A Network Analysis Study on the Internationalization of Higher Education Research

Aliakbar Akbaritabar, Giovanni Barbato

akbaritabar@DZHW.eu, giovanni.barbato@UNIMI.it

5/6/2020
UCL Virtual Workshop

Introduction: Object of Study

- Internationalization of Higher Education (HE) research, empirically analyzed through international co-authorship
- **International co-authorship**: when two or more scholars from different countries have co-authored one or more papers (Katz and Martin 1997; Kwiek 2018): steady growth + increasing authors, organizations and countries involved (Wagner, Park, and Leydesdorff 2015). However. . . social sciences less studied;
- HE research as a multidisciplinary and emergent field of study characterized by a **high fragmentation**: marked division between *policy-based* and *learning* and *teaching* studies and scholars; low communication between them; lots of **part-time researchers** (Tight 2004; Horta and Jung 2014; Kim, Horta, and Jung 2017)

Introduction: our Theoretical Approach

- Social Network Theory (SNT)
- The set of direct and indirect relationships among actors provide opportunities and constrains that affect their behaviors (Granovetter 1977)
- The nature, the characteristics and the internal dynamics of the co-authorship network shape the evolution and growth of HE research and should thus be considered as explanatory factors (Acedo et al. 2006; Wagner and Leydesdorff 2005a, 2005b; Biancani and McFarland 2013; Wagner, Park, and Leydesdorff 2015)

Three Research Goals

- **DESCRIPTIVE**: to investigate the main features of the international co-authorship network in HE research: how much HE research publications are international and collaborative $\Longrightarrow$ descriptive network analysis
- **EXPLANATORY**: to study how specific network-based mechanisms (e.g. preferential attachment and homophily) contribute to explain the growth and evolution of the international HE co-authorship network $\Longrightarrow$ Exponential Random Graph Model (ERGM), bipartite community detection: rarely applied to study HE research
- **METHODOLOGICAL**: to highlight what is the level of improvement in accuracy of network construction obtained by disambiguation of organization and author names $\Longrightarrow$ disambiguation procedures often overlooked

Literature Review: internationalization of HE research

- Tight (2007) and (2014)'s analysis: division between North American HE community and the rest of the world in terms of authors' country of origin and citation patterns: *methdological nationalism* + different development of HE systems;
- Kuzhabekova, Hendel, and Chapman (2015) (descriptive network analysis): % of internationally co-authored publications around 10%: slightly increasing but still concentrating in few Western countries;
- Avdeev (2019): intensity of international co-authorship depend on geographical distance and linguistic similarity of authors
- $\Longrightarrow$ the nature and dynamics of relationships between actors (authors or universities) have been mostly overlooked in explaining what drives internationalization of HE co-authorship network (Biancani and McFarland 2013; Kezar 2014).

Social network mechanisms as explanatory variables

- **ENDOGENEOUS FACTORS**: structural recurrent processes based on the presence of ties, that explain the formation of other ties:
  - **PREFERENTIAL ATTACHMENT** = propensity to start a new co-authorship with scholars who already present a large number of previous co-authorships (Barabâsi et al. 2002) $\Longrightarrow$ theory of cumulative advantage in science (Merton 1968)
- **EXOGENEOUS FACTORS**: nodes' attributes that can influence the formation of new ties:
  - **HOMOPHILY** = tendency of nodes to start new co-authorship with whom they present similarities on one or more attributes (Sciabolazza et al. 2017; Zhang et al. 2018)
- **NETWORK COMMUNITIES**: (Palla, Barabási, and Vicsek 2007): communities of network actors in which the patterns of collaboration may persist across or within them $\Longrightarrow$ fragmentation vs connection of the entire network

Hypotheses

- **H1: Cumulative advantage (preferential attachment)**: Scholars with higher previous collaboration have higher probability to be chosen for new collaborations.
- **H2: Citations in first 3 years**: Collaborative publications (with higher number of authors) receive higher citations i.e., have higher impact and outreach.
- **H3: Difference in total publications' count**: Scholars with similar levels of *research productivity* have higher tendency to collaborate with each other.
- **H4: Difference in first publication year**: Scholars with similar academic status and seniority have higher tendency to collaborate with each other.
- **H5: Difference in last publication year**: Scholars with similar length of publication activity (*scientific experience*) have higher tendency to collaborate with each other.
- **H6: Social closure among groups**: There are cohesive sub-groups in co-authorship network with higher tendency to collaborate *within* themselves.

