



The PhD Voice

@PhDVoice



Definition of "PhD student":

A type of person who turns coffee into research papers.

Does distance impede academic discourse?

Dr. Aneeq Sarwar

Background

- Accumulation of knowledge = better quality of life than in the past
- Codification and sharing of knowledge = cornerstone of industrial progress
- Citations make invisible knowledge flows visible and acknowledge intellectual debt (pricing for academic knowledge?)
- If so, given the universal access to the internet – citations should be 'placeless'
- Evidence for both positions

Background

- If citations are not placeless
- Geography can grant/exclude access to knowledge
- If knowledge is excludable – is it really a public good?

Limitations of existing studies

- Studies examining the economic geography of citation activity are typically about patent citations and not academic citation activity
- Research examining academic citation activity tends to focus on conventional fields (e.g., Medicine)
- Such research tends to be more qualitative, and case based (conducted more frequently by sociologists)
- Bibliographic coupling to measure knowledge relatedness

Concerns with using Bibliographic Coupling

If;

$$x_{ij} = \beta_1 \text{Distance}_{ij} + \beta_2 \text{Bibliographic Coupling}_{ij} \quad (\text{i})$$

x = decision to cite,

And;

$$\text{Bibliographic Coupling}_{ij} = \beta_1 \text{Distance}_{ij} + \beta_n C_{ij} \quad (\text{ii})$$

C = any other variable

Then: Bibliographic Coupling is not entirely endogenous

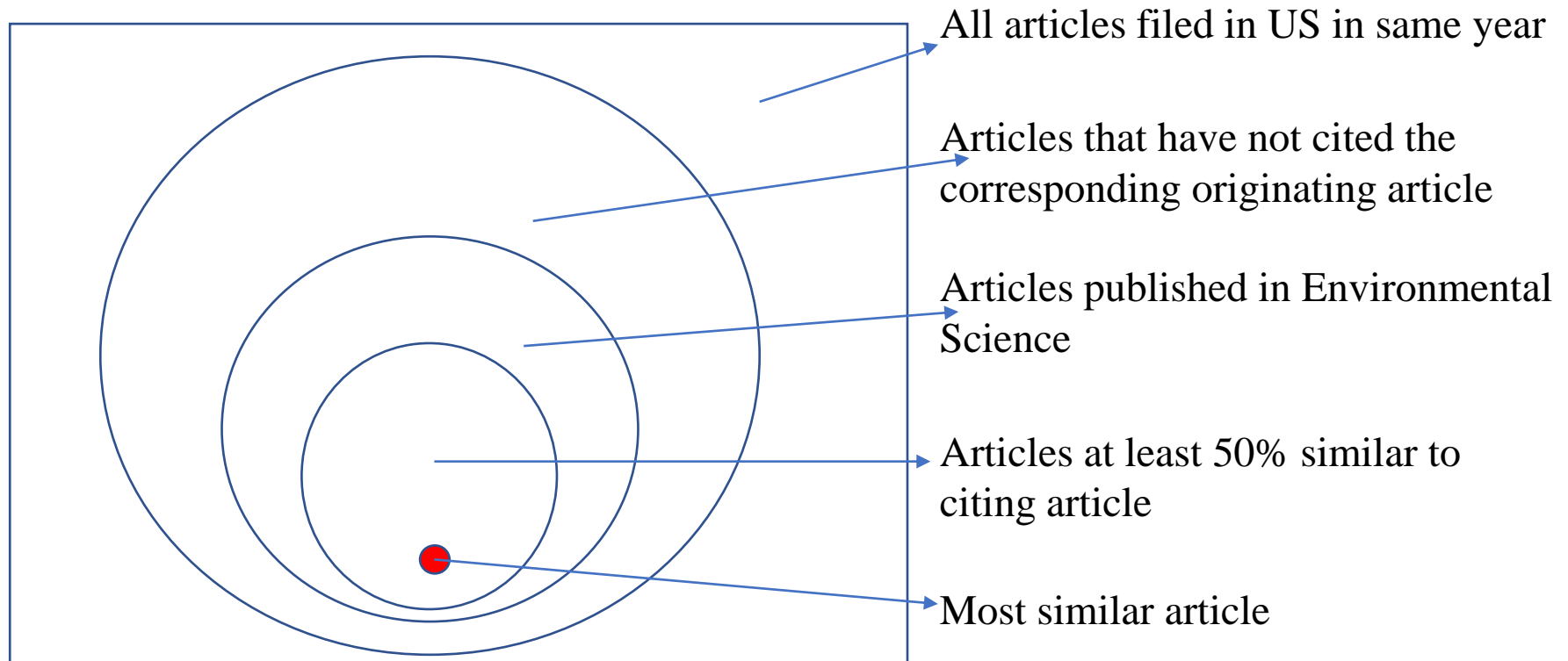
Concerns with using Bibliographic Coupling

- Will lose meaning over time
- No standardization of how references are selected in journal publications
- Not all references are equal (methods, theory, negative)
- Can not account for chain citations
- More recent fields will have limited publications and thus limited applicability

What do I bring to the debate?

