

Quantifying knowledge synchronisation in the 21st century

Jisung Yoon^{1,2}, Jinseo Park³, Jinhyuk Yun⁴ and Woo-Sung Jung^{1,5}

1. Dept. of Industrial and Management Engineering, POSTECH, Pohang, Korea
2. School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN, USA
3. Center for global R&D Analysis, Korea Institute of Science and Technology Information(KISTI), Korea
4. School of AI Convergence, Soongsil University, Seoul 06978, Republic of Korea.
5. Dept. of Industrial and Physics, POSTECH, Pohang, Korea

Jisung Yoon

@jisungyoon8

jisung.yoon92@gmail.com

Wikimedia Research Showcase
Jun. 15. 2022



Knowledge and knowledge structure

Knowledge is a familiarity, awareness, or understanding of the subject which is acquired through **experience** or **education** by perceiving, discovering, or learning.

-Wikipedia



Ref. Wikipedia

Human understanding is root of the general laws of nature that organize all experience

-Immanuel Kant

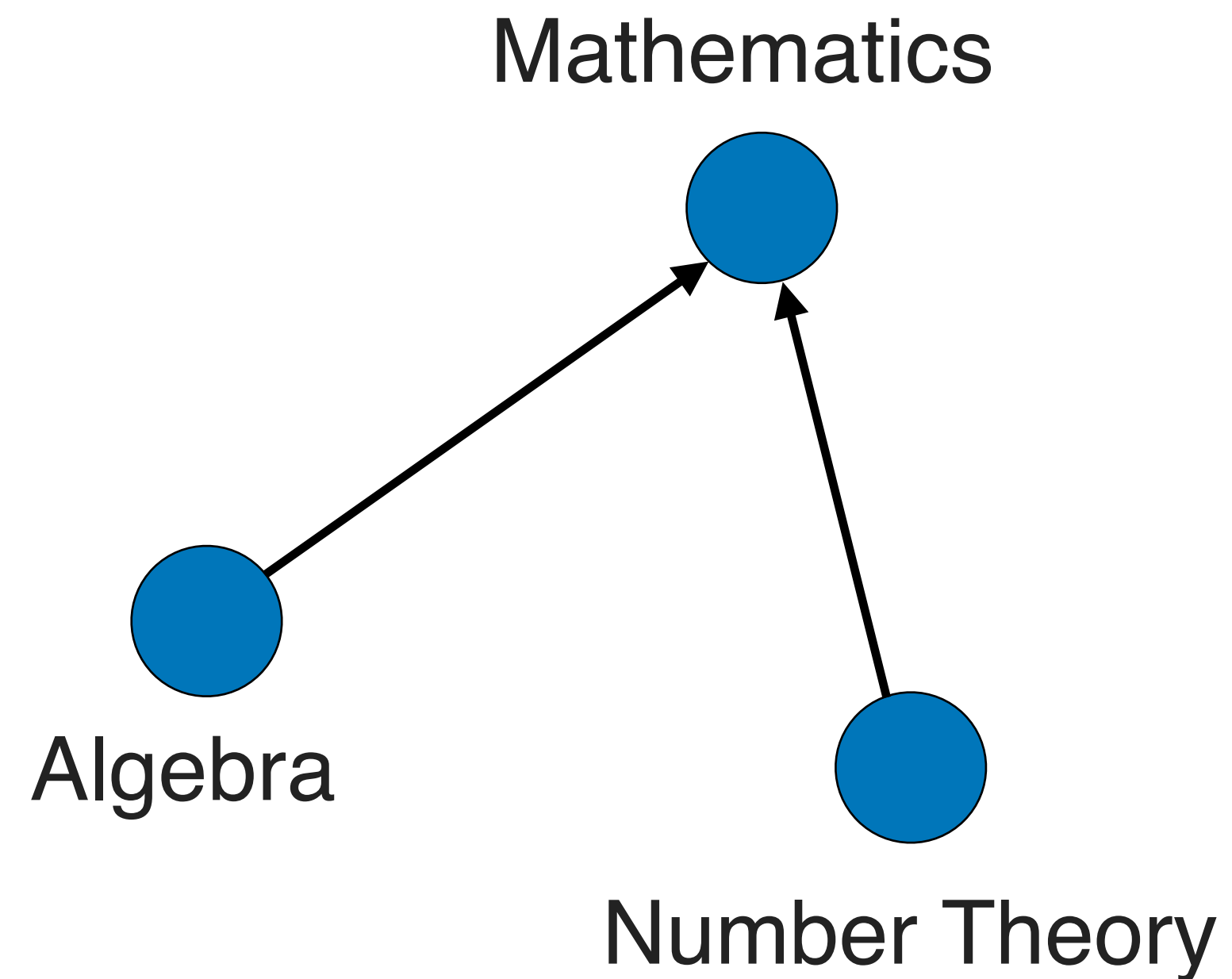
Humans conventionally acquire information through **language**, synthesizing knowledge from a flow of sensory experience (Schieffelin, B. B., & Ochs, E., 1986)

Knowledge and knowledge structure

A knowledge structure is an interrelated collection of facts or knowledge about a particular topic.

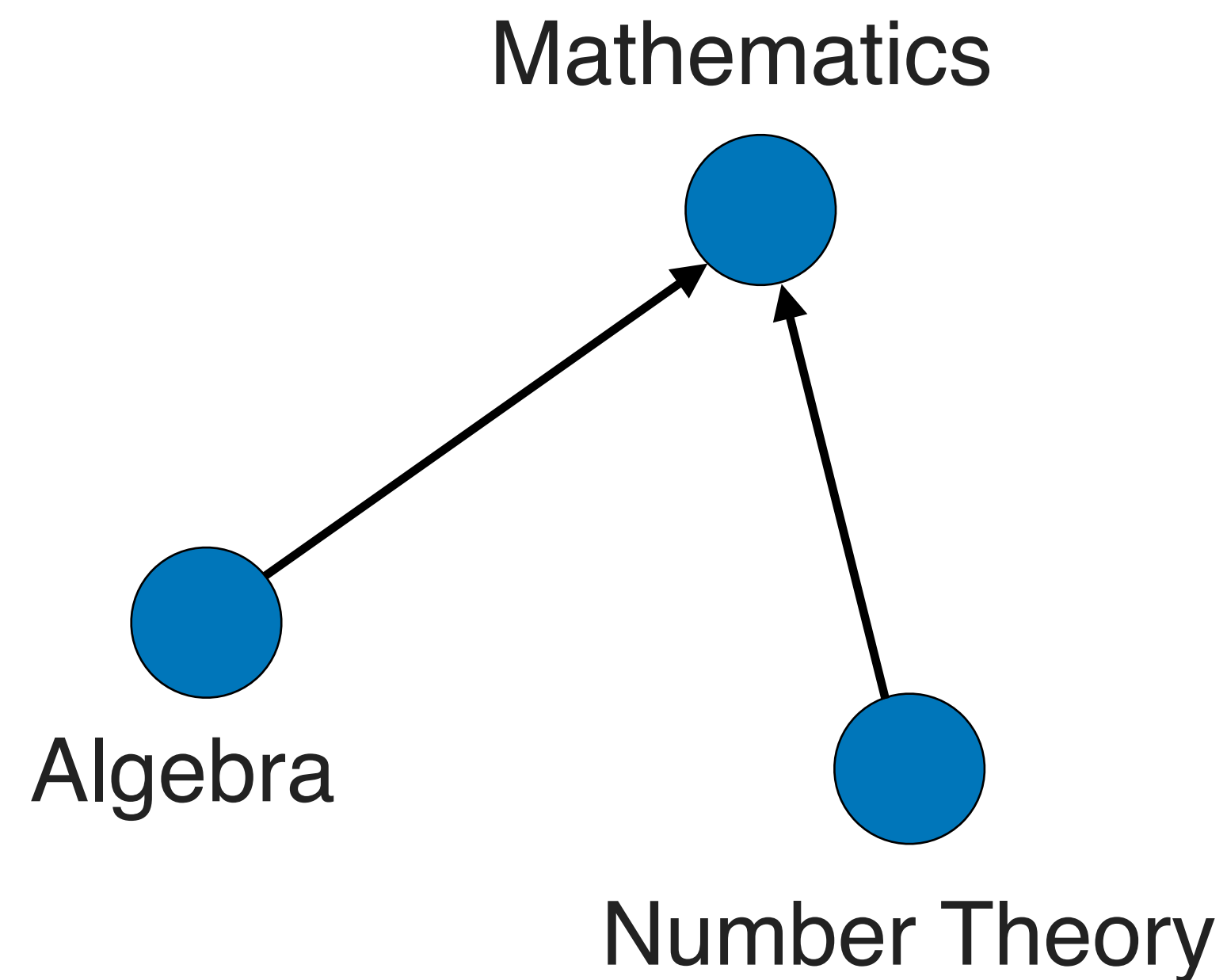
Knowledge and knowledge structure

A knowledge structure is an interrelated collection of facts or knowledge about a particular topic.



Knowledge and knowledge structure

A knowledge structure is an interrelated collection of facts or knowledge about a particular topic.



depths of wikipedia
@depthsofwiki

...