Data and Methods

- 54 HE journals, *Article* and *Review*, 1996-2018, in-house Scopus
- **GRID** (Global Research Identifier Database) 10th December 2019
- **Research Organization Registry (ROR)** [local] API 18th December 2019
- Scopus author numeric IDs
- Bipartite Network Modeling (community detection, ERGMs)
  - *bipartite*: Organization/author and paper ties
- Models probability of tie existence as dependent variable
- Constant Pots Model in `Leidenalg` library
- Emphasizes the importance of links *within* communities vs. *between*
- Communities such that the link density between the communities (*external density*) is lower than γ and the link density within communities (*internal density*) is higher than γ
- $\gamma = 7 \times 10^{-5}$, yield *90* communities with a uniform distribution of org/pubs for *organizations*

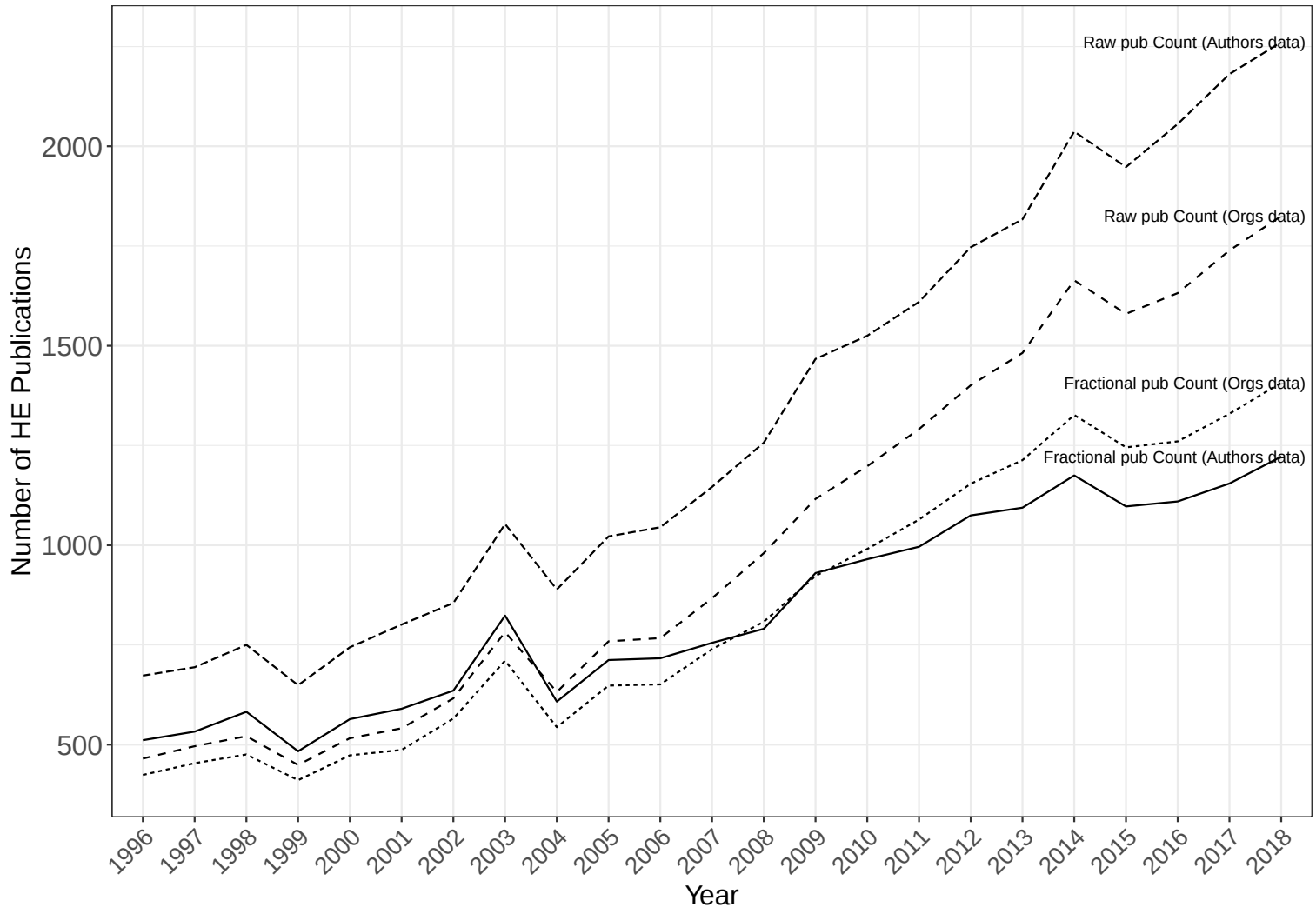
Findings: Description of networks

- High improvement in network accuracy with **disambiguation** (1/12 orgs)

