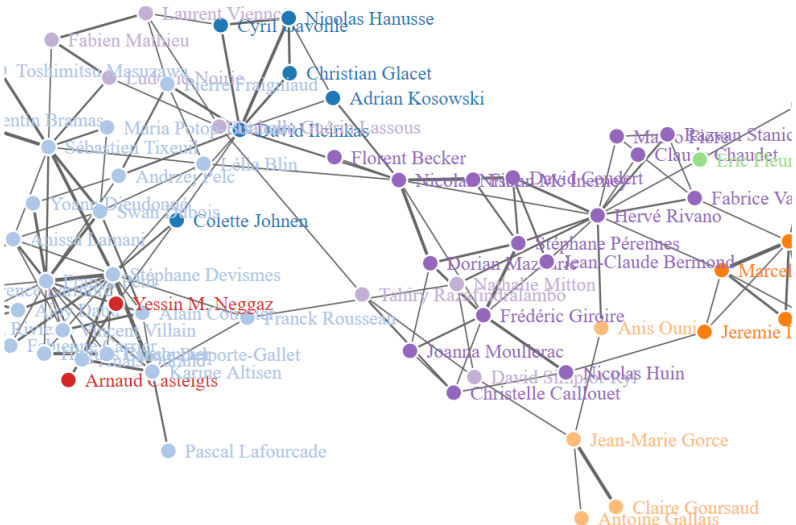
Preferential attachment model (PA)

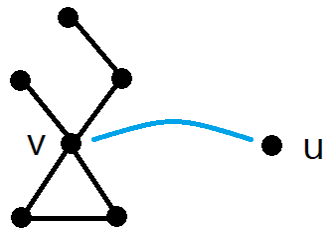
- Structure: graph. Nodes - scientists, edges - collaborations.



Preferential attachment model (PA)

- Structure: graph. Nodes - scientists, edges - collaborations.
- Growing process: preferential attachment.
  The more connected a node, the more likely to receive new links.





$$\mathbb{P}[v \text{ is chosen}] = \frac{\deg(v)}{\sum_{w \in G} \deg(w)}$$

$\deg(v)$ - number of edges incident with v
(e.g. number of collaborators)

⋆ A.L.Barabási, R.Albert, *Emergence of scaling in random networks.* Science, 286(5439):509–512, 1999

⋆ A.L.Barabási, R.Albert, *Statistical mechanics of complex networks.* Reviews of Modern Physics, 74(47):47–97, 2002

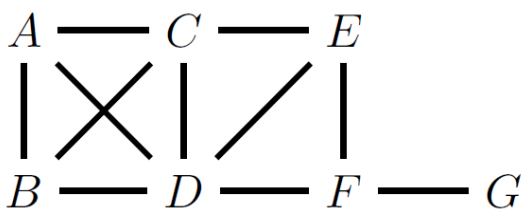
PA model against real collaboration networks

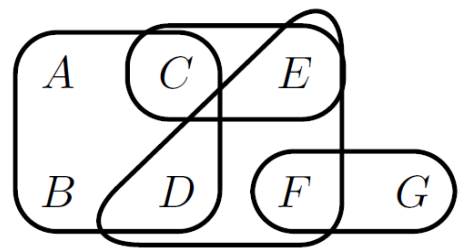
Preferential attachment model

1. Binary relations

Scientific collaboration network

1. Multiary relations





- Multiary relations model co-authorship, group collaboration...

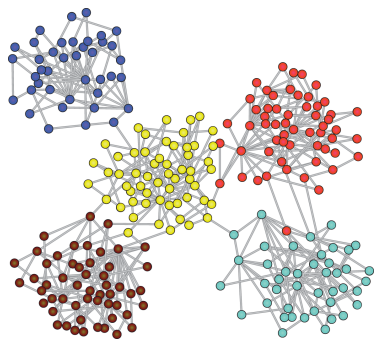
PA model against real collaboration networks

Preferential attachment model

1. Binary relations
2. Weak community structure

Scientific collaboration network

1. Multiary relations
2. Visible community structure



- Community structure models collaboration teams.