Wikipedia in English defines a crêpe as a thin pancake;
French Wikipedia defines a pancake as a thick crêpe

[Translate Tweet](#)

ake

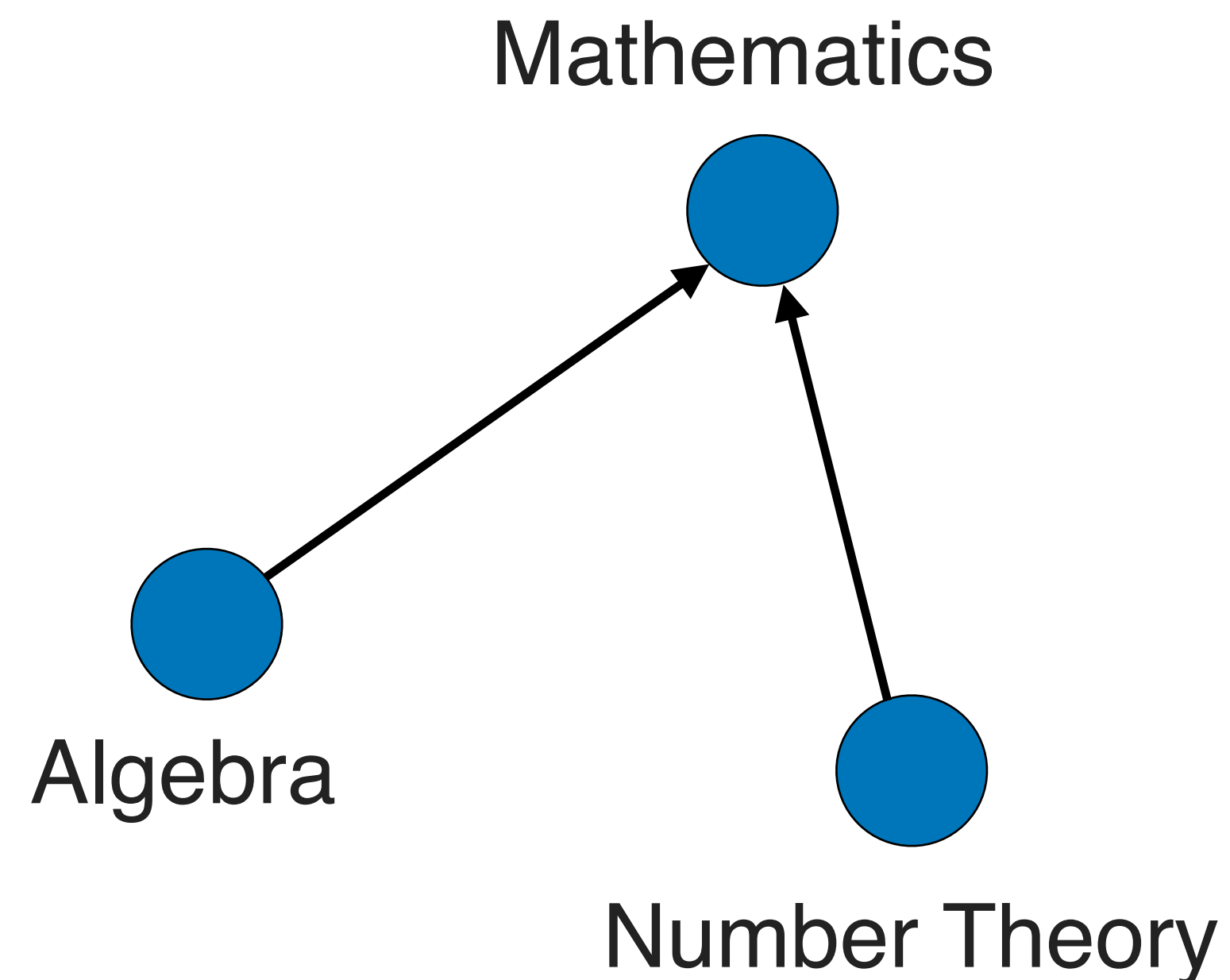
Crêpe

From Wikipedia, the free encyclopedia

cake (/ˈpæn.kɛk/ mot en français « *pancake* », /ˈpæn.kɛk/ A **crêpe** or **crepe** is a thin pancake, typically made of wheat flour, eggs, and milk. It is often served with fruit, jam, or chocolate. In French, it is pronounced /krɛp/ (listen^[1]) or /krɛp/, French: [kʁɛp] (listen), Quebec French: [kʁaɪp] (listen) is a very thin type of pancake.

Knowledge and knowledge structure

A knowledge structure is an interrelated collection of facts or knowledge about a particular topic.



 **depths of wikipedia**
@depthsofwiki

Wikipedia in English defines a crêpe as a thin pancake;
French Wikipedia defines a pancake as a thick crêpe
[Translate Tweet](#)

pancake	Crêpe
<p>pancake (\pan.kɛk\ mot en français « <i>pancake</i> », \pæn.kɛk\ mot en anglais) est un type de crêpe de petit diamètre, servie</p>	<p>From Wikipedia, the free encyclopedia</p> <p>A crêpe or crepe (/kreɪp/ (listen)^[1] or /kreɪp/, French: [kʁɛp] (listen), Quebec French: [kʁaɪp] (listen)) is a very thin type of pancake.</p>

Knowledge structure can be varied by **personalities**, **living country** or **linguistic profile** based on the social structure and education system

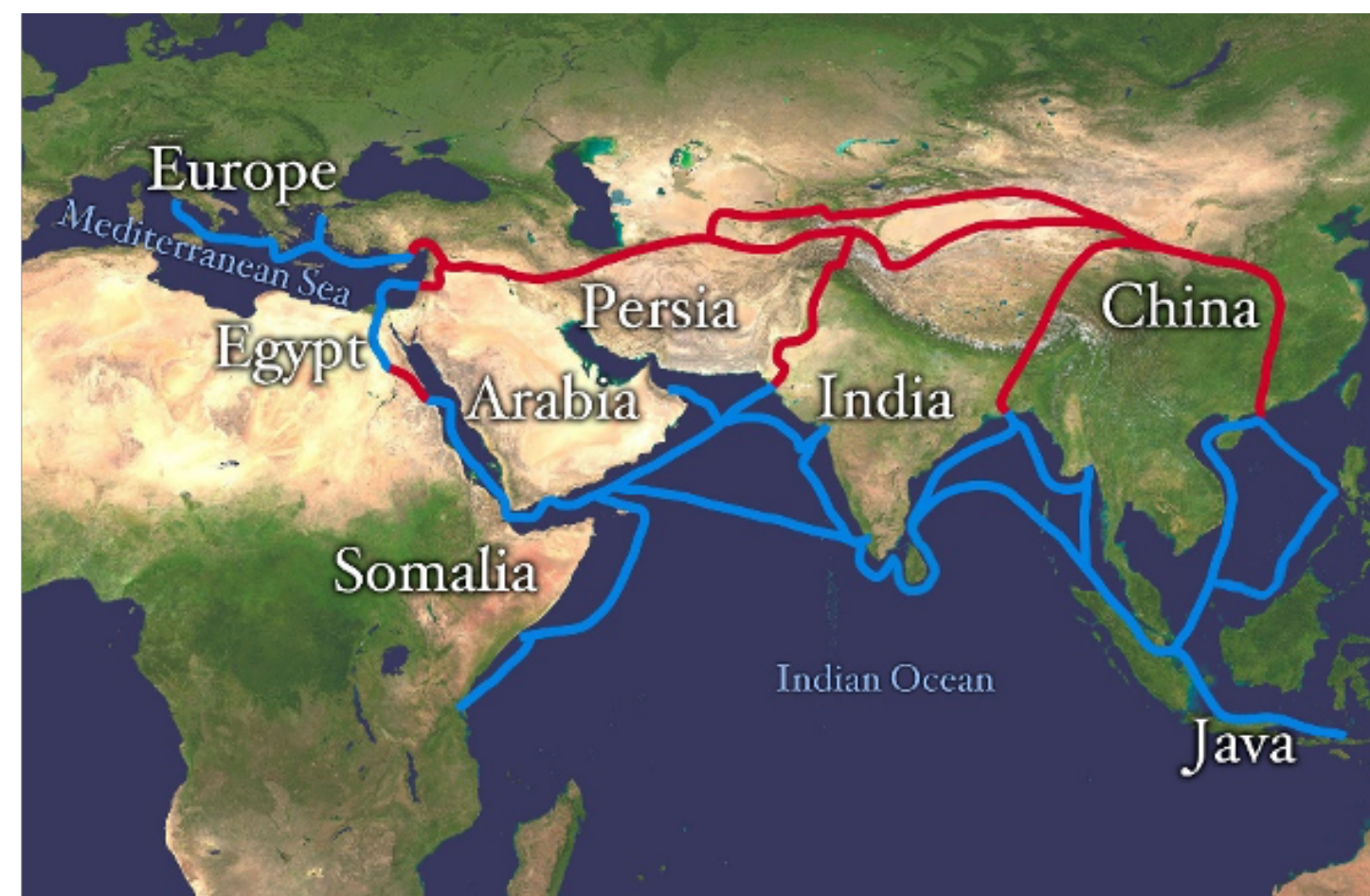
Knowledge flows through routes – knowledge dissemination

Silk Road, The route of knowledge (Andrea, A. J. ,2014)

The constant movement of peoples and goods resulted in an unprecedented transmission and exchange **of knowledge, ideas, beliefs** over three millennia.

21th century silk road (Bhandari, R., & Blumenthal, 2011)