- Measuring knowledge similarity
- Controlling for the distribution of academic activity
- Exploring climate change science

Control Selection Technique



Equation(s)

$$Y_{ij} = \beta_1 D_{ij} + \beta_2 L_{ij} + \beta_3 P_{ij} + \beta_4 B_{ij} + \varepsilon_{ij} \quad (1)$$

$$Y_{ij} = \beta_1 D_{ij} + \beta_2 L_{ij} + \beta_3 P_{ij} + \beta_4 S_{ij} + \varepsilon_{ij} \quad (2)$$

$$Y_{ij} = \beta_1 D_{ij} + \beta_2 L_{ij} + \beta_3 P_{ij} + \beta_4 B_{ij} + \beta_5 S_{ij} + \varepsilon_{ij} \quad (3)$$

$$Y_{ij} = \beta_1 D_{ij} + \beta_2 L_{ij} + \beta_3 P_{ij} + \beta_4 B_{ij} + \beta_5 S_{ij} + \beta_6 G_j + \varepsilon_{ij} \quad (4)$$

j : originating article

i : citing article

Y_{ij} : Dichotomous variable indicating presence of citation link

D_{ij} : Mean of distance between all possible author combination (based on city-to-city distance)

L_{ij} : Time difference between publication of article(s) ' j ' and ' i '

P_{ij} : Variable to indicate if both articles are published in same journal

B_{ij} : Bibliographic coupling variable based on the specification by Wuestman et al., 2019

S_{ij} : Measure of similarity between both articles based on cosine similarity

G_j : Measure of generality of originating article

ε_{ij} : Error term

Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Cited_{ij}	32,270	0.508	0.499	0	1
Distance_{ij} (D)	32,270	3180.52	2677	0	18961.9
Lag_{ij} (L)	32,270	5.772	3.417	0	14
Publication_{ij} (P)	32,270	0.204	0.403	0	1
Bibliographic Coupling_{ij} (B)	32,270	0.001	0.002	0	0.046
Similarity_{ij} (S)	32,270	0.279	0.167	0	0.886
Generality_j (G)	32,270	11.344	13.137	-1.15	107.136

Equation	0	1	2	3	4
Model	OLS	OLS	OLS	OLS	OLS
D _{ij}	-0.033***	-0.024***	-0.032***	-0.024***	-0.024***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
L _{ij}	0.004	0.030***	0.010**	0.031***	0.031***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
P _{ij}	0.038***	0.025***	0.033***	0.024***	0.024***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
B _{ij}	-	0.102***	-	0.096***	0.096***
		(0.005)		(0.005)	(0.005)
S _{ij}	-	-	0.046***	0.017***	0.017***
			(0.002)	(0.003)	(0.003)
G _j	-	-	-	-	-0.002
	-	-	-	-	(0.002)
Obs	32,270	32,270	32,270	32,270	32,270
Cluster	Yes	Yes	Yes	Yes	Yes
Std Err					

Note: : *p < 0.05, **p < 0.01, ***p < 0.001. (): robust standard errors.

Findings

- The importance of bibliographic coupling over similarity demonstrates support for the social constructivist school of thought (i.e. tools of persuasion based on perceived social hierarchies)
- Trade-off between distance and similarity

Findings

- Geography can mean significant advantages for researchers by granting access to knowledge (then geography can also exclude)
- Is knowledge a semi-public good (because it is excludable?)
- If knowledge is excludable it is also monetizable – do we still need IPRs?
- If geography affects citations – how reliable are h-indices?

Are my results generalisable?



“...it would be nice to just observe ALL of the people...”

Equation	0	1	2	3
Model	OLS	OLS	OLS	OLS
D _{ij}	-0.042***	-0.022***	-0.036***	-0.021***
	(0.002)	(0.002)	(0.002)	(0.002)
L _{ij}	0.004	0.024***	0.009**	0.025***
	(0.003)	(0.002)	(0.003)	(0.002)
P _{ij}	0.037***	-0.008***	0.024***	-0.010***
	(0.002)	(0.002)	(0.002)	(0.002)
B _{ij}	-	0.056***	-	0.053***
		(0.004)		(0.004)
S _{ij}	-	-	0.096***	0.015***
			(0.002)	(0.002)
G _j	-	-	-	-
	-	-	-	-
Obs	168,820	168,820	168,820	168,820
Cluster	Yes	Yes	Yes	Yes
Std Err				

Conclusion

- Re-evaluate performance and output based evaluation and reward mechanisms
- Location matters for start-ups
- Broader research teams?
- Knowledge is a semi-public good
- Tap into localised knowledge pool, particularly ones overseas
- More collaborations and discourse



Thank you

Please ask questions!