PA model against real collaboration networks

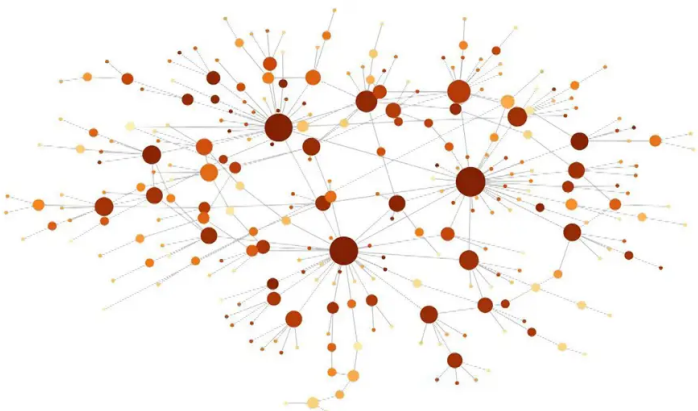
Preferential attachment model

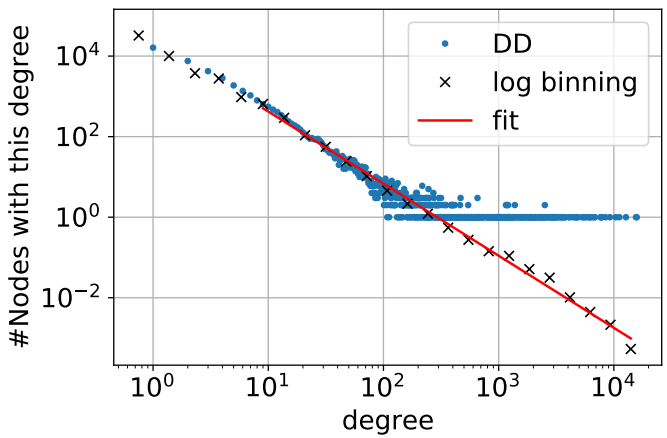
1. Binary relations
2. Weak community structure
3. Power-law degree distribution

Scientific collaboration network

1. Multiary relations
2. Visible community structure
3. Degree distribution with cut-off

Power-law: [fraction of nodes of degree k] $\sim k^{-\alpha}$, $\quad 2 < \alpha < 3$





PA model against real collaboration networks

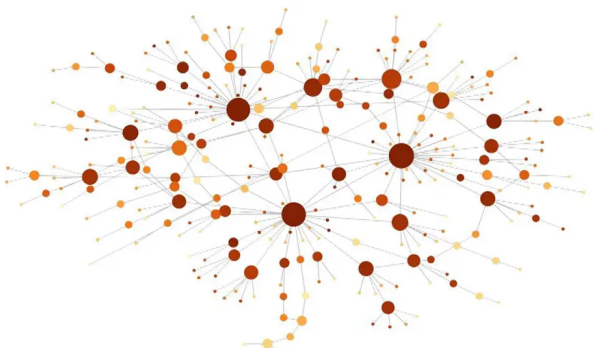
Preferential attachment model

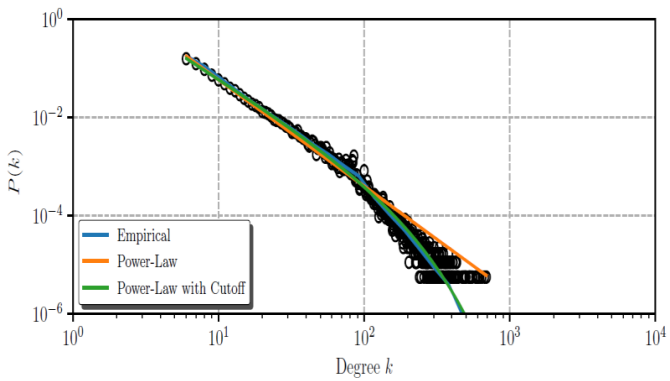
1. Binary relations
2. Weak community structure
3. Power-law degree distribution

