

# International Scientific Collaboration

as Seen From the OpenAIRE Data Perspective



**Collaborativity**

# How to measure "collaborativity" of a country?

---

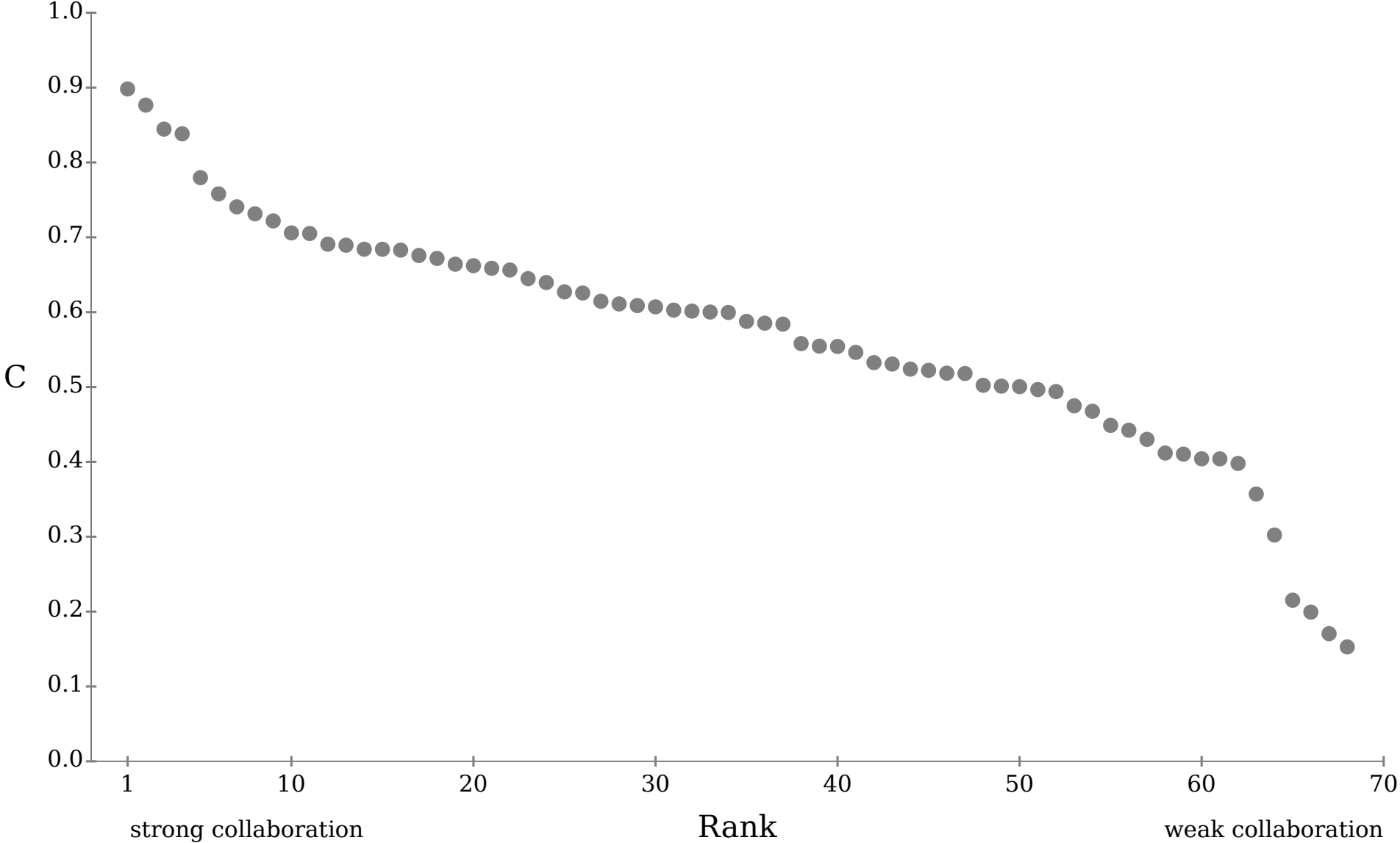
- Background
  - Some countries are more active in international collaboration than the other
  - How to quantify it?
- Proposal – collaboration coefficient C for country X

$$C(X) = \frac{\text{Number of docs. affiliated by X and at least one other country}}{\text{Number of all docs. affiliated by X}}$$