Movement of global student is 21st century silk road



Ref. Wikipedia



Ref. Wikimedia commons

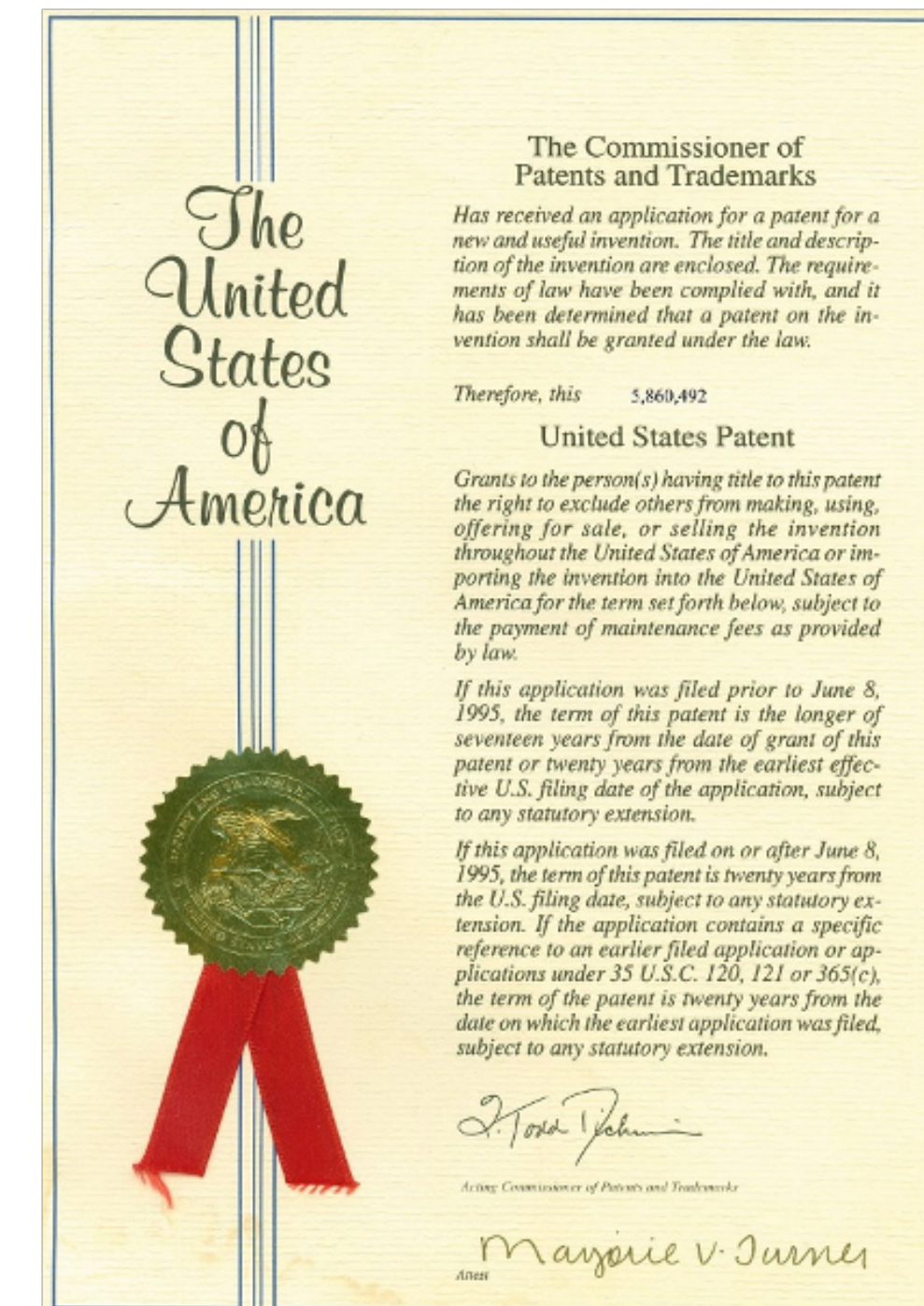
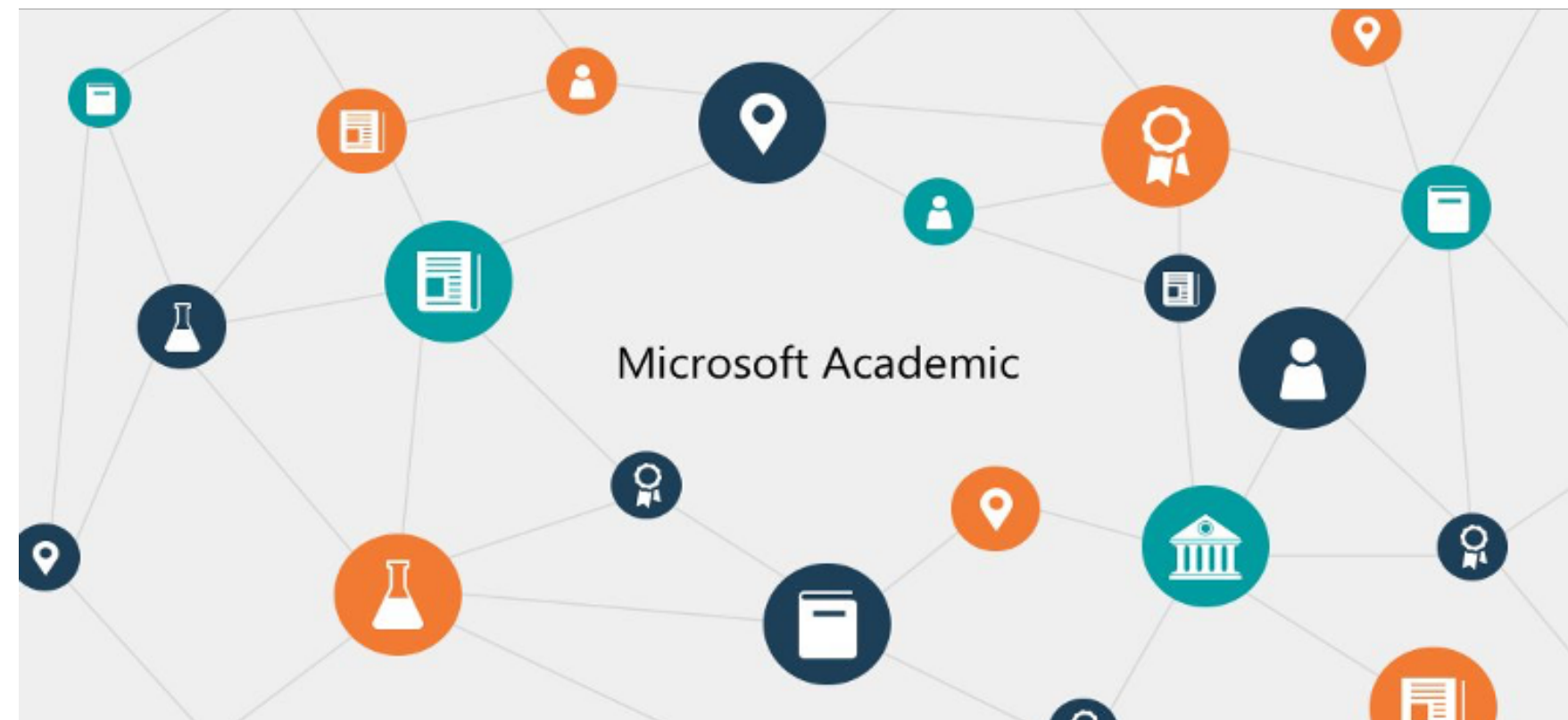
Research Questions

RQ 1: Do social interactions sincerely influence the structure of human knowledge?

RQ 2: If yes, what is the main contemporary channel of information distribution, which could be referred to as a contemporary *Silk Road*?



How to obtain the knowledge structure of specific groups?



Not proper to construct knowledge structure
of a **specific group**

How to obtain the knowledge structure of specific groups?

Wikipedia - Knowledge database emerges from **the collaboration, collective efforts, and competition** of **many individuals** and appears in consensus decision making.

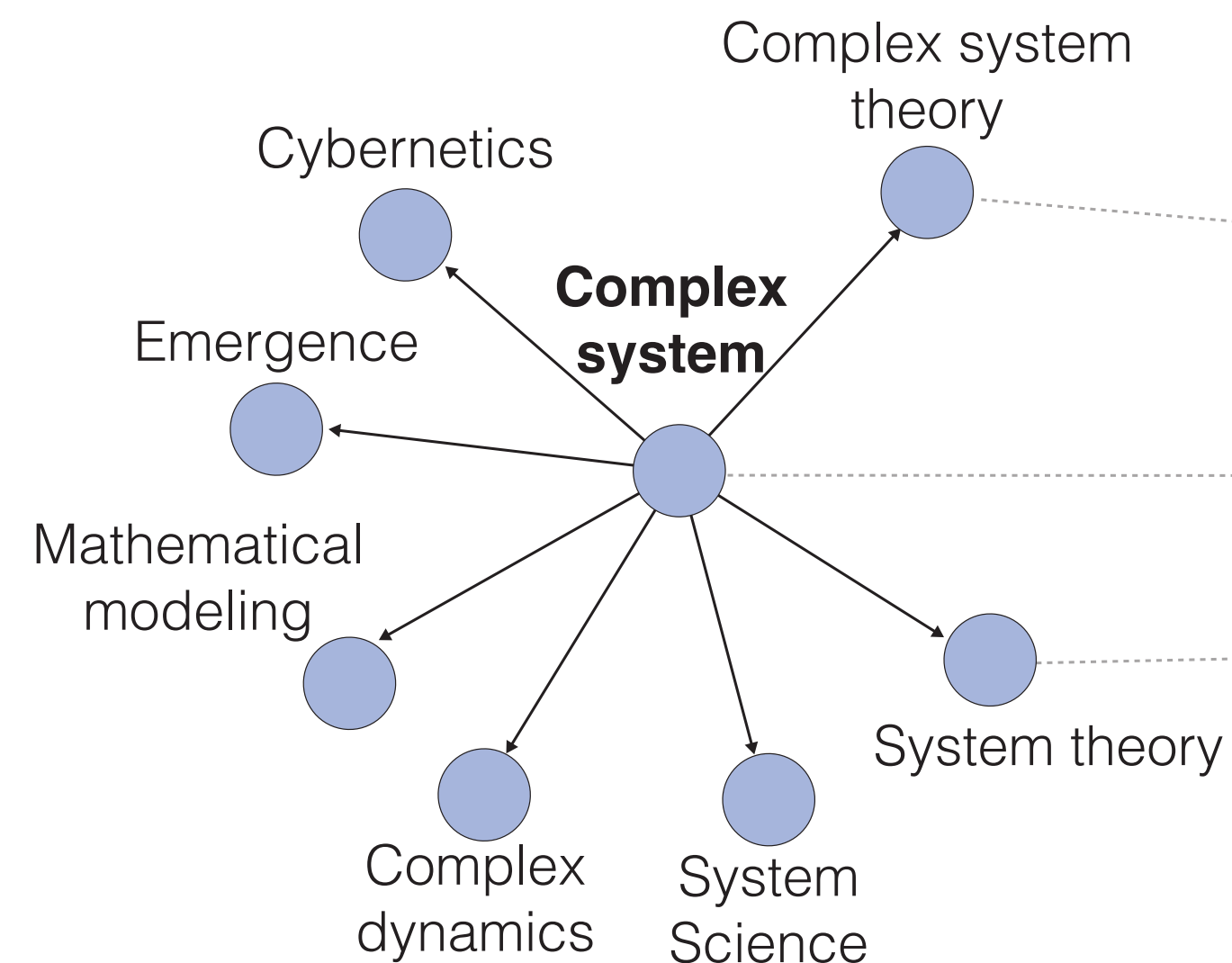


English Wikipedia

Complex system

From Wikipedia, the free encyclopedia

Categories: [Complex dynamics](#) | [Complex systems theory](#) | [Cybernetics](#) | [Emergence](#) | [Systems theory](#) | [Systems science](#) | [Mathematical modeling](#)

