The dynamics of collaboration and its implications — from careers to Europe

LANET’17 - Puebla MX

Paul Erdös collaboration network

EU high-skilled mobility ‘brain-drain’ network

Alexander M. Petersen
Management of Complex Systems Department, mist.ucmerced.edu
Science is a multi-scale system with emergent complexity. A very practical question is How to measure scientific output and impact at various scales while accounting for systemic heterogeneity.

- Science
- Country
- Institution
- Lab / Team
- Individual
- Paper

Institutional factors
Organizational factors
Social/Behavioral/Cognitive factors

The evolution of careers from a collaboration perspective

Paul Erdös (1913-1996): collaboration network at career age 10, 30, present day*
Ego collaboration network: quantifying *dynamic & heterogenous* patterns of collaboration within scientific careers

Sir Andre K. Geim

\# publications, \( N_i (2012) = 217 \)

\( S_i = 303 \) coauthors

*The average copublication duration*

\( \langle L_i \rangle = 2.1 \) years, \( \langle K_i \rangle = 3.7 \) pubs.

- Measuring the duration \( L_{ij} \) of the tie (time b/w 1st and last copublication)
- Measuring the intensity \( K_{ij} \) of the tie (# of copublications)
- Measuring the net scientific impact \( C_{ij} \) of the tie (net citation tally for pubs. between \( i \) and \( j \))

How important are academic “Life partners”?
- Division/Diversity of labor
- Breadth/Depth of expertise
- Risk/Reward sharing
- Ethics of credit distribution & free-riding
Mathematical definition for extreme outlier in an exponential distribution:

“super tie” threshold \( K^c_i = (\langle K_i \rangle - 1) \ln(S_i) \)

- \( N_i(2010) = 909 \) publications
- \( S_i = 541 \) coauthors
- \( \langle K_i \rangle = 5.7 \) papers

Comparison of the biology data yields a K-S statistic indicating significant disagreement with the exponential distribution, while the physics datasets indicate excellent agreement with the exponential distribution.
Do super-ties correlate with higher citations?

Unit of analysis: publication $p$

Hierarchical “fixed effects” model: 473 researchers indexed by $i$

Dependent variable $z_{i,p} = \text{the citation impact of publication } p \text{ normalized to baseline citation levels defined by other papers published in the same year } y$.

$z_{i,p} = \left( \frac{\ln c_{i,p,y} - \langle \ln c_y \rangle}{\sigma[\ln c_y]} \right)$

This measure maps $c_{i,p,y}$ to a stable normal distribution $N(0,1)$ >> appropriate for comparing citation impact across time.

$R_{i,p}$: A super-tie indicator variable = 1 if at least one of the coauthors is a super tie, and 0 otherwise. 52% of publications have $R=1$.

$N_i(t_p)$: number of papers up to year $t_p$ ≈ prestige measure

$S_i(t_p)$: number of distinct coauthors up to year $t_p$ ≈ collaboration radius measuring access to new/old team members

$a_{i,p}$: number of coauthors ≈ proxy for coordination costs and technology level

$t_p$: publication year of $p$, measured as a career age, accounting for aging and cumulative advantage effects, learning and prestige

Fixed-effects model - measures each researcher against his/her baseline $z_{i,p}$

$z_{i,p} = \beta_R R_{i,p} + \beta_a \ln a_{i,p} + \beta_t t_{i,p} + \beta_N \ln N_i(t_p) + \beta_S \ln S_i(t_p) + \beta_i + \epsilon_{i,p}$

On average:
- 1 in 25 collaborators qualify as a super-tie
- 1 in 2 publications include a super-tie
The significant + value of super-ties

Comparing publications with and without a super-tie — within a researcher’s publication portfolio (i.e. fixed effects) — compared to the author’s publications with R=0 (the counterfactual), the publications with R=1 have 0.2 $\sigma_z$ higher citations.

In terms of real citations, this citation boost corresponds to a roughly 20% citation increase at the publication level!

Emphasizes who in addition to how many coauthors

Plausible explanations: compounding self-citations, reputation arising from larger formal and informal social network; added value of skill complementarity, trust, conviction, commitment, experience, collocation, moral support, risk-profit sharing

Quantifying the impact of weak, strong, and super ties in scientific careers. PNAS, 2015
Is Europe Evolving Toward an Integrated Research Area?