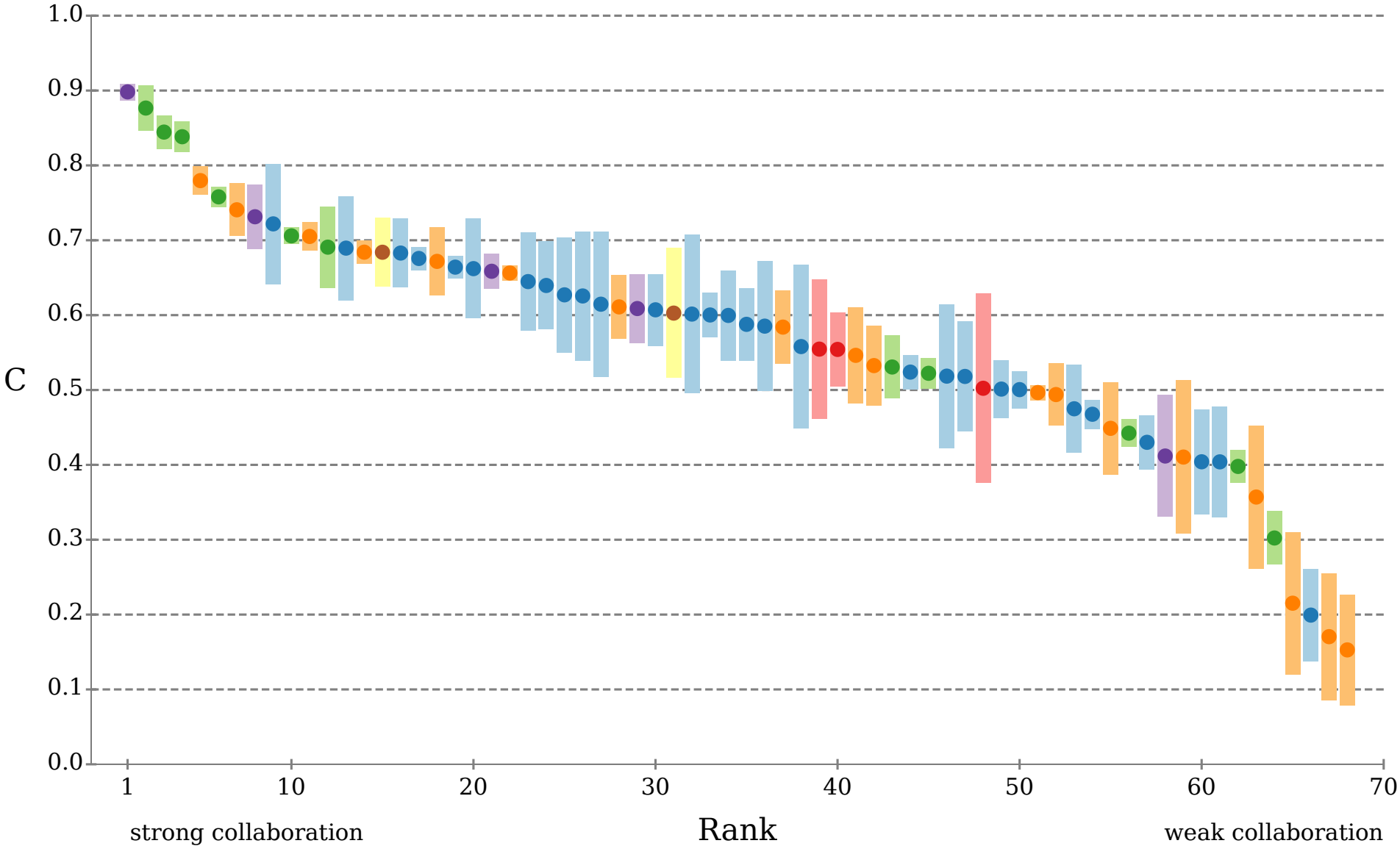
- $0.0 \leq C \leq 1.0$
- $C=0.0$  – no documents authored within international collaboration
- $C=1.0$  – only documents authored within international collaboration