Scientific collaboration network

1. Multiary relations
2. Visible community structure
3. Degree distribution with cut-off

Power-law with cutoff: [frac. of nodes of deg. k] $\sim k^{-\alpha}\gamma^{k}, \quad 0 < \gamma < 1$





- Cutoff may indicate the retirement of researchers.

New models of collaboration networks

Preferential attachment model

1. Binary relations
2. Weak community structure
3. Power-law degree distribution

Scientific collaboration network

1. Multiary relations
2. Visible community structure
3. Degree distribution with cutoff

- F. Giroire, N. Nisse, M. Sulkowska, T. Trolliet, *Preferential attachment hypergraph with high modularity*, under review.
  - provides a tool for analysing inter-team research collaboration
- F. Giroire, N. Nisse, K. Ohulchanskyi, M. Sulkowska, T. Trolliet, *Preferential attachment hypergraph model with node ageing*, under review.
  - model implementing phenomena such as retirement or career change

New models of collaboration networks

Preferential attachment model

1. Binary relations
2. Weak community structure
3. Power-law degree distribution

Scientific collaboration network

1. Multiary relations
2. Visible community structure
3. Degree distribution with cutoff

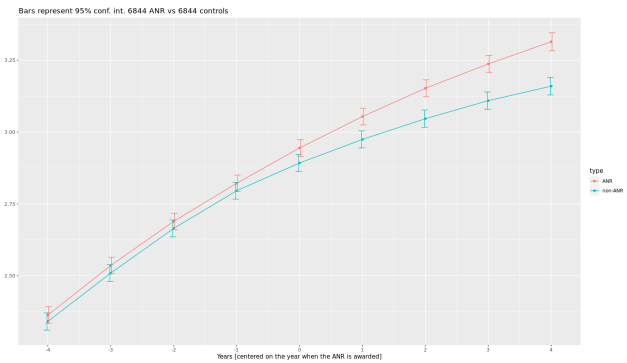
- F. Giroire, N. Nisse, M. Sulkowska, T. Trolliet, *Preferential attachment hypergraph with high modularity*, under review.
  - provides a tool for analysing inter-team reserach collaboration
- F. Giroire, N. Nisse, K. Ohulchanskyi, M. Sulkowska, T. Trolliet, *Preferential attachment hypergraph model with node ageing*, under review.
  - model implementing phenomena such as retirement or career change

New models of collaboration networks

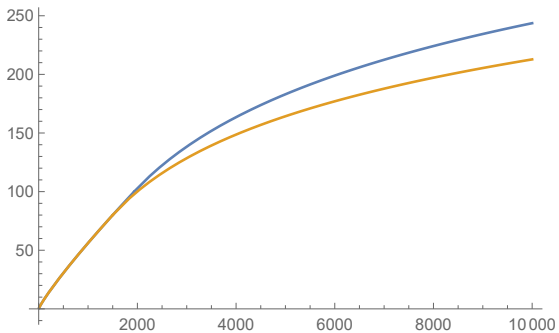
- F. Giroire, S. Pérennes, T. Trolliet, *A random growth model with any real or theoretical degree distribution*, International Conference on Complex Networks and their Applications (COMPLEX NETWORKS), (Madrid, Spain), Dec. 2020; journal version under review.
  - flexible tool for modeling scientific networks with arbitrary degree distribution

New models of collaboration networks

- F. Giroire, S. Pérennes, T. Trolliet, *A random growth model with any real or theoretical degree distribution*, International Conference on Complex Networks and their Applications (COMPLEX NETWORKS), (Madrid, Spain), Dec. 2020; journal version under review.
  - flexible tool for modeling scientific networks with arbitrary degree distribution
- A. Corsini, F. Giroire, N. Nisse, M. Pezzoni, M. Sulkowska, T. Trolliet, *Dynamic graph model with varying attachment function*, in preparation
  - graph model for comparing the productivity of granted and non-granted researchers