If there are any
comments/suggestions I am
happy to discuss them by
email at
asarwar@swin.edu.au

Appendix

Correlation Matrix

	C_{ij}	D_{ij}	L_{ij}	P_{ij}	B_{ij}	S_{ij}	G_j
Cited_{ij} (C)	1						
Distance_{ij} (D)	-0.069	1					
Lag_{ij} (L)	-0.005	0.064	1				
Publication_{ij} (P)	0.077	-0.031	-0.133	1			
Bibliographic Coupling_{ij} (B)	0.202	-0.113	-0.273	0.157	1		
Similarity_{ij} (S)	0.1	-0.051	-0.151	0.129	0.347	1	
Generality_j (G)	-0.001	0.006	-0.075	0.001	0.027	0.073	1

- **Key paper:** Wuestman, M. L., Hoekman, J., & Frenken, K. (2019). The geography of scientific citations. *Research Policy*, 48(7), 1771–1780. <https://doi.org/10.1016/j.respol.2019.04.004>
- Dependent Variable: Citation Link (dummy)
- Independent variables: Bibliographic Coupling, Location Couplings, Times Cited, Shared Authors, Authors Product, References Product, Interaction terms
- Control selection: all others - weighted sample
- Field: Life Sciences and Medicine
- Key Finding: “when knowledge relatedness is high, the effect of geographical co-location on citation likelihood is non-existent.”
- N = 5 million
- Model: Logistic Regression Model

Generality

$$G_j = \sum_{k=1}^{N_l} \left(\frac{N \text{ citing}_k}{N \text{ citing}_l} \right)^2$$

k = patent class (subject field from web of science for the journal of publication in the case of a journal article), j = observation, l = all patent classes.

However, the measure discussed above can suffer from bias which results in the diversification measures becoming biased downwards when the total number of citations to and from an article are small (Hall, 2005). Hall (2005) provides a method to correct this bias as:

$$\hat{\eta} = \frac{N \cdot G_j - 1}{N - 1}$$

$D_{ij} \times S_{ij}$	0.008***	-	-	-	-
	(0.002)	-	-	-	-
$S_{ij} \times G_j$	-	0.001	-	-	-
	-	(0.002)	-	-	-
$D_{ij} \times B_{ij}$	-	-	0.029***	-	-
	-	-	(0.005)	-	-
$S_{ij} \times B_{ij}$	-	-	-	-0.050***	-
	-	-	-	(0.002)	-
$B_{ij} \times G_j$	-	-	-	-	0.021***
	-	-	-	-	(0.004)
Obs	32,270	32,270	32,270	32,270	32,270
Cluster Std Err	Yes	Yes	Yes	Yes	Yes

Note: : *p < 0.05, **p < 0.01, ***p < 0.001. (): robust standard errors.

Equation	1	2	3	4
Model	Probit	Probit	Probit	Probit
D _{ij}	-0.062***	-0.081***	-0.062***	-0.062***
	(0.008)	(0.008)	(0.008)	(0.008)
L _{ij}	0.082***	0.026**	0.084***	0.084***
	(0.009)	(0.009)	(0.009)	(0.009)
P _{ij}	0.066***	0.084***	0.063***	0.063***
	(0.008)	(0.007)	(0.008)	(0.008)
B _{ij}	0.301***	-	0.288***	0.288***
	(0.019)	-	(0.019)	(0.019)
S _{ij}	-	0.117***	0.038***	0.038***
		(0.007)	(0.008)	(0.008)
G _j	-	-	-	-0.006
	-	-	-	(0.007)
Obs	32,270	32,270	32,270	32,270
Cluster Std Err	Yes	Yes	Yes	Yes

Note: : *p < 0.05, **p < 0.01, ***p < 0.001. (): robust standard errors.

	$C_{ij} = 1$						$C_{ij} = 0$					
Variables	D	L	P	B	S	G	D	L	P	B	S	G
Distance (D)	1						1					
Lag (L)	0.080	1					0.047	1				
Publication (P)	-0.023	-0.155	1				-0.030	-0.107	1			
Bibliographic Coupling (B)	-0.109	-0.321	0.151	1			-0.093	-0.224	0.138	1		
Similarity (S)	-0.050	-0.159	0.124	0.356	1		-0.038	-0.142	0.122	0.315	1	
Generality (G)	0.009	-0.075	-0.005	0.027	0.073	1	0.004	-0.075	0.008	0.024	0.071	1