# Coefficient C for different countries

---



# Coefficient C for different countries



# Analysis of C coefficient results

---

- For main stream countries  $0.4 \leq C \leq 0.8$
- Countries with high values of C
  - Leaders/outliers – Peru, Kenya, Tanzania, Uganda
  - Often countries with poor economy and low scientific production
  - However, small and rich countries also score high (see, Hong Kong, Switzerland, Belgium)
- Countries with low values of C
  - Leaders/outliers – Iran, South Korea, Turkey, India, Nigeria, Japan
  - Mostly countries isolated because of the alphabet, geography etc.
  - Often with considerable total number of publications produced

# Pairwise collaboration

# Measuring collaboration between two countries

---

- First idea
  - To characterize collaboration between countries X and Y, use number of papers affiliated by both countries  $N(X,Y)$
- Problems
  - Very strongly influenced by the most productive countries (i.e., with highest total publications count  $N(X)$ )
  - Difficult to interpret, because of the large mismatches between number of publications
  - Difficult to compare  $N(X_1,Y_1)$  and  $N(X_2,Y_2)$ , when all involved countries  $X_1,Y_1,X_2,Y_2$  are different



# Most frequent countries coauthoring with Poland

---

## Most frequent countries coauthoring with PL

X	N(X;PL)
US	2253
DE	2164
UK	1351
FR	1205
IT	893
ES	718
NL	556
CH	527
RU	514
SE	513

## Most productive countries

X	N(X)
US	317014
UK	134986
DE	114231
CN	97366
FR	74015
IT	70547
JP	68061
IN	66977
CA	61387
AU	44936

- Note high overlap of countries collaborating with Poland with most productive countries world-wide

# Useful reference

---

## THE MEASUREMENT OF INTERNATIONAL SCIENTIFIC COLLABORATION

TERTTU LUUKKONEN,\* R. J. W. TUISSEN,\*\* O. PERSSON,\*\*\* G. SIVERTSEN\*\*\*\*

*\*The Academy of Finland, P. O. Box 57, SF-00551 Helsinki (Finland)*

*\*\*Centre for Science and Technology Studies, University of Leiden,  
P. O. Box 9555, 2300 RB Leiden (The Netherlands)*

*\*\*\*Inforsk Group, University of Umeå, S-901 87 Umeå (Sweden)*

*\*\*\*\*Institute for Studies in Research and Higher Education, Munthesgate 29, N-0260 Oslo 2 (Norway)*

(Received February 15, 1993)

A growing science policy interest in international scientific collaboration has brought about a multitude of studies which attempt to measure the extent of international scientific collaboration between countries and to explore intercountry collaborative networks. This paper attempts to clarify the methodology that is being used or can be used for this purpose and discusses the adequacy of the methods. The paper concludes that, in an analysis of collaborative links, it is essential to use both absolute and relative measures. The latter normalize differences in country size. Each yields a different type of information. Absolute measures yield an answer to questions such as which countries are central in the international network of science, whether collaborative links reveal a centre - periphery relationship, and which countries are the most important collaborative partners of another country. Relative measures provide answers to questions of the intensity of collaborative links.