Impact of ANR funding
on the number of publications.
Scopus data (1990-2018, all disciplines)



Impact of funding
modeled by PA graph with
varying attachment function

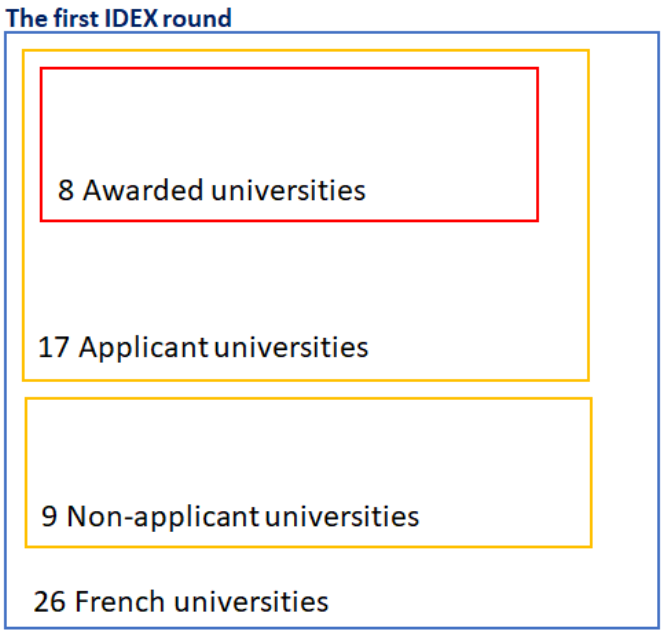
Outline of the talk

1. **Datasets.**
   Scopus database, French and European grants databases, Theses.fr repository, European Patent Office.
2. **Complex network theory approach.**
   Preferential attachment hypergraph models, node trajectory analysis.
3. **Econometric approach.**
   Difference-in-differences, multivariate regression analysis, word2vec algorithm.
4. **Conclusions.**

The impact of IDEX funding

Research questions:

1. What's the impact of **applying for** IDEX funding on researchers' outcomes?
2. What's the impact of **being awarded** IDEX funding on researchers' outcomes?



- A. Corsini, M. McCabe, P. Musso, M. Pezzoni, *The impact of IDEX funding: evidence on French researchers' activities*, in preparation
  - time frame: 2006-2015
  - methodology: difference-in-differences

The impact of IDEX funding

Researcher's outcome measures:

- publications
- citations (citation weighted publications)
- interdisciplinarity (interdisciplinary publications)
- collaborations within university (within university co-authors)
- collaborations in France (national co-authors)
- international collaborations (international co-authors)
- patenting (at least one patent)
- mentoring (at least one PhD student)
- fundraising ability (at least one ANR grant)