Metrics	Non disamb orgs	Disamb orgs	Non disamb authors	Disamb authors SID	Disamb authors orgs
Number of connected components	22,203	928	14,033	12,958	1,241
Number of biparitite nodes	79,507	28,397	97,848	73,586	4,557
Number of biparitite edges	58,289	34,751	97,113	67,936	3,391
Number of biparitite nodes G	11,091	26,013	24,975	21,524	54
Number of biparitite nodes G %	14	92	26	29	1
Number of biparitite edges G	11,757	33,289	33,488	26,865	58
Number of biparitite edges G %	20	96	34	40	2
Density G	0	0	0	0	0
Number orgs/authors	48,076	4,122	66,417	42,170	3,028
Number orgs/authors G	5,977	3,080	14,994	10,503	28
Number papers	31,431	24,275	31,431	31,416	1,529
Number papers G	5,114	22,933	9,981	11,021	26

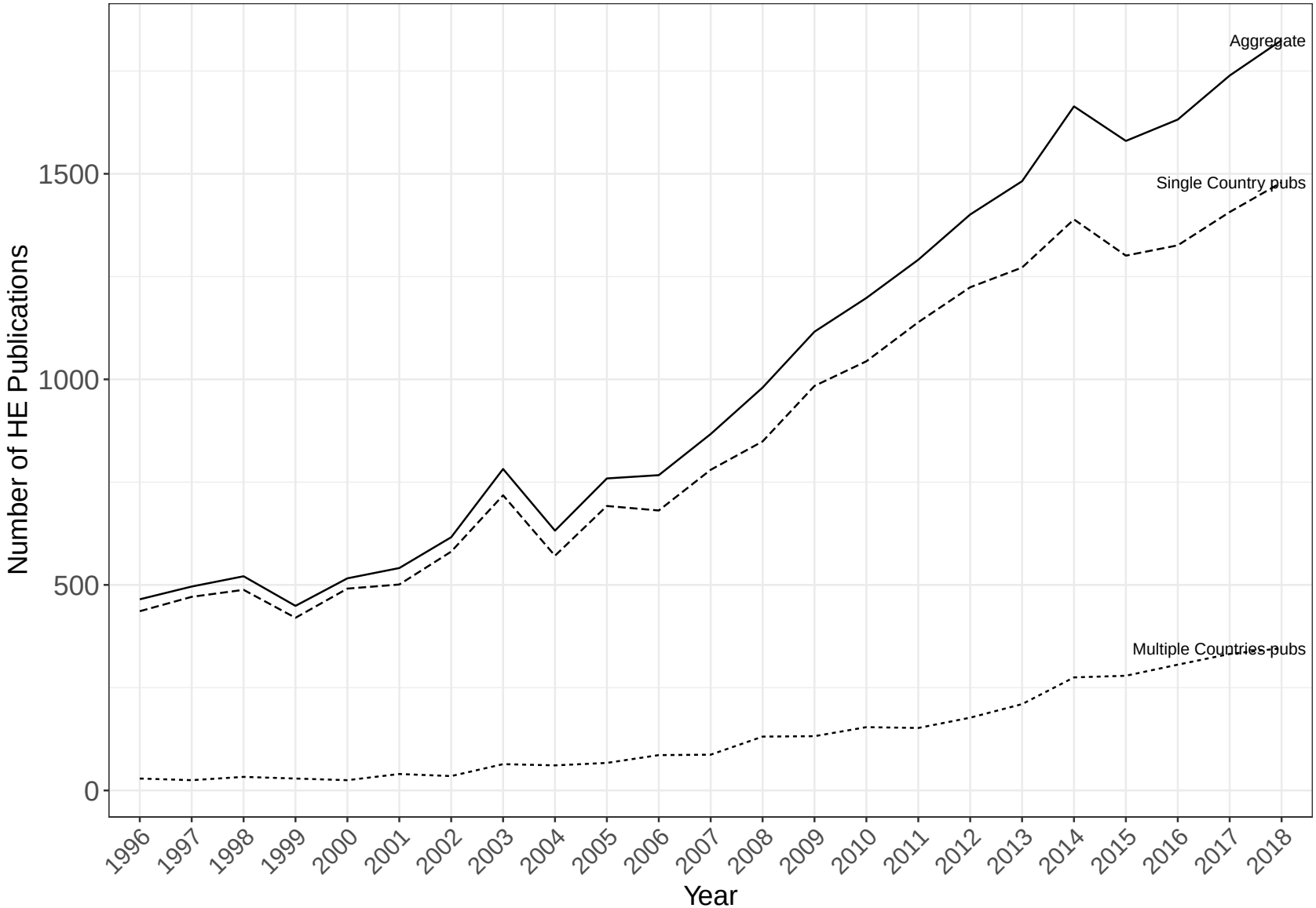
Publications in HE research

- Higher collaborative work in recent years
- Is this signal of internationalization?