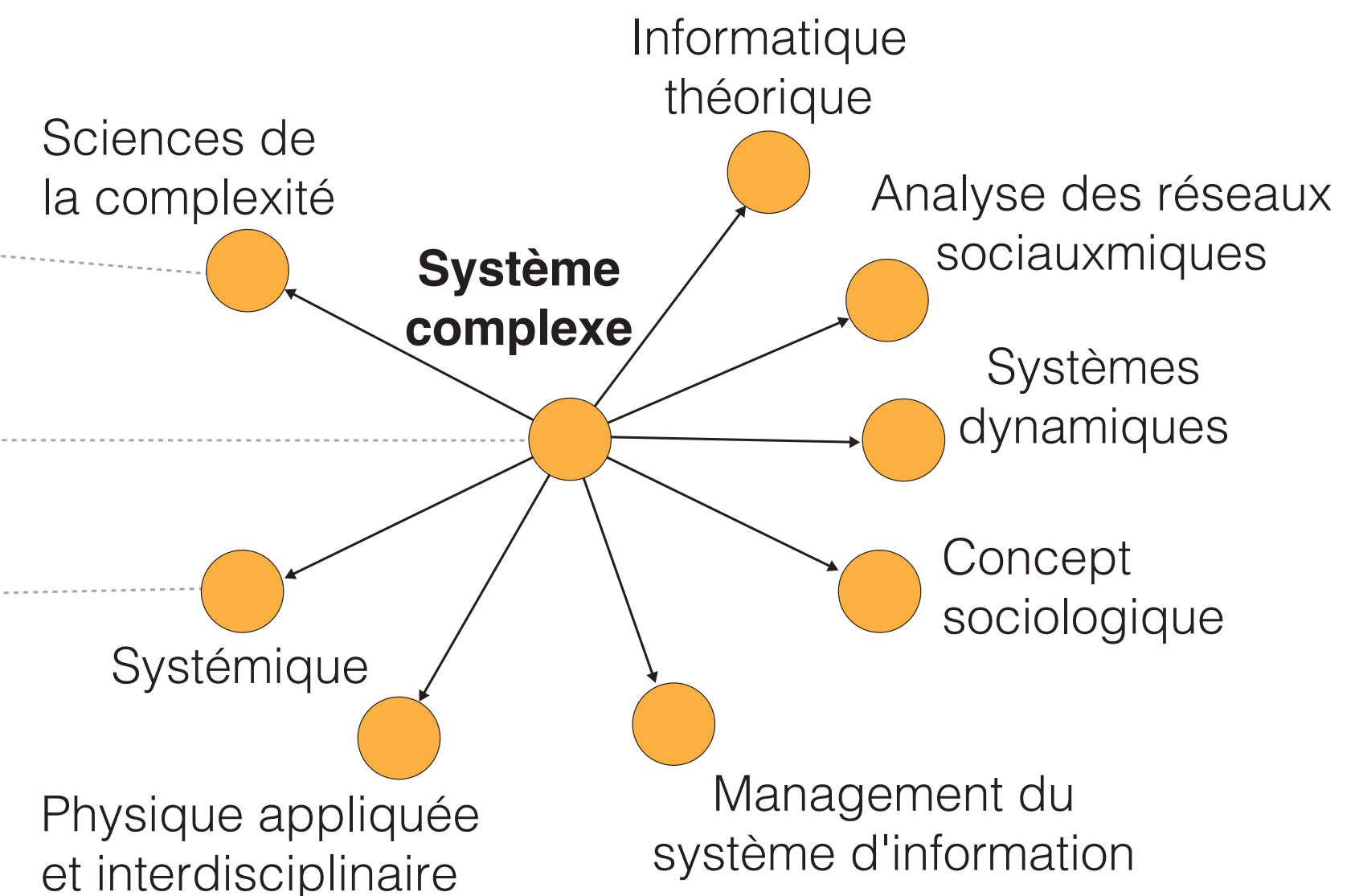


French Wikipedia

Système complexe

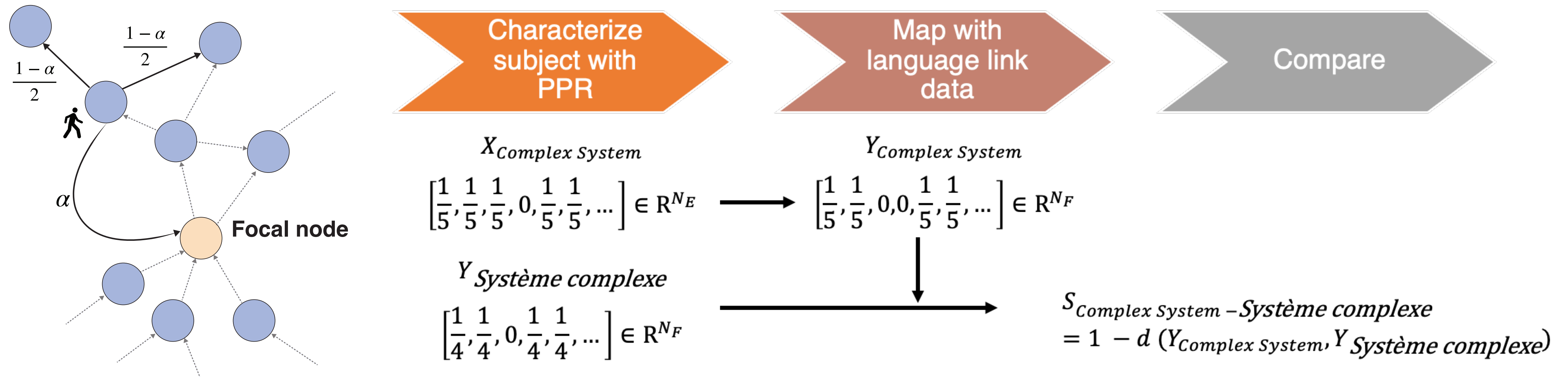
Pour diverses théorie de la complexité, voir [Théorie de la complexité](#) (page d'homonymie).

Catégories : [Sciences de la complexité](#) | [Systémique](#) | [Informatique théorique](#) | [Physique appliquée et interdisciplinaire](#) | [Management du système d'information](#) | [Systèmes dynamiques](#) | [Concept sociologique](#) | [Analyse des réseaux sociaux](#) [+]



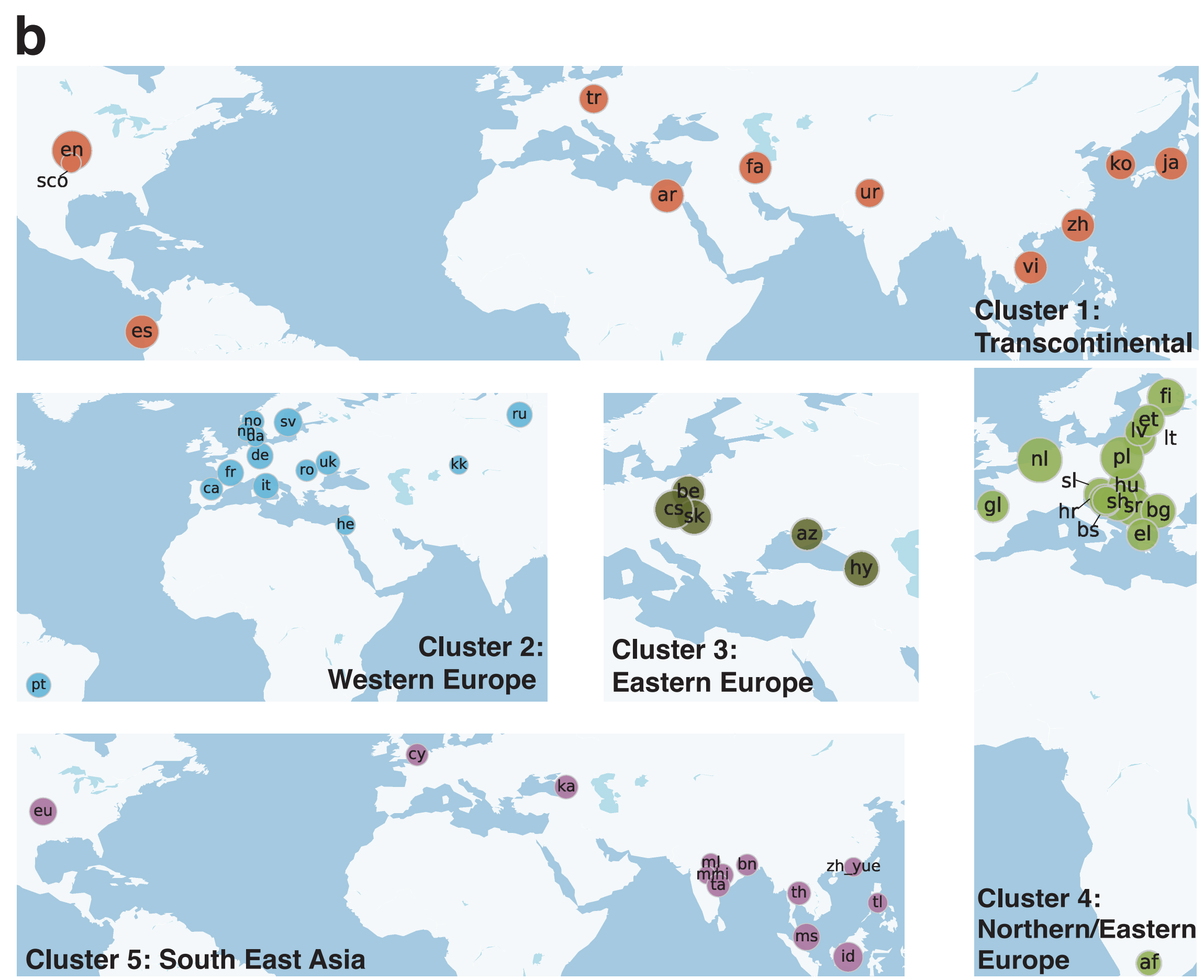
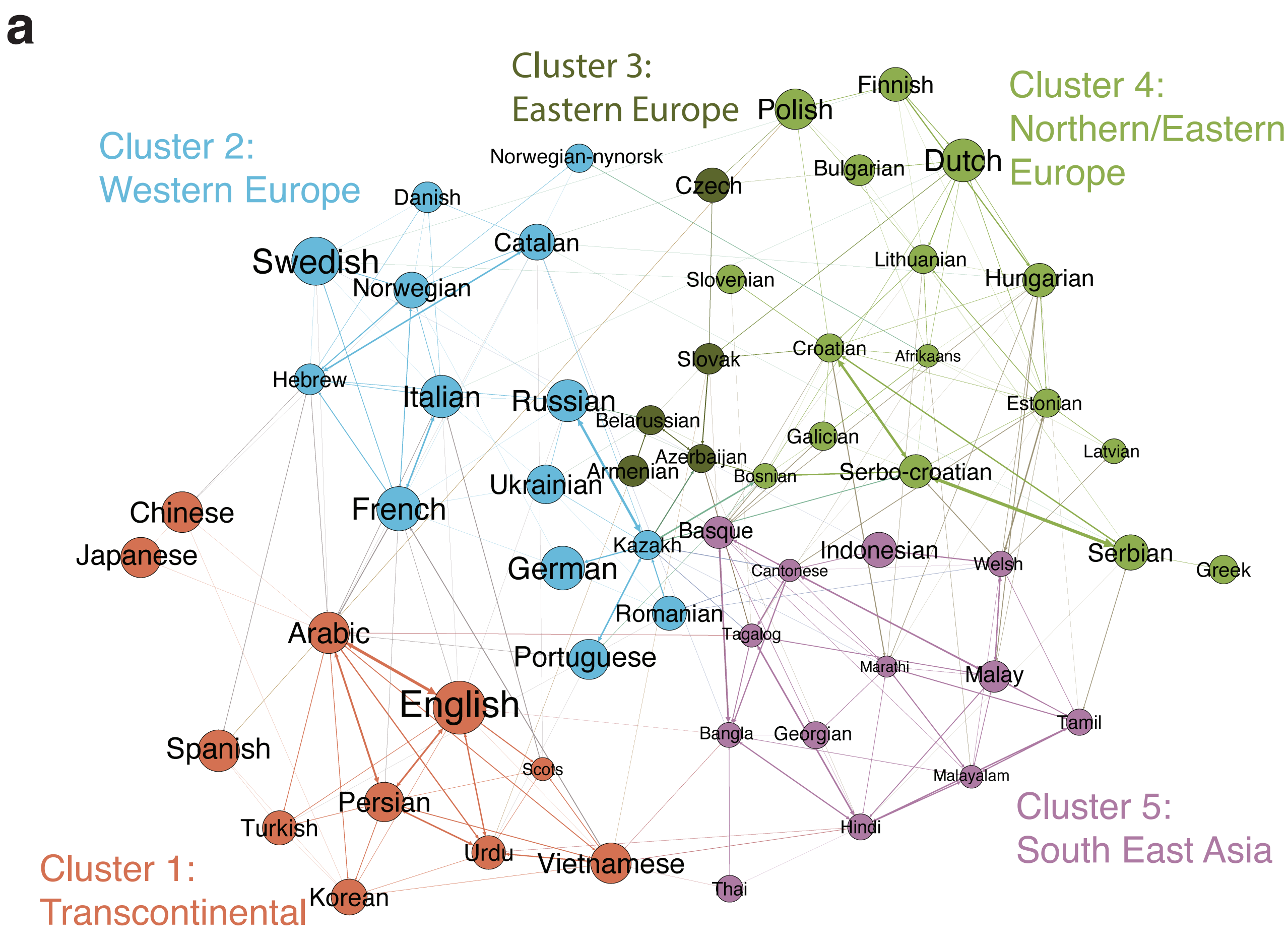
Knowledge structure similarity from genealogy vector of scientific concept