**Scientometrics, 28, 15-36 (1993)**

# Normalized measures of collaboration

---

- Salton's measure

$$S(X, Y) = \frac{N(X, Y)}{\sqrt{N(X) N(Y)}}$$

- Jaccard's measure

$$J(X, Y) = \frac{N(X, Y)}{N(X) + N(Y) - N(X, Y)}$$

- Properties

- Both measures normalized to [0,1] 😊
- However, no direct interpretation 😞
- Comparison between different countries still questionable 😞

# Normalized measures of collaboration for PL

---

X	S(X;PL)	N(X;PL)
DE	0.04576	2164
FR	0.03166	1205
US	0.02860	2253
UK	0.02628	1351
CZ	0.02614	300
ES	0.02476	718
RU	0.02427	514
IT	0.02403	893
UA	0.02362	212
BE	0.02214	430

X	J(X;PL)	N(X;PL)
DE	0.01644	2164
FR	0.01304	1205
RU	0.01225	514
ES	0.01161	718
CZ	0.01154	300
BE	0.01119	430
SE	0.01101	513
CH	0.01017	527
IT	0.01001	893
DK	0.00992	354

- Comparison of the top-S and top-J countries for PL
  - J punishes stronger for the total number of publications (UK,US)
  - 7 out of 10 countries repeat in top-10 for both J & S, but rankings differ
  - Still correlated to most productive countries

# Collaboration between two countries - reloaded

---

- Another approach to measuring collaboration
  - What would be the number of  $N(X,Y)$ , if the collaborators were selected independently and randomly

$$N^*(X,Y) = \frac{N(X)N(Y)}{N}$$

- The ratio of real  $N(X,Y)$  and expected  $N^*(X,Y)$  makes a good candidate for measure of collaboration strength

$$T(X,Y) = \frac{N(X,Y)}{N^*(X,Y)}$$

- Properties
  - Explicitly decouples from the most productive countries 😊
  - Has clear interpretation 😊
  - Can be compared between different countries 😊

# One more take at collaboration with PL

**Best scores in collaboration with PL**

X	T(X;PL)	N(X;PL)
UA	3.042	212
CZ	2.632	300
FI	1.389	261
RU	1.324	514
BE	1.317	430
DK	1.269	354
GR	1.180	226
AT	1.157	312
DE	1.118	2164
SE	1.101	513

**Worst scores in collaboration with PL**

X	T(X;PL)	N(X;PL)
JP	0.271	313
AU	0.376	286
CA	0.409	425
US	0.420	2253
UK	0.591	1351
PT	0.649	261
NL	0.733	556
IT	0.747	893
CH	0.949	527
FR	0.961	1205

- Insights from best & worst T scores
  - No visible correlation with most productive countries
  - High-scoring countries are usually close to Poland, whereas low scoring are distant
  - Contrary to the previous findings, collab. with giants (UK, US) is weak

Only countries with  $N(X,Y) > 200$  were included

# Top T-scoring connections worldwide

---

X	Y	T(X;Y)	N(X;Y)
EG	SA	24.785	720
MY	SA	14.163	718
MY	TH	9.695	360
DK	NO	5.319	923
FI	SE	4.840	1278
NO	SE	4.605	1335
AU	NZ	4.462	1029
FI	NO	4.100	479
DK	SE	4.019	1576
MW	UK	3.940	331
RU	UA	3.886	317
FR	TN	3.665	362
DK	FI	3.568	564
BE	NL	3.370	2515
CN	HK	3.367	998