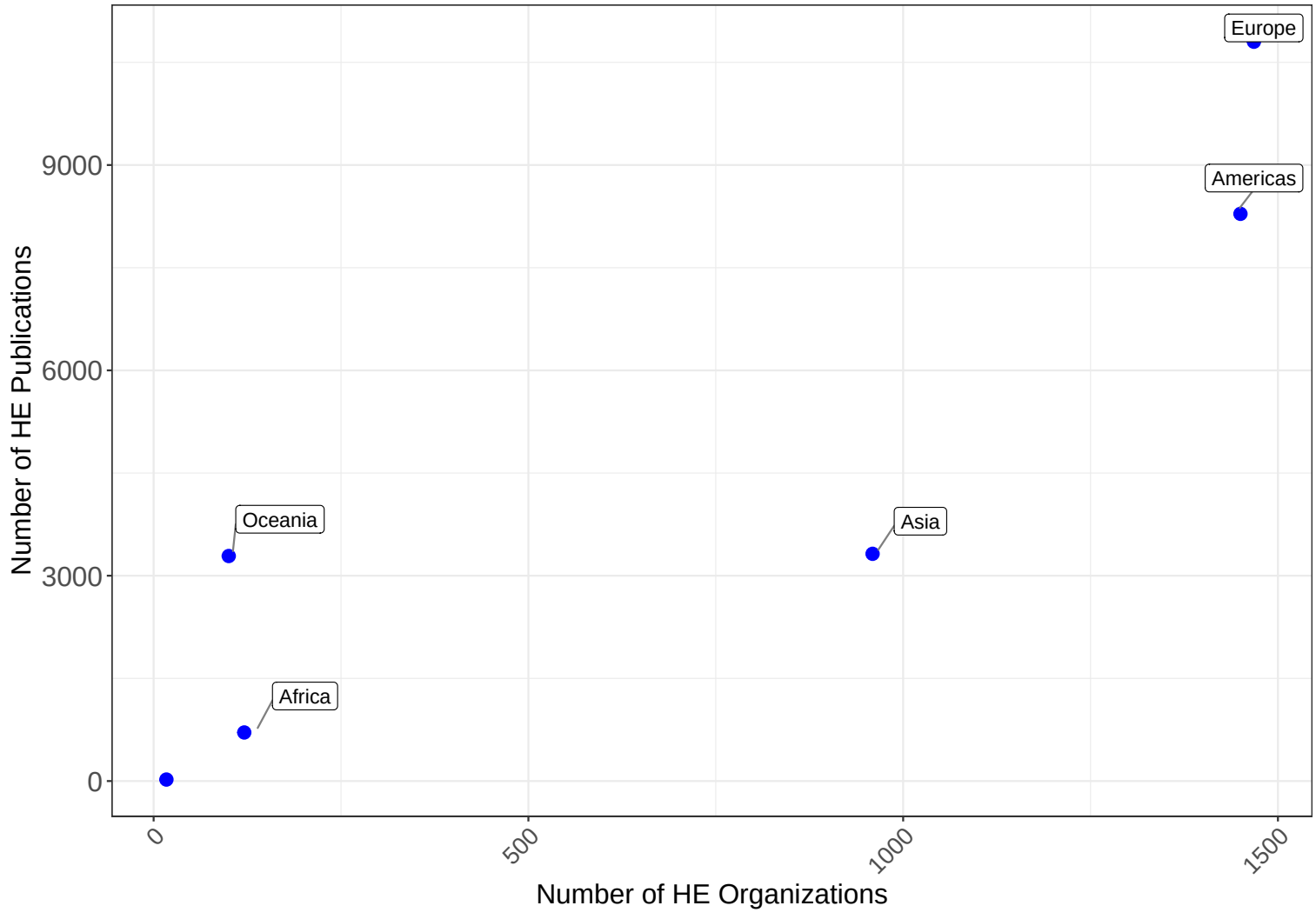
Single authored vs. Multiple country publications