How does this question manifest in the cross-border mobility/collaboration activities in Europe?
Quantifying the impact of EU policies on cross-border R&D integration

EU Horizon 2020 Impact Assessment: one of the 5 key objectives is to “encourage cross-border training and career development, and supporting research infrastructures”

The EU spends ~ 10% of government level R&D budget on programs with explicit cross-border criteria, compared to < 1% for non-EU countries

Thus, EU initiatives aimed at integrating the “European Research Area” (ERA) serves as a “treatment”
Methods and Data

Complex networks approach

Geocoded data
(a) 2.4 million patents filed in the EPO and
(b) 0.26 million scientific publications

4 patent networks
(i) co-inventor
(ii) co-applicant
(iii) citations
(iv) mobility

1 publication network
(v) co-author

NUTS3 regions ≈ province/district/county size
Intra-country -vs- Cross-border Networks

Methods and Data

**Geocoded data**
(a) 2.4 million patents filed in the EPO and
(b) 0.26 million scientific publications

4 patent networks
(i) co-inventor
(ii) co-applicant
(iii) citations
(iv) mobility
(v) co-author

EU countries

non-EU countries
Comparing the community structure of the 2009 EU-15 and US coinventor networks

Q1: Are the scientific borders in the EU any different than the geo-political borders?

Q2: has there been an intensification in cross-border R&D activity in Europe vis-a-vis other OECD countries (control group used to quantify the “treatment effect”)

Community structure of the 2009 EU-15 and USA coinventor network.
Communities are labeled by their most-central region and were generated using the Louvain algorithm — communities are sub-sets of nodes (NUTS3 regions) more strongly linked to one another than to nodes outside.
Quantifying the impact of EU policies on cross-border R&D integration

Q: do “treated” EU countries have different cross-border collaboration patterns than “untreated” non-EU countries above global trends

Null hypothesis $H_0(t \mid t_0)$: With respect to base year $t_0$, there is relative increase in # of links $L_t$ in year $t$

- $\tau_t > 0$ : accept $H_0(t \mid t_0)$
- $\tau_t \leq 0$ : reject $H_0(t \mid t_0)$
Q2: Is there any positive trend in the rate of cross-border activity within the EU — relative to the world?

**Causal (DiDiD) model:** We measured the effect of EU institutional integration policies by quantifying the relative rate of cross-border links — comparing a) **within** versus **across**-borders links, b) **EU** vs **Non-EU** links, c) and across time.

**Patent co-inventor network:**
Additional cross-border links per region pair

**Why stagnation since 2004??**

*Is Europe Evolving Toward an Integrated Research Area?*
*Science, 2014*

*The evolution of networks of innovators within and across borders: Evidence from patent data* *Research Policy, 2015*
Divergence in Eastern - Western integration within the global science system

With globalization, the rate of international collaboration has largely been increasing, however there is considerable regional variation.

For example, comparing the decade before and after 2004, while Western Europe and North America experienced a 36-42% increase in the rate of cross-border collaboration (per publication), Eastern Europe and Asia have experienced much slower 9% growth.

These diverging trends point to the importance of historical, socio-technological, and geographic factors underlying the globalization of science.

So why have Western and Eastern Europe followed different cross-border collaboration paths?

Estimating cross-border collaboration rates in Europe under the counterfactual — no 2004 EU enlargement — using the Synthetic Control Method (SCM)

**RUBIN CAUSALITY MODEL**

If you give a little kid a balloon, how do you really know it makes them happy? And happier by how much?

\[ W = 1 : \text{State in which receives balloon} \]
\[ W = 0 : \text{State in which does not receive balloon} \]

\[ Y(1) : \text{Outcome of child } i \text{ if } W=1 \]
\[ Y(0) : \text{Outcome of same child } i \text{ if } W=0 \]

Causal “treatment” effect = \( Y(1) - Y(0) \)

**Fundamental Challenge**: How to measure the counterfactual outcome in a world where there is only one reality — i.e. only one observed outcome, either \( Y(1) \) or \( Y(0) \) ?

26 non-EU control group countries:
AR, AM, AZ, BY, CA, CN, CO, CU, IN, IL, JP, KZ, KW, KG, MG, MX, MN, PA, RU, RS, SG, KR, TT, TR, UA, US

Country-level control/matching variables:
[Scimago] Cross-border pubs, Total pubs, Citations
[World Bank] GDP per capita, Govt. Expenditure on R&D

What would have happened had there NOT BEEN a 2004 expansion of the European Union??

Counterintuitively, there would have been MORE cross-border integration had there been no European Union enlargement!!
FIG. S3: Synthetic Control Method – Cross-border publications. Counterfactual estimates of the number of cross-border publications, $\hat{Y}_i$, for each of the 2004 and 2007 new EU entrants. Each solid line indicates the (real) observed number of cross-border publications by year. Each dashed line indicates the synthetic estimates had the country not entered the EU. The SCM explanatory variables used to estimate $\hat{Y}_i$ are the total number of publications ($\log_{10} D_{Alli,t}$), the normalized citations ($R_{Alli,t}$), the per-capita GDP ($\log_{10} GDP_{pci,t}$), and government expenditure on R&D, $e_{i,t}$; “All” indicates the total across all subject areas ($s$).