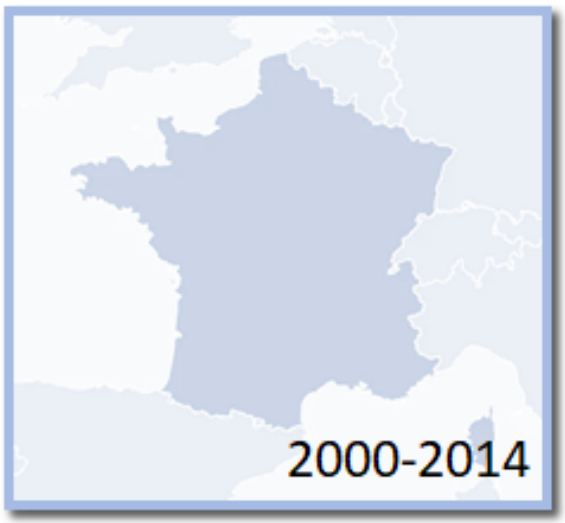
The impact of IDEX funding

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Publications	Citation Weighted Publications	Interdisciplinary Publications	Within University Co-authors	National Co-authors	International Co-authors	At least one Patent	At least one PhD Student	At least one ANR Grant
After 2011*Applicant	**0.045**	**-0.010**	**0.015**	**0.22*** **	**0.18**	**0.12**	**0.0018**	**0.0090*** **	**-0.0015**
After 2011*Awarded	**-0.038***	**-0.36** **	**-0.0043**	**0.069**	**0.20** **	**0.59*** **	**-0.00078**	**-0.0094*** **	**-0.0014**
Researcher controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Department controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	1.25***	0.76	0.54***	0.79***	0.29	-1.04***	0.010***	0.15***	0.0047
Observations	331,450	331,450	331,450	331,450	331,450	331,450	331,450	331,450	331,450
R-squared	0.039	0.011	0.026	0.032	0.045	0.051	0.005	0.022	0.066
N of researchers	33,145	33,145	33,145	33,145	33,145	33,145	33,145	33,145	33,145
Researcher fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

- field heterogeneity across Health Sciences, Life Sciences, Physical Sciences

What makes a PhD student productive?

- A. Corsini, M. Pezzoni, F. Visentin, *What makes a productive PhD student?*, under review
  - $\sim$ 77 000 students graduated from French universities
  - mathematics, engineering, physics, medicine-biology-chemistry
  - time frame: 2000-2014
  - methodology: multivariate regression analysis, word2vec algorithm



What makes a PhD student productive?

PhD student's productivity measures:

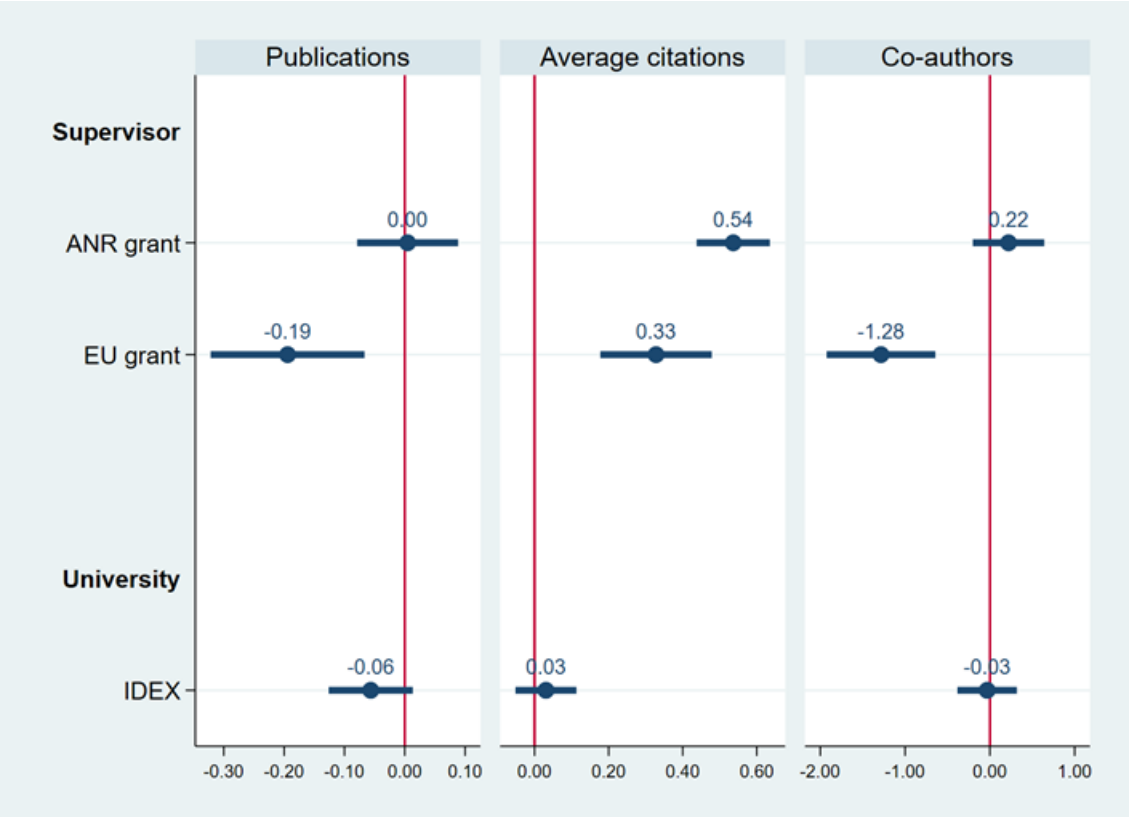
- number of publications
- average yearly citations received per publication
- number of distinct co-authors