- No, the majority are still single country co-authorship



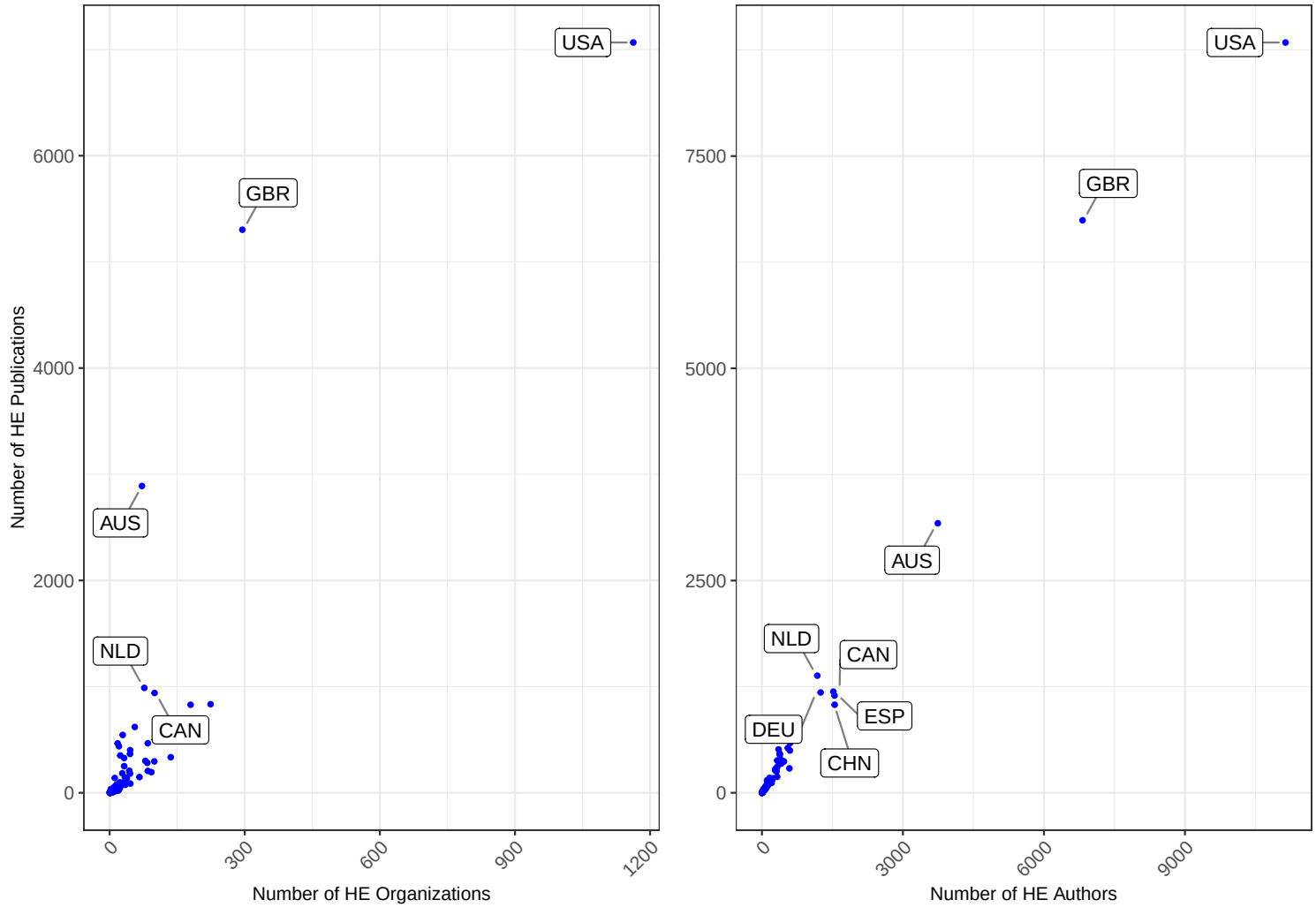
HE organizations vs. HE publications (continents)

- Europe is more prolific in the aggregate picture; Asia vs. Oceania.