Quantify surrounding contexts of the given subject leveraging network representation



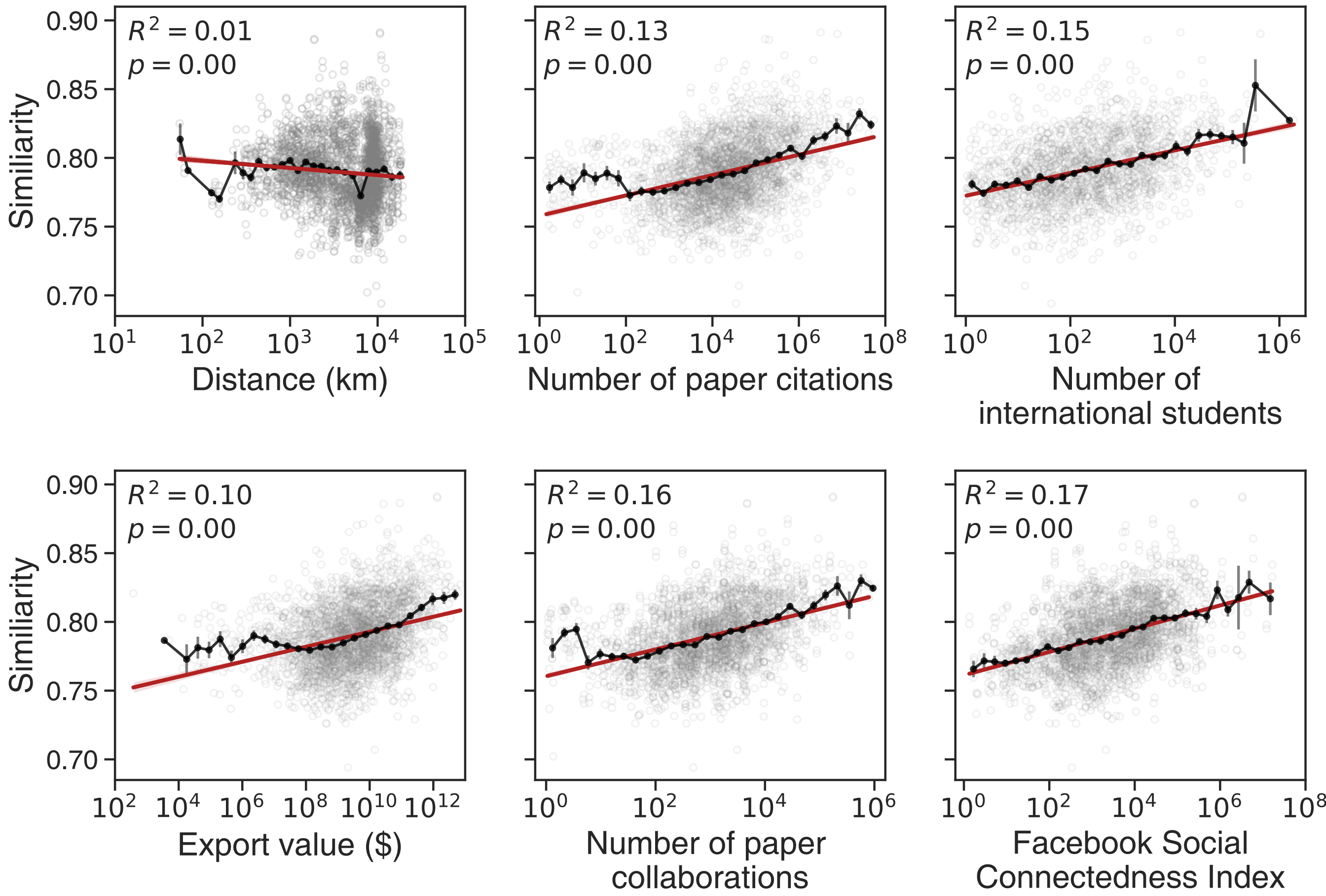
We define the similarity between two knowledge structure as the average value of all subject similarity between two languages

Geographical proximity still influences, but socio-economic interaction shape the knowledge structure



Geographical proximity affects the similarity of knowledge structure across language usage groups.

Geographical proximity still influences, but socio-economic interaction shape the knowledge structure



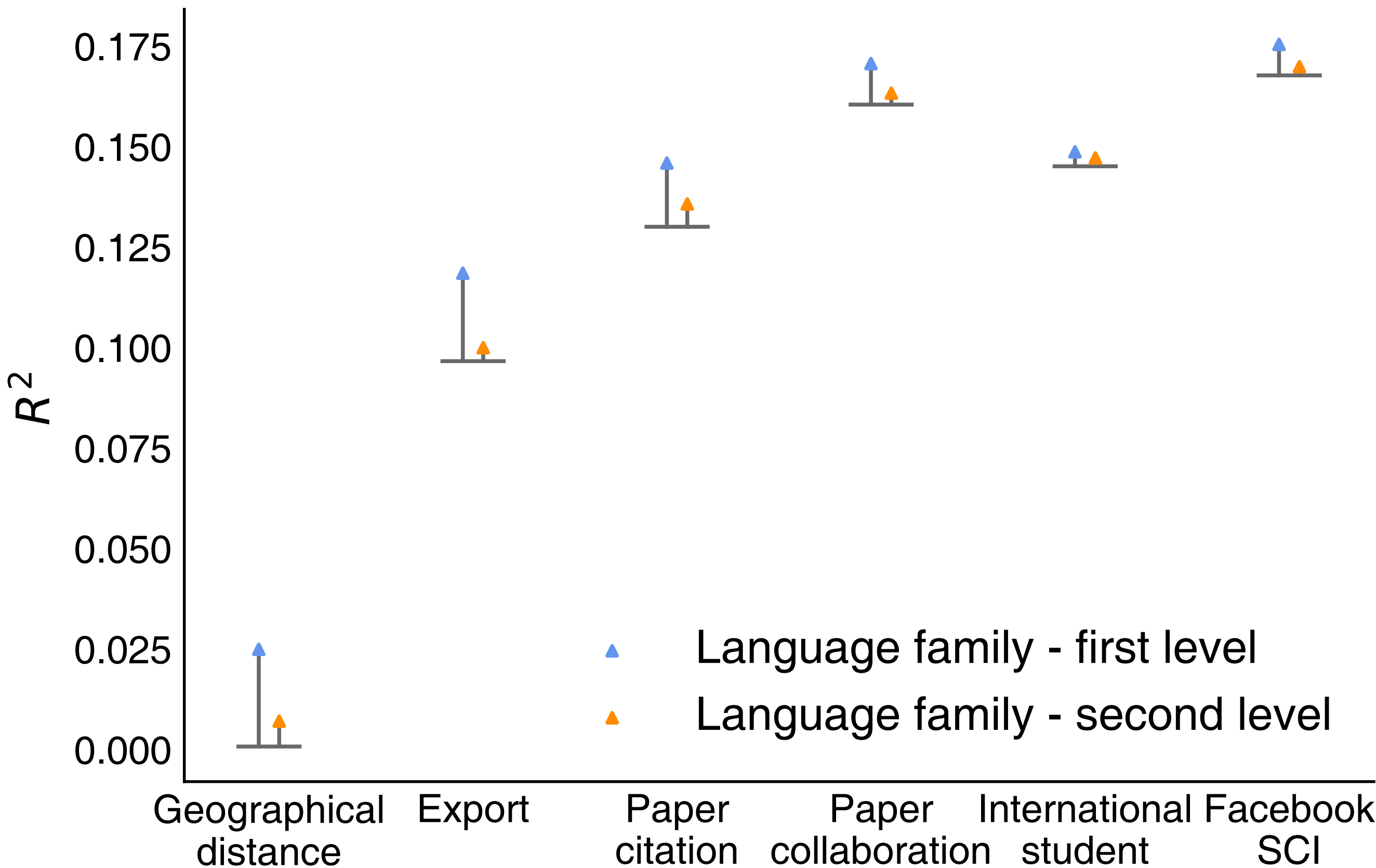
We found degree of association between socio-economic interactions and knowledge structure to occur in the following order:

- Geographical distance
- Export
- Weak knowledge dissemination
- Soft-power movement
- Strong knowledge dissemination
- Mutual friendship on the web

Weak

Strong

Geographical proximity still influences, but socio-economic interaction shape the knowledge structure



**Including language relatedness:
Language Family**

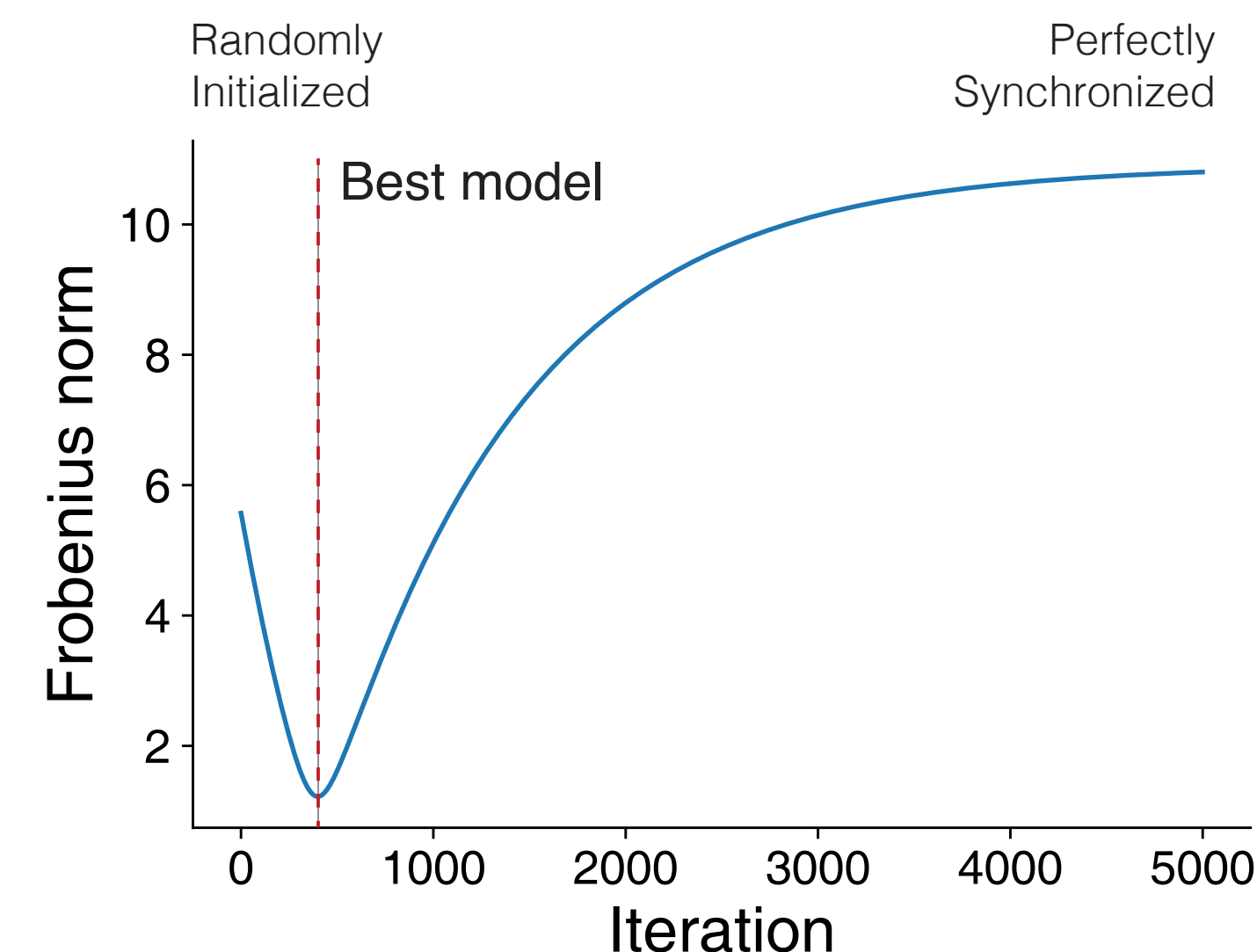
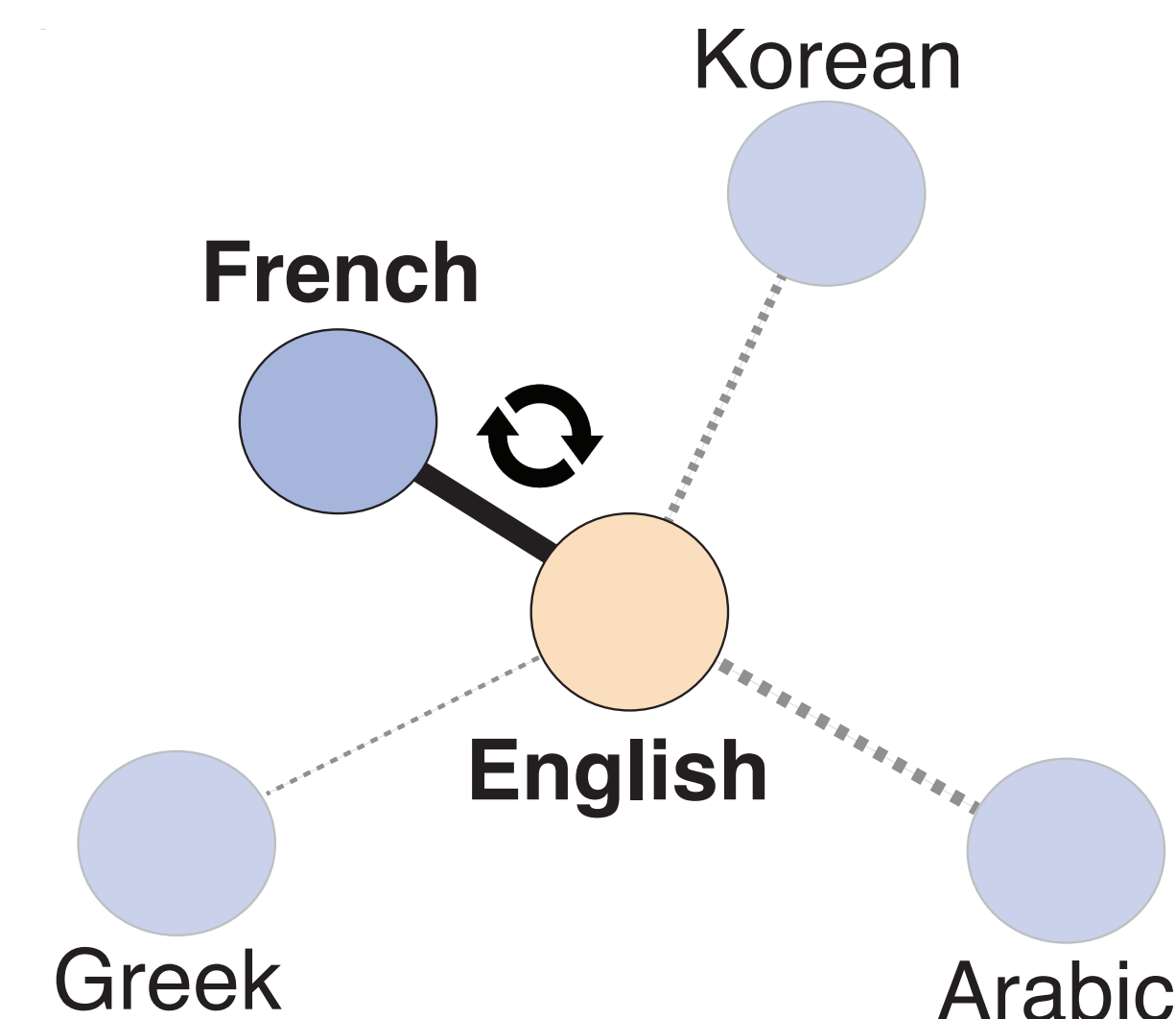
Suggesting international student or Facebook SCI may already encode the majority of language relatedness

Mechanistic model for the knowledge dissemination

Our empirical analysis reveals that

- i) knowledge structures are more likely to be similar if interactions exist between language usage groups
- ii) the degree of association in knowledge structures varies based on the types of interactions

