On the Heels of Giants:

### Internal Network Structure and the Race to Build on Prior Innovation

Nicholas S. Argyres, Washington University in St. Louis Luis A. Rios, Purdue University Brian S. Silverman, University of Toronto

> Presented at RIETI November 19, 2024

### But first, a bit about me

I frequently study the organization of <u>internal</u> R&D, often with Nicholas Argyres and Luis Rios

- "R&D, Organization Structure, and the Development of Corporate Technological Knowledge" with Nicholas Argyres, *Strategic Management Journal*, 2004
- 2. "Organizational Change and the Dynamics of Innovation" with Nicholas Argyres and Luis Rios, *Strategic Management Journal*, 2020
- 3. "On the Heels of Giants: Internal Network Structure and the Race to Build on Prior Innovation"

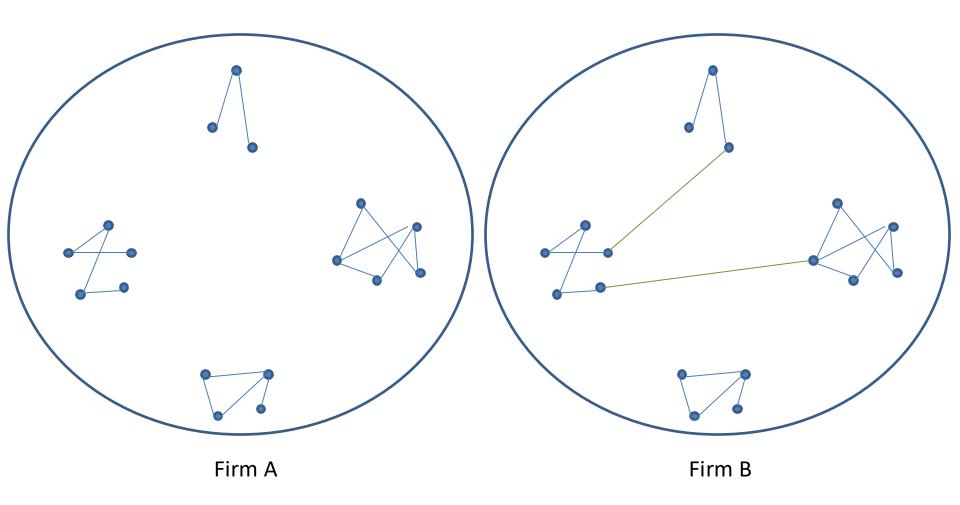
with Nicholas Argyres and Luis Rios, *Strategic Management Journal*, 2<sup>nd</sup> revise-and-resubmit]

### Background

- Proprietary technological knowledge often forms the basis for a firm's competitive advantage
- Developing and sustaining such an advantage requires building on the firm's own knowledge before rivals do.
  - "...a race between an innovator and the ability of the imitating firm [to] reverseengineer...the substantive technology." (Kogut & Zander 1992: 393)
  - "Generative appropriability" (Ahuja, Lampert & Novelli 2013)
- Literature emphasizes complementary assets, IP enforcement, non-compete agreements, location choice
  - Teece 1986; Mitchell 1991; Cohen et al. 2000; Agarwal et al. 2009; Marx et al. 2009; Somaya 2012; Alcacer & Chung 2007; Bloom et al. 2013
- Less on intra-organizational characteristics (Exceptions: Liebeskind 1996; Zhao 2006)
- Intrafirm inventor networks help determine innovation directions and outcomes (Reagans & McEvily 2003; Singh 2005; Sorenson et al. 2006; Singh & Agrawal 2011; Moreira et al. 2018)
  - But focus is on diffusion, absorption and impact, not appropriability or speed

### **Research question and findings**

- Does the structure of a firm's internal inventor network affect its generative appropriability? If so, then how?
- Sample: All 1,391 firms appearing in COMPUSTAT between 1986 and 2013 that had at least 250 patents during the sample period
- Result: More integrated ("small-worldy") internal networks → greater and faster generative appropriation
  - More integrated networks → higher self-citation rate, as proportion of overall citations
  - Relationship is stronger immediately after patent application, declining over time
  - Self-citation is positively associated with value appropriation



Firm B's network is more highly connected than Firm A's network

### Inventor networks and innovation

- Intrafirm inventor networks help determine innovation directions and outcomes (Singh 2005; Sorenson et al. 2006; Singh & Agrawal 2011; Zhao 2006)
- Key attributes of network: brokerage and closure
  - Networks rich in structural holes facilitate *creation* of new innovations... whereas more closed networks facilitate their *diffusion* (Hansen 1999; Reagans & Zuckerman 2001; Obstfeld 2005)
- Whole networks that are more "integrated" are characterized by high levels of brokerage and closure, and...
  - Generate more impactful innovations (Guler & Nerkar 2012; Argyres et al. 2020)
  - Expedite the absorption of external knowledge (Moreira et al. 2019) and knowledge recombination (Carnabuci & Operci 2013)
  - Implication: more integrated inventor networks → more rapid leveraging of knowledge to build upon existing innovations

### Internal networks and appropriability: Evidence from practice

- "...an idea in one area may be able to be translated into another therapeutic area. Quite often an indication may be unsuccessful in one therapeutic domain but have legs in another, however with the wrong structure (they) may not be able to take advantage of this."
  - quoted in Balachandron & Eklund 2019

- Merck developed Boceprevir to fight hepatitis C, but was unable to quickly identify its potential in fighting COVID.
- Pfizer used Boceprevir in its COVID vaccine -- realized approx. \$17 billion from it in 2022.