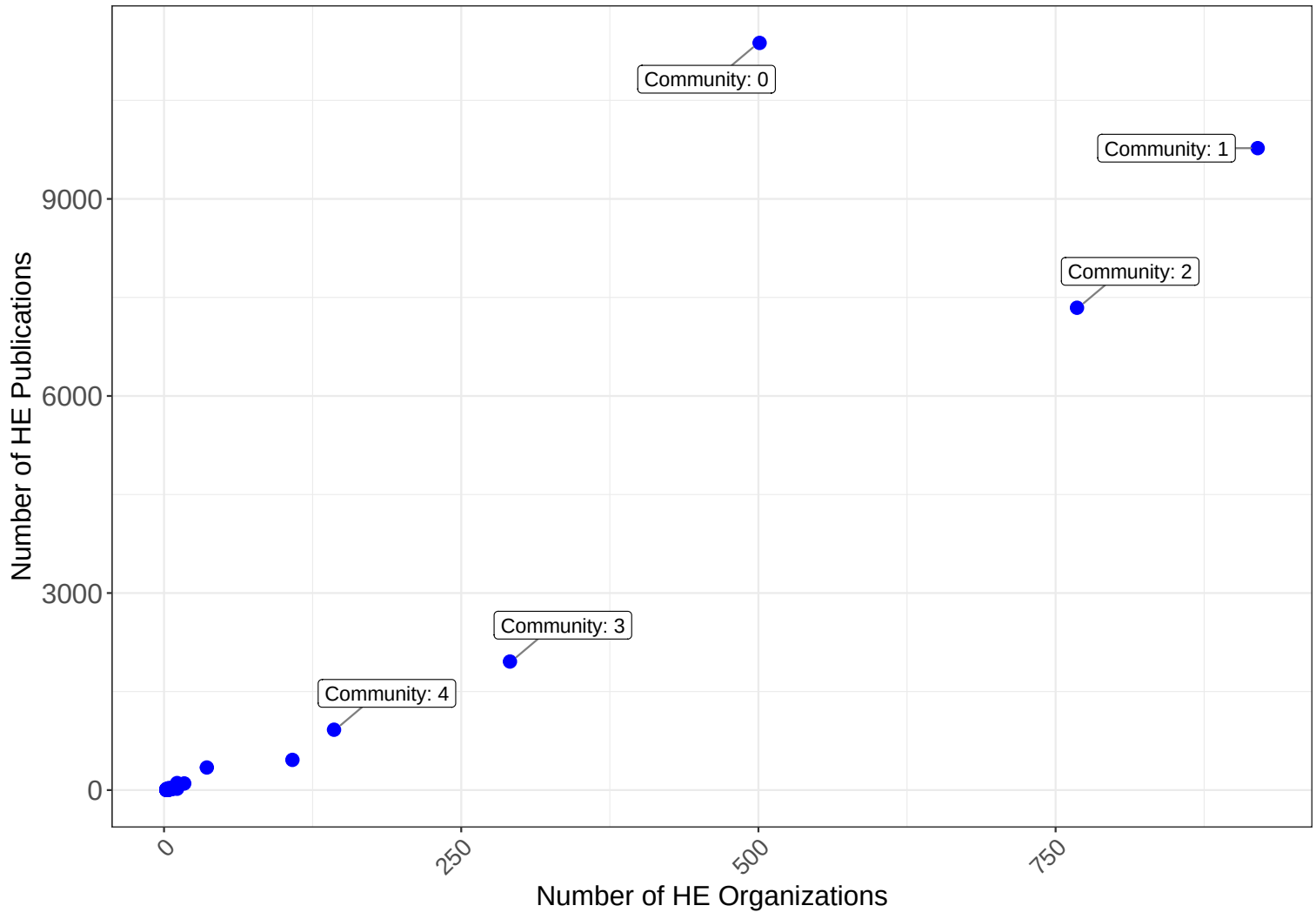
HE organizations/authors vs. HE publications (countries)

- In single country view, US is the most prolific
- GBR and AUS, smaller HE population, relatively prolific