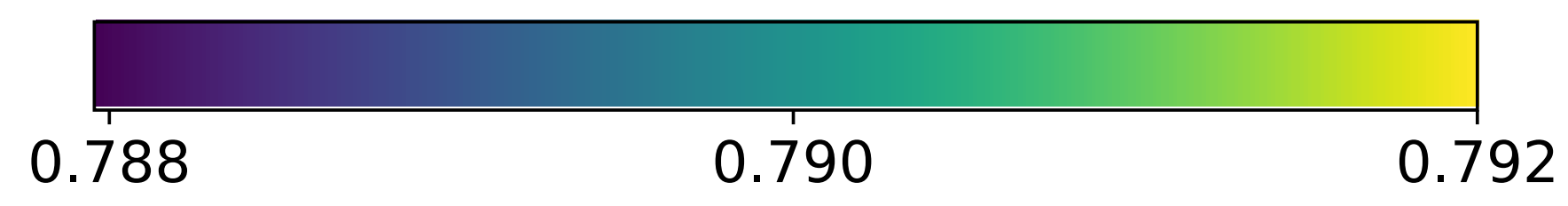
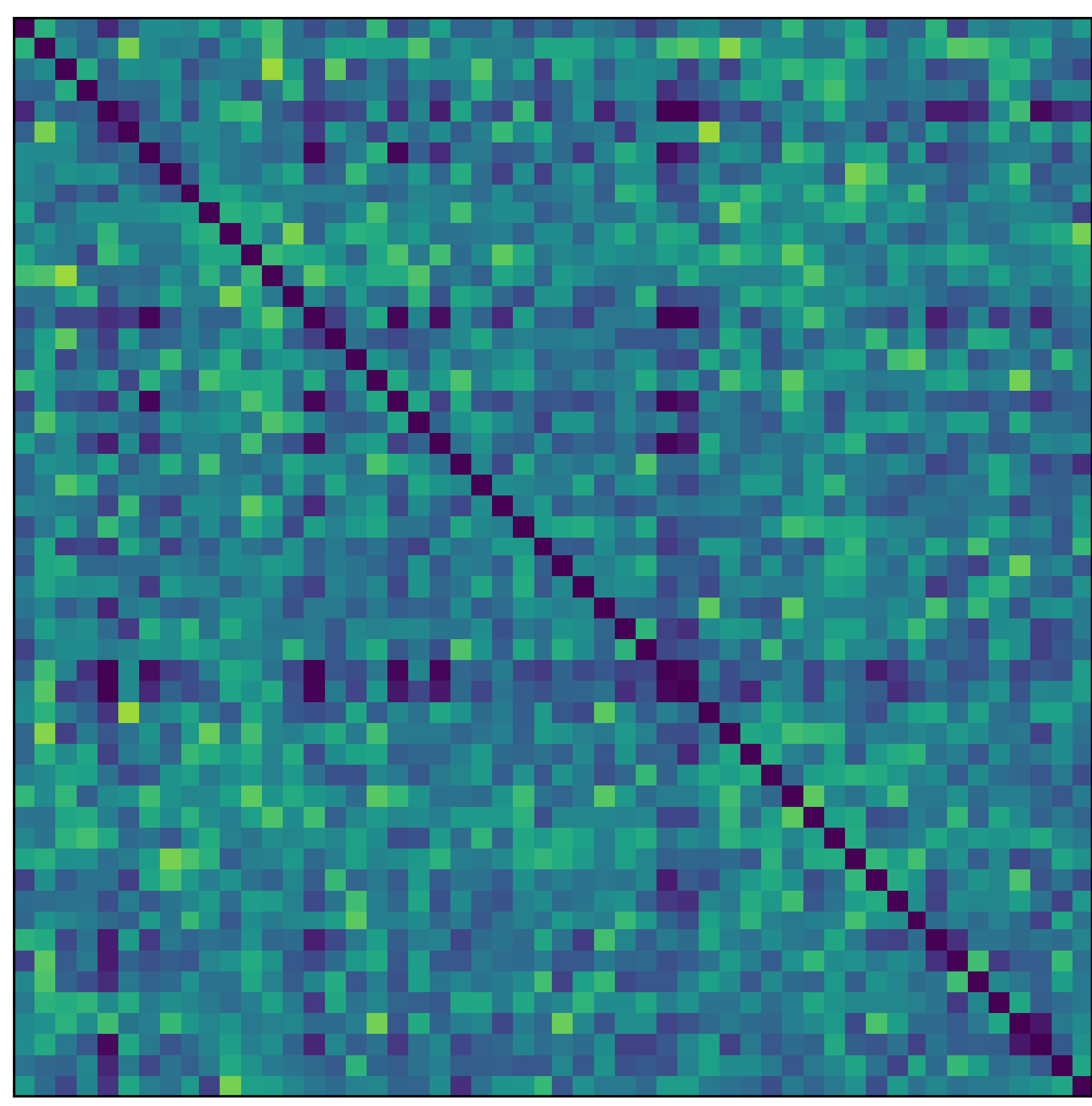
We build the mechanistic model for the knowledge motivated by the Kuramoto model



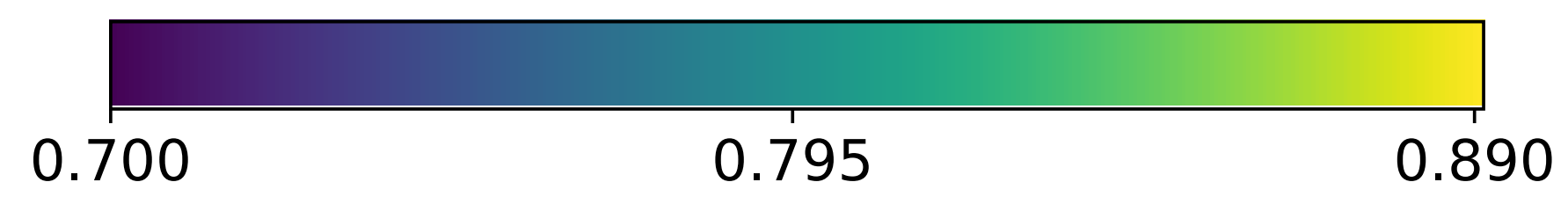
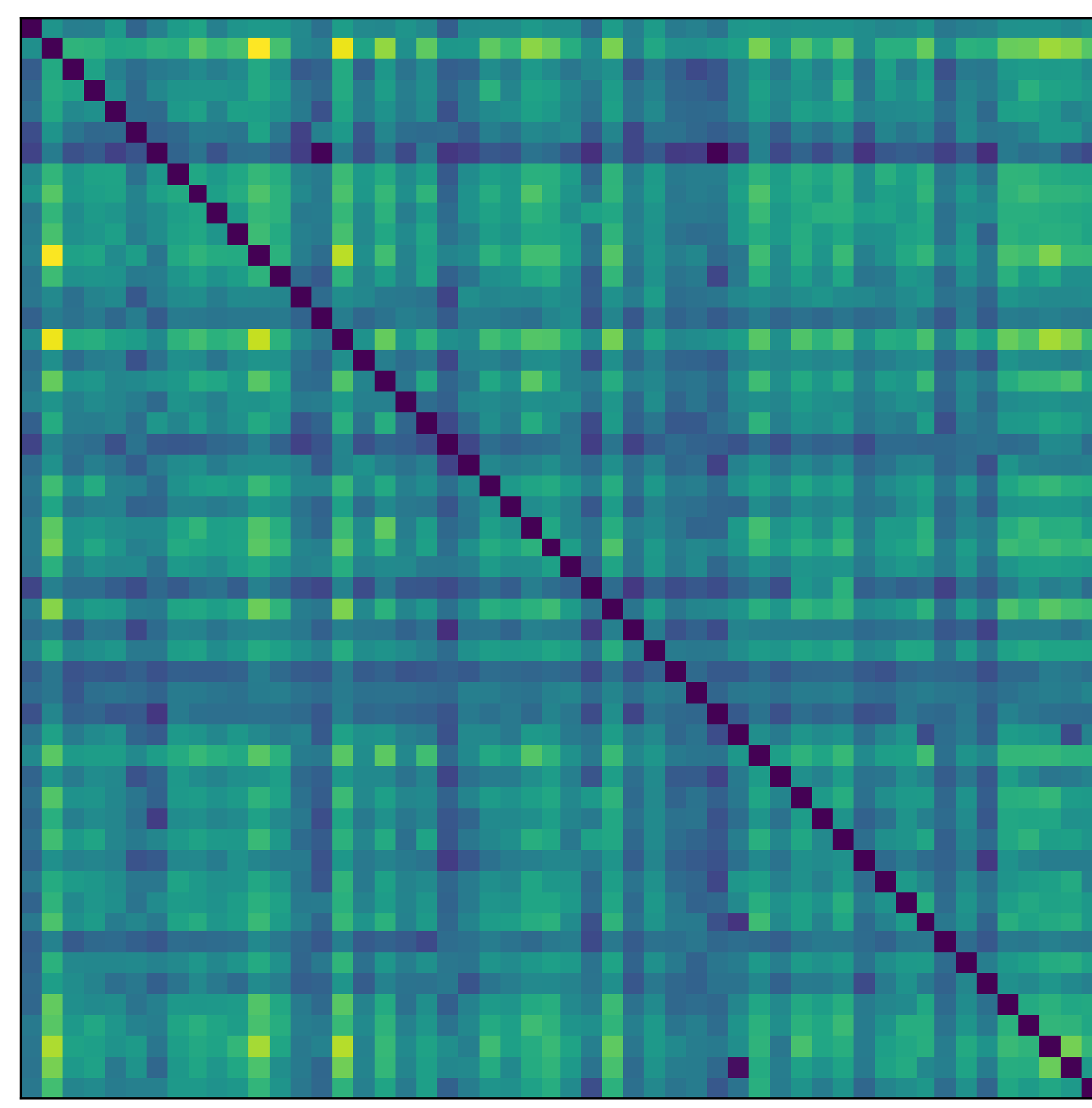
$$S^{model}(t^*) = \operatorname{argmin}_t ||S^{model}(t) - S^{empirical}||_F,$$