($\%$) is the percent difference between the net area under the real and synthetic curves after the entry year (indicated by each dashed vertical line), serving as a basic estimate of the net impact of the 2004 enlargement on each country. ($\%$) $> 0$ for 11 of the 12 countries; the mean and standard deviation of the individual values are $h_i \pm \sigma_i = 22 \pm 29$ percent.

Twelve 2004/2007 Entrant Countries
(A) SCM results for the fraction $f_t$ of cross-border publications and (B) the total number $\chi_t$ of cross-border publications. The solid curves represent the real data, while the dashed curves represent the estimates for the counter-factual scenario of no 2004 EU enlargement.

- Note that the $\chi_t$ that represent the incumbent pre-2004 EU countries are divided by 10 in order to facilitate visualizing all the curves on the same scale.
- $\delta$ and $\delta(\%)$ represent the difference between the real and synthetic curves after 2004, providing estimates of the “2004 EU Entry” effect on cross-border European integration.
- (C,D) Estimation of the significance level of the SCM results using the “permutation test”
Unintended consequence: there would have been more cross-border integration without EU enlargement.

Why?? Brain drain: largely from Eastern to Western European countries.

The micro-level mechanism connecting x-border collab. & brain drain:

Before 2004 EU enlargement:
- East-West collaboration
- East to West mobility

After 2004 EU enlargement:
- X

Quantifying the negative impact of brain drain on the integration of European science.

Science Advances, 2017
What is the marginal impact of ‘brain-drain’ ($B_i$) on the international collaboration rate ($f_i$) of the average country within each EU group?

The impact of brain drain on cross-border collaboration is even more negative for the new 2004 EU members.
Human & Social capital perspectives on the value of EU membership

Gateway hub for Eastern countries

Gateway hub for Northern countries

Forward-looking questions:
What will be the impact of a ‘Hard Brexit’ on:
(a) the import/export of high-skilled labor between the UK & EU? (human capital)
(b) the social capital (i.e. research networks) in Europe?
Centrality in the high-skilled mobility network — before and after the 2004 enlargement
How might the ‘hardness’ of Brexit affect EU high-skilled mobility networks?

High-skilled labour mobility in Europe before and after the 2004 enlargement. J. Royal Society Interface, 2017
Reorganization of high-skilled mobility pathways in Europe in a hypothetical “Hard Brexit” scenario

Empirical: real data

Hypothetical: Brexit scenario with severe restrictions on international mobility to the UK
Thank you!

A special thanks to my collaborators:
Omar Doria, Andrea Morescalchi, Fabio Pammolli, Orion Penner, Michelangelo Puliga

Papers available at:
http://physics.bu.edu/~amp17/

- **Quantifying the impact of weak, strong, and super ties in scientific careers.**


- **Quantifying the negative impact of brain drain on the integration of European science.** O. A. Doria Arrieta, F. Pammolli, A. M. Petersen. Science Advances, 2017

- **High-skilled labour mobility in Europe before and after the 2004 enlargement.** A. M. Petersen & M. Puliga. J. Royal Society Interface, 2017
The dynamics of collaboration and its implications — from careers to Europe

Collaboration in science is intrinsically interpersonal, and as a result, the networks of (in)formal relations are characteristically dynamic. In this talk I will discuss recent work on how these dynamics impact career paths, with implications as far-reaching as the evolution of entire national research systems. In the first part I will focus on the remarkably wide variation of collaborative strengths within research careers. In order to demonstrate the added value of long-term interpersonal partnership on career outcomes, I will present the results of a within-career (i.e. researcher fixed-effects) regression model showing that publications authored by a given scientist that include her strongest collaborators have higher citation impact relative to those publications that do not. These results point to the advantage of “super” social ties characterized by trust, conviction, and commitment. In the second part I will discuss the aggregate implications of collaboration dynamics at the level of the European Research Area (ERA) — a longstanding vision of the European Union to develop a competitive and integrated innovation system through directed cross-country policies. In order to measure the EU’s progress towards the establishment of the ERA, we analyzed the rate of international publication for 32 European countries using data extracted from millions of academic publications from 1996 to 2012. We then used the EU 2004/2007 enlargement, a large policy intervention representing a multi-country and multi-stage “quasi-experiment”, to provide causal insights into the interaction between two types of cross-border activity: human mobility and international collaboration. Our results reveal a counterintuitive result — that the twelve 2004/2007 entrant EU countries would have had higher rates of cross-border collaboration had they not joined the EU — thereby identifying an unintended consequence of labor market integration in Europe. Together, these results identify East-to-West European brain drain as a mechanism underlying the stalled integration of the ERA.