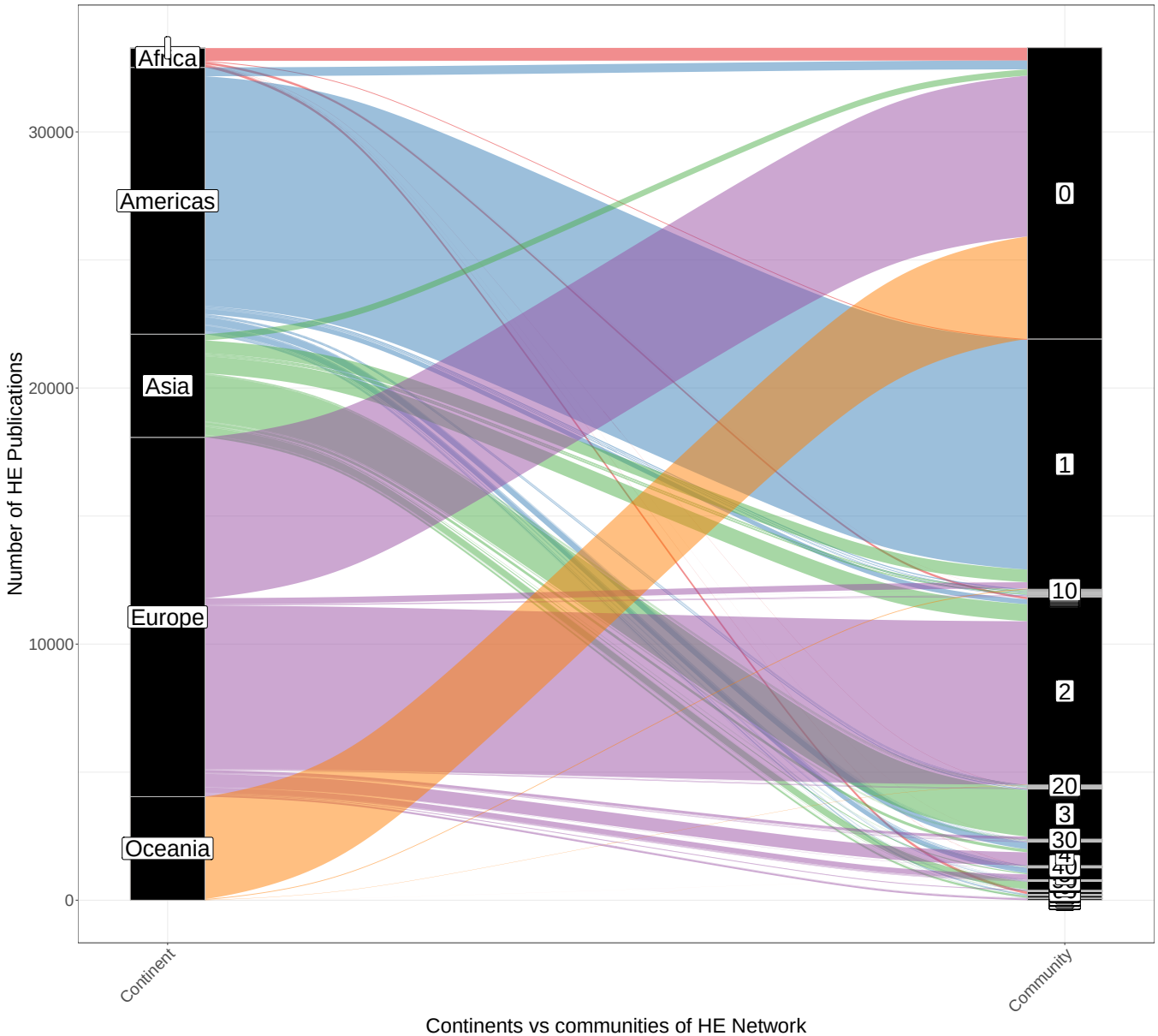
Composition of HE communities

- H6: Larger communities not necessarily more prolific (see 0)



Composition of HE communities (2/2)

- H6: Fragmentation of HE network



Bipartite ERGM Results

- H1: Tendency towards higher preferential attachment in *authors* and lower preferential attachment in *organizations*
- H2: Publications with higher number of *citations* are results of more collaborative work (in line with literature)
- H3: Higher tendency of collaboration between most prolific organizations with lower prolific ones
- H5: Similar date of latest publication (e.g., publishing until recently) increased the probability of ties in *authors* and decreased it in *organization* (i.e., different activity levels collaborate)

Table 3: Bipartite ERGMs results explaining effect of organization and author attributes and structural variables on co-authorship tie existence

	Giant component of HE co-authorship					
	Author level			Organization level		
	(1)	(2)	(3)	(4)	(5)	(6)
Ties	−8.570*** (0.008)	−7.159*** (2.407)	−6.850*** (0.085)	−7.332*** (0.032)	−18.079** (7.305)	−34.093*** (0.417)
Preferential attachment (H1)	0.328*** (0.041)		1.416*** (0.001)	−6.677*** (0.071)		−6.584*** (0.003)
Citations in first 3 years (H2)		0.001 (0.001)	0.001 (0.001)		0.035*** (0.001)	0.009*** (0.001)
Difference in total pubs (H3)		0.0004 (0.002)	0.001 (0.003)		−0.001 (0.001)	−0.003** (0.001)
Difference in first pub (H4)		0.003 (0.002)	0.004* (0.002)		−0.002 (0.002)	−0.005 (0.004)
Difference in last pub (H5)		−0.003* (0.002)	−0.005** (0.002)		0.006* (0.004)	0.019*** (0.004)
Akaike Inf. Crit.	479,013.800	481,103.300	479,454.100	78,870.930	94,285.810	74,916.810
Bayesian Inf. Crit.	479,063.500	481,186.100	479,570.100	78,919.150	94,366.180	75,029.320