Mechanistic model for the knowledge dissemination

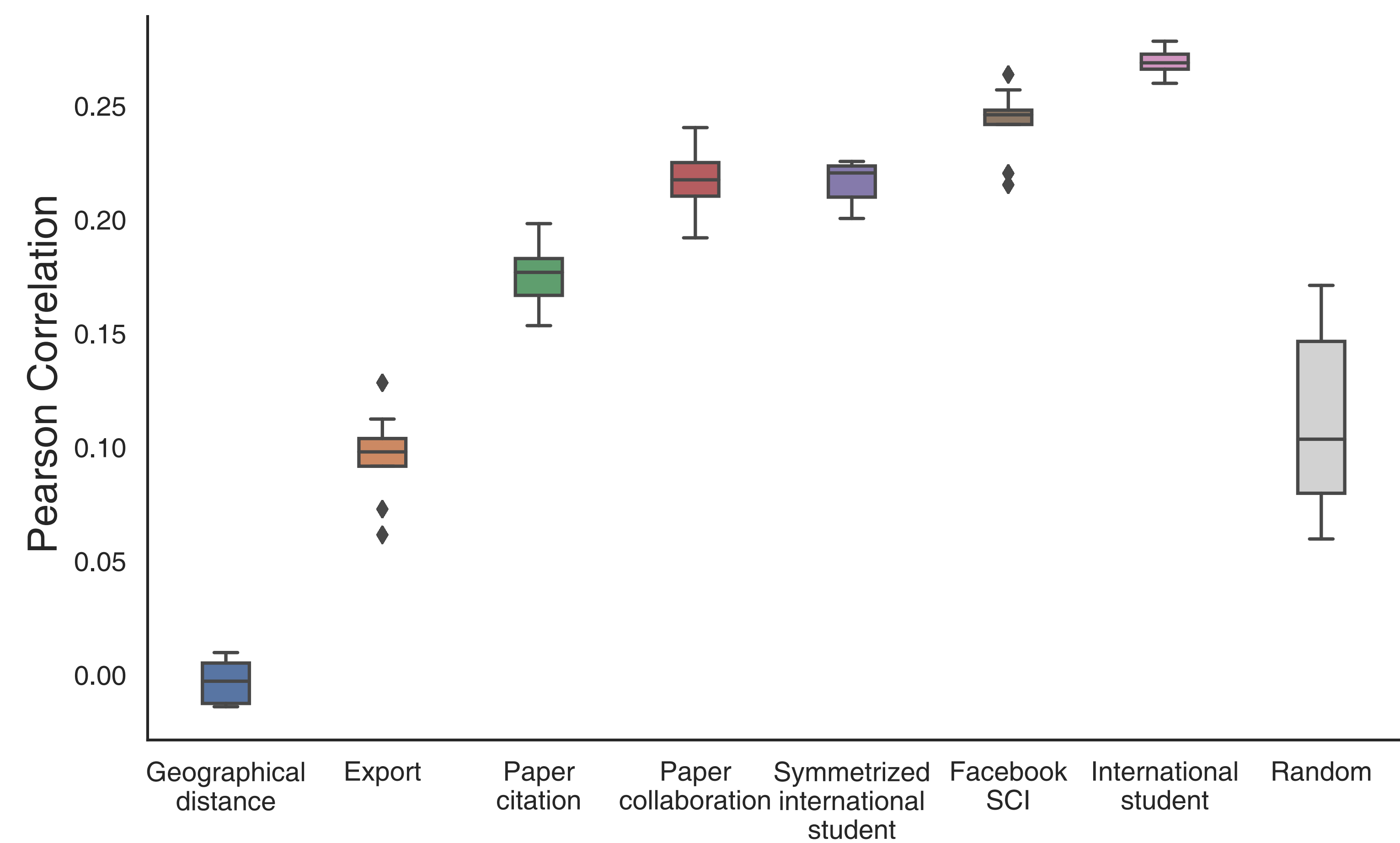
Facebook SCI reconstruction



Knowledge structure simlairty



Mechanistic model for the knowledge dissemination



Our model is the simplest replica, using only a single route of exchange, but implies that **social interaction can shape the structure of human knowledge.**

Summary

We use a multi-lingual Wikipedia linkage to evaluate the similarity of knowledge structure from different language groups.

Utilizing the network representation, we discover a plausible modular structure of languages language usage groups and found that **cooperative scientific research** and **social ties** can explain the synchronization of knowledge structures rather than geographical proximity.

Furthermore, we successfully regenerate the similarity of empirical knowledge structures from various socio-economic ties and uncover **the potential mechanism underlying synchronization of the knowledge structure.**



Collaborators



Jisung Yoon
POSTECH



Jinseo Park
KISTI



Jinhyuk Yun
Soongsil Univ.



Woo-Sung Jung
POSTECH



Preprint: <https://arxiv.org/abs/2202.01466>

Jisung Yoon
@jisungyoon8
jisung.yoon92@gmail.com

**Thanks for
your attention!**

Appendices

Extracting backbone of the network

- Extracting Backbone of similarity network
 - Relative similarity

$$r_{ij} = \frac{\frac{s_{ij}}{\sum_j s_{ij}}}{\frac{\sum_j s_{ij}}{\sum_i \sum_j s_{ij}}} = \sum_i \sum_j s_{ij} * \frac{s_{ij}}{\sum_j s_{ij} * \sum_j s_{ij}}$$

$$r_{ij} = Strength_{total} * \frac{s_{ij}}{strength_{out} \text{ of } i * strngth_{in} \text{ of } j}$$

- Select edges that higher than threshold
 - we select 1.04 which network fully connected to one weakly-connected component

Multivariate analysis

Supporting Table 1: Interrelationship of knowledge structure across language usage groups reveals the impact of socio-economic interactions — multivariate regression analysis

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Facebook SCI	0.0124***	0.0081***	0.0094***	0.0099***	0.0120***	0.0097***
Paper collaboration		0.0068***	−0.0014	−0.0012		−0.0018
Paper citation			0.0067***	0.0094***		0.0091***
Export				−0.0032***		−0.0026***
Geographical distance				−0.0006		0.0006
Language family					0.0050***	0.0048***
Intercept	0.6604***	0.6550***	0.6460***	0.6645***	0.6600***	0.6560***
N	2,572	2,572	2,572	2,572	2,572	2,572
AIC	-10604.65	-10654.71	-10659.45	-10666.44	-10617.97	-10678.05
R ²	0.169	0.185	0.188	0.191	0.174	0.195
Adjusted R ²	0.168	0.185	0.189	0.189	0.173	0.193

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

- Model 1** Knowledge similarities ~ Facebook SCI
- Model 2** Knowledge similarities ~ Facebook SCI + Paper collaboration
- Model 3** Knowledge similarities ~ Facebook SCI + Paper collaboration + Paper citation
- Model 4** Knowledge similarities ~ Facebook SCI + Paper collaboration + Paper citation + Export + Geographical distance
- Model 5** Knowledge similarities ~ Facebook SCI + Language family
- Model 6** Knowledge similarities ~ Facebook SCI + Paper collaboration + Paper citation + Export + Geographical distance + Language family

Multivariate analysis

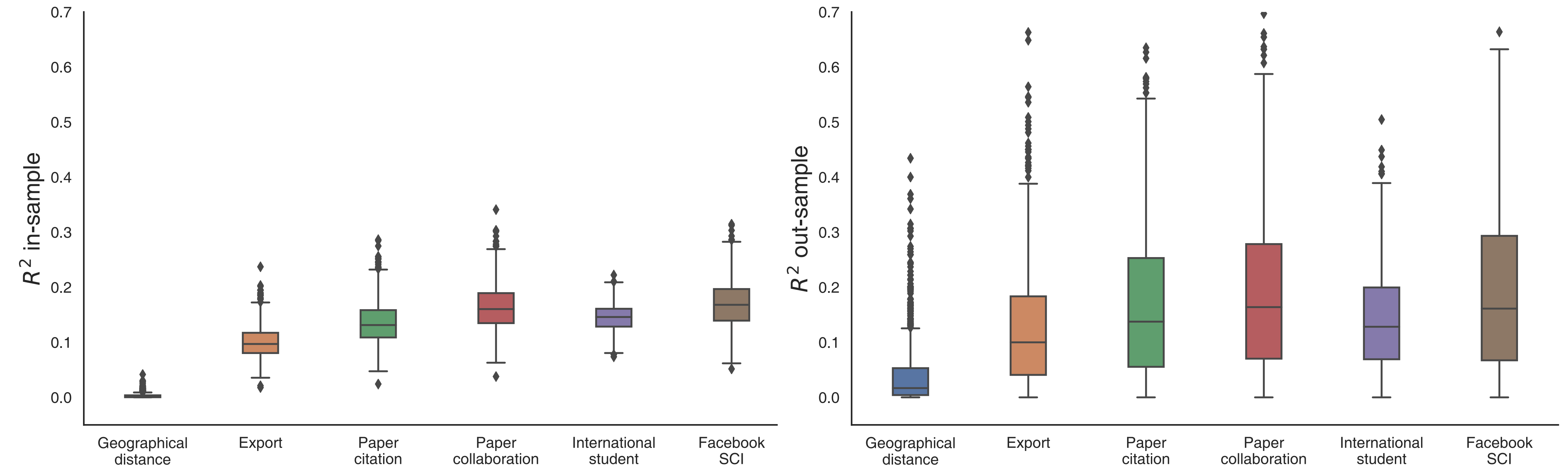
Supporting Table 1: Interrelationship of knowledge structure across language usage groups reveals the impact of socio-economic interactions — multivariate regression analysis

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Facebook SCI	0.0124***	0.0081***	0.0094***	0.0099***	0.0120***	0.0097***
Paper collaboration		0.0068***	−0.0014	−0.0012		−0.0018
Paper citation			0.0067***	0.0094***		0.0091***
Export				−0.0032***		−0.0026***
Geographical distance				−0.0006		0.0006
Language family					0.0050***	0.0048***
Intercept	0.6604***	0.6550***	0.6460***	0.6645***	0.6600***	0.6560***
N	2,572	2,572	2,572	2,572	2,572	2,572
AIC	-10604.65	-10654.71	-10659.45	-10666.44	-10617.97	-10678.05
R ²	0.169	0.185	0.188	0.191	0.174	0.195
Adjusted R ²	0.168	0.185	0.189	0.189	0.173	0.193

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

- Model 1** Knowledge similarities ~ Facebook SCI
- Model 2** Knowledge similarities ~ Facebook SCI + Paper collaboration
- Model 3** Knowledge similarities ~ Facebook SCI + Paper collaboration + Paper citation
- Model 4** Knowledge similarities ~ Facebook SCI + Paper collaboration + Paper citation + Export + Geographical distance
- Model 5** Knowledge similarities ~ Facebook SCI + Language family
- Model 6** Knowledge similarities ~ Facebook SCI + Paper collaboration + Paper citation + Export + Geographical distance + Language family

Out-sample prediction



To demonstrate that the associations depicted in Fig. 3 are genuine and not the result of correlations between observed points, we train the model with a randomly chosen subset of the languages (in-sample) and calculate the R^2 on the remaining languages (out-sample).

Mapping the country level statistics onto the language

- Basically, socio-economic data are county to country data.
- For our analysis, we develop a method that projects county to country data to language to language data.
- Language projection method
 - $Y_{l \rightarrow L} = A_{L \rightarrow C}^T * X_{C \rightarrow C} * A_{L \rightarrow C}$, Language projected data
 - $X_{C \rightarrow C} \in R^{N_C * N_C}$, Country to country socio-economic data
 - $A_{L \rightarrow C} \in R^{N_C * N_L}$, Country to language matching matrix(Ronen et al, 2014)
 - e.g.) South Korea → 100% Korean
 - e.g.) United States → 82.1% English, 10.7% Spanish