



## Research collaboration network of Hungarian PhD holders – How to find the ideal collaboration partner for your scientific career?

Zsófia Vida<sup>1</sup>, István Péter Járay<sup>2</sup>, Balázs Lengyel<sup>2,3</sup>, Sándor Soós<sup>1</sup> <sup>1</sup>MTA Library and Information Center, <sup>2</sup>KRTK, <sup>3</sup>BCE

#### NKFI 116163 project:

MA 1826 K

"Career models and career advancement in research and development. Different patterns and inequalities in labour market opportunities, personal network building and work-life balance"

Innovation and Development in Linked Regions Session 2020 Conference of the Hungarian Regional Science Association

## "You can be successful at any age"



Sinatra R, Wang D, Deville P, Song C, Barabási A-L (2016) Quantifying the evolution of individual scientific impact. Science

# Science policy problem: The success of young scholars

According to FKA survey:

Wage problems: The salary isn't enough without research scholarships

How can you obtain research scholarships?

- How will someone be a successful researcher?
- What does help a young researcher to have high citation impact?

### Research collaborations are more and more important

- Creating new knowledge instead of little science, big science became usual (de Solla Price, 1979)
- The number of co-authored papers and the average number of authors per paper increased (Beaver, 2001, Glänzel & Schubert 2004, Wuchty et al., 2007)
- The unit of creating new knowledge is the research group instead of a single author (Ziman, 1994)



Wuchty S, Jones BF, Uzzi B(2007) The increasing dominance of teams in production of knowledge. Science

## Example1: Diversity in Hungarian movies' collaboration networks





Juhász S, Tóth G, Lengyel B (2019) PLOS ONE, forthcoming

C

В

## Example 2: Diverse and cohesive interaction in inventor



Tóth G, Lengyel B (2019) The Journal of Technology Transfer, forthcoming

## Cohesive versus diverse collaboration patterns



• Easier to reach and to learn complex knowledge

Aral S (2016) The future of weak ties. American Journal of Sociology

## Alternative hypotheses

#### H1: Diverse networks help young scholars to reach high citation impact

- New and innovative ideas
- Contacts in many communities



H2: Cohesive networks help young scholars to reach high citation impact

- Expert skills
- Strong embeddedness in the community

## Data

#### www.doktori.hu

16,151 PhD holders in 2017

- Researcher ID
- Title of thesis
- Year of defence
- Research area of doctoral school
- Name of supervisor

#### www.mtmt.hu

We can identify 9,415 PhD holders in MTMT database

They have 272,954 publications

- Authors ID and Coauthors author ID
- Year of publications
- Name of journals
- Citation number in 2017

#### Web of Science

- Reach better impact measures
- Identify foreign co-authors and esspecially their publications



## Data 2

- Second download from MTMT database in 2020.
- Contains:
  - PhD holders' publications
  - PhD holders' co-authors' publications
- Summary: via articles between 1990-2018
  - 41,110 authors who are registered in MTMT database
  - 1,057,977 articles

## Number of PhD holders and their publications



Number of PhD defenders between 1993-2017

Number of publications between 1990-2018

## Collaboration network and explanatory variable

• Network transformation from co-authorship network

Paper 1 -> Ego + A + B Paper 2 -> Ego + B + C + D Paper 3 -> Ego + A + F Paper 4 -> Ego + A + E

- Weighted growing network (year by year)
- Network indicators were counted every year to every PhD holder

in 2-step weighted Ego networks:

- Degree
- Global clustering (with and withouth the Ego)
- Burst's constraint
- Density (with and without the ego)
- Betweenness centrality
- High global clustering, high constraint and high density value indicates a cohesive network sturcture where the Ego's coauthors collaborate each other.



## OLS specification and control variables

$$\begin{split} HIV_{t+8} &= \alpha + \beta_1 HIV_t + \beta_2 DEG_t + \beta_3 GCLUS_t + \beta_4 (DEG_t \times GCLUS_t) + \beta_5 PAP_t \\ &+ \beta_6 PAP_{t+8} + D_t + DISC + \varepsilon \end{split}$$

$$\begin{split} &HIV_{t+8} \\ &= \alpha + \beta_1 HIV_t + \beta_2 DEG_t + \beta_3 GCLUS_t + \beta_4 (DEG_t \times GCLUS_t) \\ &+ \beta_5 \Delta DEG_{t+u} + \beta_6 \Delta GCLUS_{t+u} + \beta_7 PAP_t + \beta_8 PAP_{t+8} + D_t + DISC + \varepsilon \end{split}$$

	Dependent variable:			
	Citations	8 years po	st-defence	
	M1	M2	M3	
	(1)	(2)	(3)	
log(citations_cum_2 + 1)	0.928***	0.921***	0.919***	
	(0.008)	(0.009)	(0.009)	
<pre>log(papers_cum_2 + 1)</pre>	-0.112***	-0.128***	-0.134***	
	(0.016)	(0.017)	(0.017)	
<pre>log(papers_cum_8 - papers_cum_2 + 1)</pre>	0.322***	0.324***	0.325***	
	(0.011)	(0.011)	(0.011)	
<pre>log(degree_2 + 1)</pre>		0.041** (0.017)	0.028 (0.018)	
<pre>log(global_clustering_no_ego_2 + 1)</pre>			0.277*** (0.084)	
Constant	1.745***	1.755***	1.696***	
	(0.413)	(0.412)	(0.412)	
Observations	2,424	2,424	2,424	
R2	0.946	0.946	0.947	
Adjusted R2	0.945	0.945	0.945	
Note:	*p<0.	1; **p<0.05	; ***p<0.01	