proxies calculated between $t-3$ *and* $t+1$*, where* t *is the defense year*

77 143 PhD students	Mean	Sd	Min	Max
Publications	2.37	2.99	0.00	20.00
Average citations	2.11	3.51	0.00	98.14
Co-authors	8.93	15.37	0.00	200.00

What makes a PhD student productive?

Funding effect on PhD students' productivity



NOTE: The figure shows the coefficients of OLS estimates.

Bars represent 99% confidence intervals.

Outline of the talk

1. **Datasets.**
   Scopus database, French and European grants databases,
   Theses.fr repository, European Patent Office.
2. **Complex network theory approach.**
   Preferential attachment hypergraph models, node trajectory analysis.
3. **Econometric approach.**
   Difference-in-differences, multivariate regression analysis,
   word2vec algorithm.
4. **Conclusions.**

Conclusions

Research outcomes of SNIF project:

- Creation of a big, unique database for studying the impact of research funding.
- Development of new mathematical tools for analyzing scientific collaboration networks.
- Specific conclusions:
  - The IDEX application phase favors the creation of new research teams within the university.
  - At the universities awarded with IDEX, researchers enlarge their national and international network.
  - IDEX funding does not affect PhD students' productivity. Supervisors' individual ANR and EU grants boost the Ph.D. students' work visibility.

Conclusions

Other outcomes of SNIF project:

- Triggering collaboration within UCA between I3S, Inria, Gredeg and Skema.
- Triggering international collaboration with Maastricht University and Wrocław University of Science and Technology.
- SNIF participants have been able to obtain additional funding to continue their research effort:
  - Visiting professorship supported by the Graduate School of Business & Economics (GSBE) at Maastricht University.
  - European Patent Office grant. DOC-TRACK: STEM Doctoral Graduates and inventive activities in four European countries.
  - Postdoc position at I3S laboratory.
  - Erasmus+ scholarship.

References

1. A. Corisni, F. Giroire, N. Nisse, M. Pezzoni, M. Sulkowska, T. Trolliet, *Dynamic graph model with varying attachment function*, in preparation.
2. A. Corsini, M. McCabe, P. Musso, M. Pezzoni, *The impact of IDEX funding: evidence on French researchers' activities*, in preparation.
3. A. Corsini, M. Pezzoni, F. Visentin, *What makes a productive PhD student?*, under review.
4. F. Giroire, N. Nisse, K. Ohulchanskyi, M. Sulkowska, T. Trolliet, *Preferential attachment hypergraph model with node ageing*, under review.
5. F. Giroire, N. Nisse, M. Sulkowska, T. Trolliet, *Preferential attachment hypergraph with high modularity*, under review.
6. F. Giroire, S. Pérennes, T. Trolliet, *A random growth model with any real or theoretical degree distribution*, International Conference on Complex Networks and their Applications (COMPLEX NETWORKS), (Madrid, Spain), Dec. 2020; journal version under review.

Thank you!