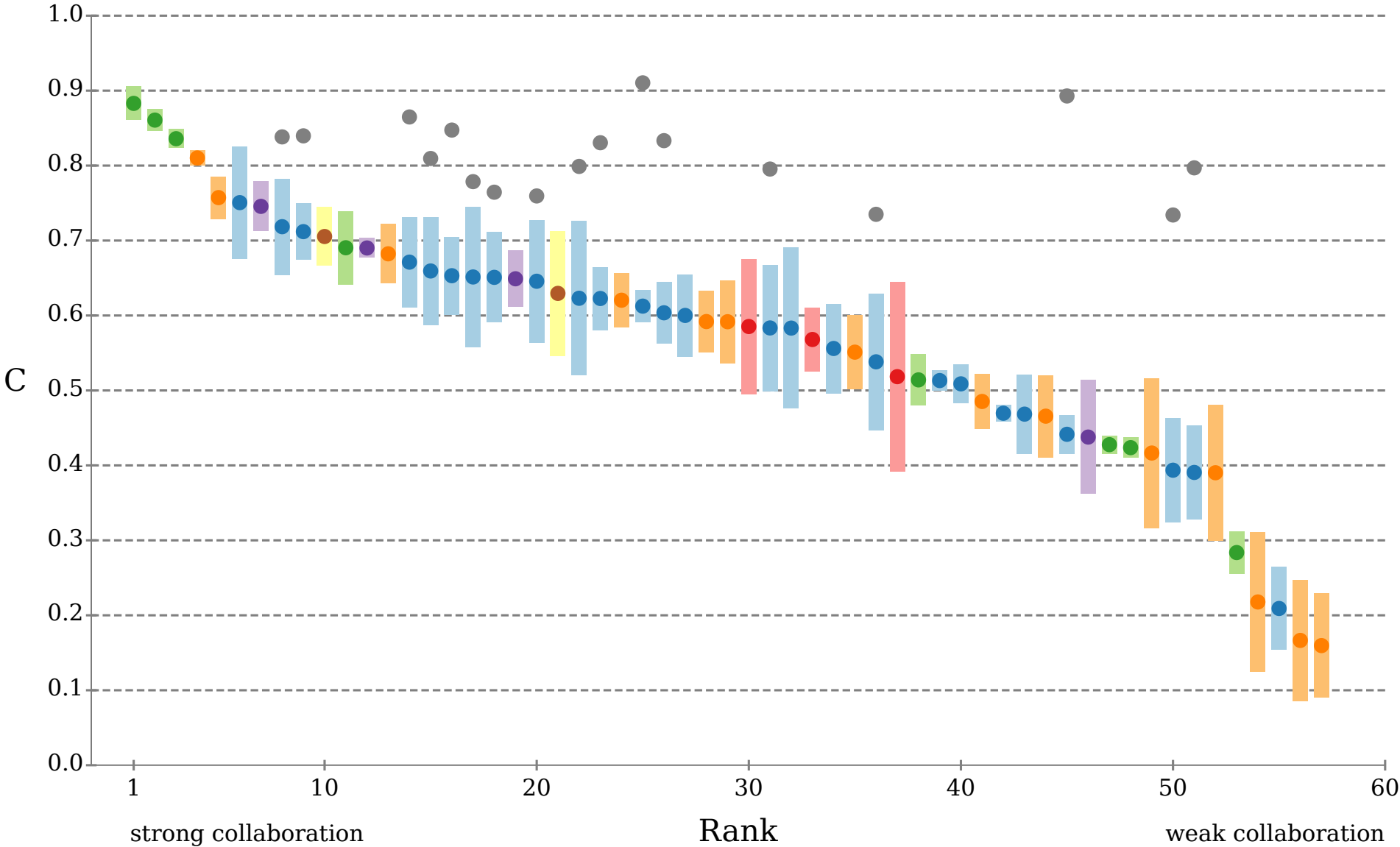
- Many expected connections
  - Nordic countries DK, NO, FI, SE
  - Australia and New Zealand
  - Russia and Ukraine
  - Some post-colonial links, e.g, Malawi-UK, France-Tunisia
  - Belgium and the Netherlands
  - China and Hong-Kong

Only countries with  $N(X,Y) > 300$  were included

**FP7**



# Collaboration coefficient for publications from FP7



## Poland influence of the FP7 on collaboration

---

Country	FP7	ALL	Ratio
CL	8.145	1.857	4.386
NO	3.671	0.999	3.675
US	1.221	0.391	3.123
FI	2.610	1.316	1.983
RU	2.621	1.658	1.581
DK	1.858	1.187	1.565
SE	1.472	1.025	1.436
GB	0.857	0.619	1.384
NL	1.031	0.745	1.384
AT	1.458	1.231	1.184
CH	1.026	0.889	1.154
DE	1.127	1.043	1.081
BE	1.389	1.38	1.007
IT	0.832	0.833	0.999
FR	0.851	0.943	0.902
ES	0.970	1.108	0.875

**Thank you!**