Note: *p<0.1; **p<0.05; ***p<0.01

Discussion and Conclusions 1/3

- **FIRST RQ**: *how much are HE research publications international and collaborative?*
- Increasing degree of *team science* (Wuchty, Jones, and Uzzi 2007): from single-author to multi-authors publications
- But still a highly fragmented field: 928 disconnected components, 90 sub-groups of giant component
- Low degree of internationalization: publications by authors from a single country are more prevalent than multiple country ones (13.2%)
- Majority of internationally co-authored papers come from western countries: Europe and Americas have 3 times more HE publications than others (Kuzhabekova, Hendel, and Chapman 2015)

Discussion and Conclusions 2/3

- **SECOND RQ**: *how network mechanisms contribute to explain evolution of the international HE co-authorship network?*
- Preferential attachment (H1): is an explanatory mechanism at the author level, while its effect is lower at organization level; confirms literature on other discipline (Wagner and Leydesdorff 2005b; Zhang et al. 2018)
- There are minority of *key authors* while in the *organization* level there is not
- Citations in first 3 years (H2): Collaborative publications (higher number of authors) have higher impact and outreach
- Homophily effects: Hetrophily of total pubs and latest active years.
- Community detection (H6): several cohesive sub-groups that present a higher number of co-authorship ties within themselves rather than between them $\Longrightarrow$ atomization of the HE field (confirms the literature)

Discussion and Conclusions 3/3

- **THIRD RQ**: what is the level of improvement in accuracy of network construction obtained by disambiguation of organization and author names?
- Out of each 12 organizations, 1 is reliable
- Scopus IDs, out of each 2 authors, 1 is reliable

Limitations

- In the data collection process: publications on HE outside HE journals are not included
- Journals not indexed in Scopus are not included
- We only cover **successful** scientific collaboration that results in a *publication* while as outlined by Katz and Martin (1997), scientific collaboration can take multiple forms (e.g., funding proposals co-authorship, patents)
- Further homophily effects: Gender, academic level, continents, countries, etc. not implemented due to non-converging models

Thanks for your attention!

- akbaritabar@DZHW.eu
- giovanni.barbato@UNIMI.it

Part of bibliography

Acedo, Francisco José, Carmen Barroso, Cristobal Casanueva, and José Luis Galán. 2006. "Co-Authorship in Management and Organizational Studies: An Empirical and Network Analysis." *Journal of Management Studies* 43 (5). Wiley Online Library: 957–83.

Avdeev, Stanislav. 2019. "International Collaboration in Higher Education Research: A Gravity Model Approach." *Higher School of Economics Research Paper No. WP BRP* 54.

Barabâsi, Albert-Laszlo, Hawoong Jeong, Zoltan Néda, Erzsebet Ravasz, Andras Schubert, and Tamas Vicsek. 2002. "Evolution of the Social Network of Scientific Collaborations." *Physica A: Statistical Mechanics and Its Applications* 311 (3-4). Elsevier: 590–614.

Biancani, Susan, and Daniel A McFarland. 2013. "Social Networks Research in Higher Education." In *Higher Education: Handbook of Theory and Research*, 151–215. Springer.

Granovetter, Mark S. 1977. "The Strength of Weak Ties." In *Social Networks*, 347–67. Elsevier.

Horta, Hugo, and Jisun Jung. 2014. "Higher Education Research in Asia: An Archipelago, Two Continents or Merely Atomization?" *Higher Education* 68 (1). Springer: 117–34.

Katz, J. S., and B. R. Martin. 1997. "What Is Research Collaboration?" *Research Policy* 26 (1): 1–18.

Kezar, Adrianna. 2014. "Higher Education Change and Social Networks: A Review of Research." *The Journal of Higher Education* 85 (1). Taylor & Francis: 91–125.

Kim, Yangson, Hugo Horta, and Jisun Jung. 2017. "Higher Education Research in Hong Kong, Japan, China, and Malaysia: Exploring Research Community Cohesion and the Integration of Thematic Approaches." *Studies in Higher Education* 42 (1). Taylor & Francis: 149–68.

Kuzhabekova, Aliya, Darwin D Hendel, and David W Chapman. 2015. "Mapping Global Research on International Higher Education." *Research in Higher Education* 56 (8). Springer: 861–82.

Kwiek, Marek. 2018. "International Research Collaboration and International Research Orientation: Comparative Findings About European Academics." *Journal of Studies in International Education* 22 (2). SAGE Publications Sage CA: Los Angeles, CA: 136–60.

Merton, Robert K. 1968. "The Matthew Effect in Science: The Reward and Communication Systems of Science Are Considered." *Science* 159 (3810). American Association for the Advancement of Science: 56–63.

Palla, Gergely, Albert-László Barabási, and Tamás Vicsek. 2007. "Quantifying Social Group Evolution." *Nature* 446