	Dependent variable:			
	Citations	8 years po	st-defence	
	M1	M2	M3	
	(1)	(2)	(3)	
<pre>log(citations_cum_2 + 1)</pre>	0.928***	0.921***	0.919***	
	(0.008)	(0.009)	(0.009)	
<pre>log(papers_cum_2 + 1)</pre>	-0.112***	-0.128***	-0.134***	
	(0.016)	(0.017)	(0.017)	
<pre>log(papers_cum_8 - papers_cum_2 + 1)</pre>	0.322***	0.324***	0.325***	
	(0.011)	(0.011)	(0.011)	
<pre>log(degree_2 + 1)</pre>		0.041** (0.017)	0.028 (0.018)	
log(global_clustering_no_ego_2 + 1)			0.277*** (0.084)	
Constant	1.745***	1.755***	1.696***	
	(0.413)	(0.412)	(0.412)	
Observations	2,424	2,424	2,424	
R2	0.946	0.946	0.947	
Adjusted R2	0.945	0.945	0.945	
Note:	*p<0.	1; **p<0.05	; ***p<0.01	

		Dependent	vaniable	
	Dependent Variable:			
	Citations 8 years post-defence			
	Sciences (1)	Life Sci (2)	Engineer (3)	Social Sci (4)
log(citations cum 2 + 1)	0.953***	0.861***	0.868***	0.885***
	(0.017)	(0.018)	(0.034)	(0.025)
log(papers_cum_2 + 1)	-0.199***	-0.162***	0.066	-0.154***
	(0.036)	(0.036)	(0.058)	(0.050)
log(papers_cum_8 - papers_cum_2 + 1)	0.337***	0.330***	0.267***	0.416***
	(0.021)	(0.021)	(0.040)	(0.037)
<pre>log(degree_2 + 1)</pre>	0.033	0.087**	-0.067	0.041
	(0.035)	(0.039)	(0.071)	(0.051)
<pre>log(global_clustering_no_ego_2 + 1)</pre>	0.540***	0.467*	0.259	-0.140
	(0.168)	(0.274)	(0.371)	(0.221)
		0 077	0 00C***	***
Constant	(0 350)	(0.630)	(0.320)	2.298***
	(0.333)	(0.039)	(0.329)	(0.011)
Observations	512	672	185	382
R2	0.958	0.926	0.945	0.927
Adjusted R2	0.956	0.923	0.937	0.921
Note:		*p<0.1	; **p<0.05	; ***p<0.01

		Dopondont	vaniable	
			variable:	
	Citations 8 years post-defence			efence
	Sciences	Life Sci	Engineer	Social Sci
	(1)	(2)	(3)	(4)
log(citations_cum_2 + 1)	0.953***	0.861***	0.868***	0.885***
	(0.017)	(0.018)	(0.034)	(0.025)
log(papers_cum_2 + 1)	-0.199***	-0.162***	0.066	-0.154***
	(0.036)	(0.036)	(0.058)	(0.050)
$\log(\text{namers cum 8} - \text{namers cum 2} + 1)$	0.337***	0.330***	0.267***	0.416***
105(babe) 2_cam_0 babe) 2_cam_2 + 1)	(0.021)	(0.021)	(0.040)	(0.037)
$\log(degree 2 + 1)$	0 022	0 007**	0 067	0 011
log(degree_z + 1)	(0.035)	(0.039)	(0.071)	(0.051)
	• •			
<pre>log(global_clustering_no_ego_2 + 1)</pre>	0.540***	0.467*	0.259	-0.140
	(0.168)	(0.274)	(0.371)	(0.221)
Constant	1.653***	0.837	0.926***	2.298***
	(0.359)	(0.639)	(0.329)	(0.611)
Observations	512	672	185	382
R2	0.958	0.926	0.945	0.927
Adjusted R2	0.956	0.923	0.937	0.921
Note:		*p<0.1	; **p<0.05	; ***p<0.01

## Conclusion

- The recognition of tender age researchers is mainly determined by the quality of the publications created during and near after PhD study years
- Instead of study years, it is more important to publish after obtaining a PhD degree
- During PhD study years and the next few years after obtaining a PhD degree, cohesive collaboration networks tend to facilitate success (H2)

## Limitations and future plans

- We see only co-author who are registered in MTMT database, so we can't see foreign co-authors.
- Using raw citation numbers.
- What is the role of the supervisor?
- What is the role of foreign co-authors?
- Do researchers have a connection to "important" authors? using other network indicators
- What are the PhD holders' geographical mobility like?

## Thank you!

Zsófia Vida, István Péter Járay, Balázs Lengyel, Sándor Soós

vida.zsofia@konyvtar.mta.hu istvanpeterjaray@gmail.com lengyel.balazs@krtk.mta.hu soos.sandor@konyvtar.mta.hu