### Internal networks and appropriability

- Integrated internal network facilitate flows of knowledge and information:
  - Innovations generated elsewhere in the large firm
    - Reagans & Zuckerman 2001; Obstfeld 2005
  - Uncodified knowledge
    - Polanyi 1962; Teece 1982; Winter 1987
  - Prior failures
    - Nelson & Winter 1982; Sitkin 1992; Eggers 2012; Khanna et al. 2016
- H1: Greater integration of a firm's internal inventor network is associated with greater generative appropriability

### Inventor networks and the duration of appropriability

- Everything diffuses eventually (Jaffe et al. 1993; Sorenson et al. 2006)
- Innovating firms have limited time in which to pre-emptively build on their own knowledge (Kogut & Zander 1992; Ceccagnoli 2009)
  - Exacerbated for firms with highly integrated networks, as the greater number of people exposed to the knowledge → more potential conduits for leakage
  - Ideally (from a focal firm's perspective), by the time the initial knowledge fully diffuses, the firm will already be building on its own follow-on innovations (Ahuja et al. 2013)

H2: The positive relationship between the integration of a firm's internal inventor network and its generative appropriability is strongest immediately after the initial invention, and weakens as the initial invention ages

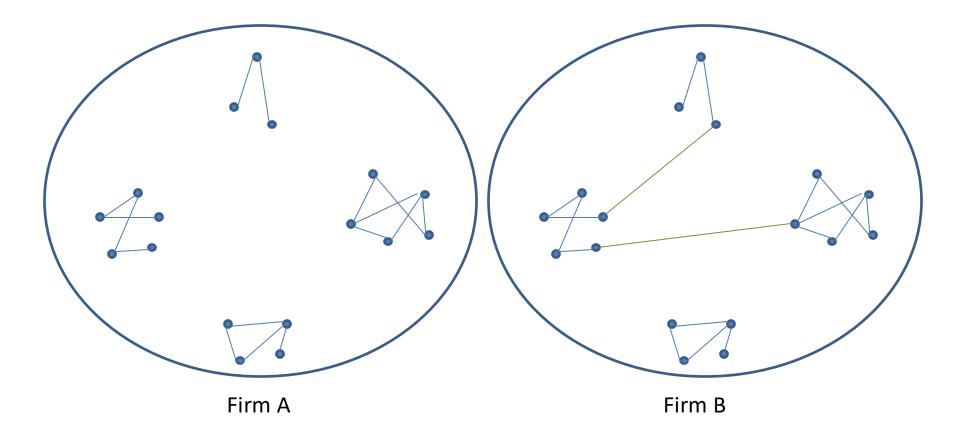
- Co-invention network properties
  - Patent ownership and bibliometric dataset constructed by matching EPO's PATSTAT, USPTO, Bureau VanDjik's ORBIS database, Lee Fleming's Berkeley data project, and NBER dataset
- Innovative outcomes, e.g. patent citations
  - PATSTAT bibliometrics; examiner-added citations removed
- Characteristics of firms
  - COMPUSTAT
- Sample
  - All 1,391 firms appearing in COMPUSTAT between 1986 and 2013 that had at least 250 patents during the sample period
  - All patents for these firms with application dates between 1986 and 2013, aggregated at the patent family level: > 400K patent families
  - Citations to these patents through 2019

### Co-invention network properties

- Patent ownership and bibliometric dataset constructed by matching EPO's PATSTAT, USPTO, Bureau VanDjik's ORBIS database, Lee Fleming's Berkeley data project, and NBER dataset
- Innovative outcomes, e.g. patent citations
  - PATSTAT bibliometrics; examiner-added citations removed
- Characteristics of firms
  - COMPUSTAT

## - Sample

- All 1,391 firms appearing in COMPUSTAT between 1986 and 2013 that had at least 250 patents during the sample period
- All patents for these firms with application dates between 1986 and 2013, aggregated at the patent family level: > 400K patent families
- Citations to these patents through 2019

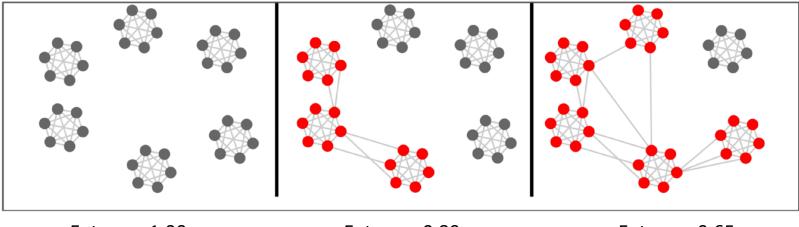


Firm B's Giant Component is larger than Firm A's Giant Component

 $Giant_{kt} = \frac{NumInvLargestComponent_{kt}}{NumInvFirm_{kt}}$ 

### **Network Integration, measure 2: Entropy**

You can think of this as an inverse Herfindahl based on cluster sizes



Entropy=1.00 Giant=1/6 (6/36) Entropy=0.89 Giant=1/2 (18/36) Entropy=0.65 Giant=5/6 (30/36)

Usually, Giant and Entropy are strongly negatively correlated...

 $H = -\sum_{c=1}^{C} [(Nc/N) * log(Nc/N)], \qquad Entropy_{kt} = H_{norm} = \frac{H}{logC}$ 

### Small-world measure of "near-decomposability" (based on Watts 2001)

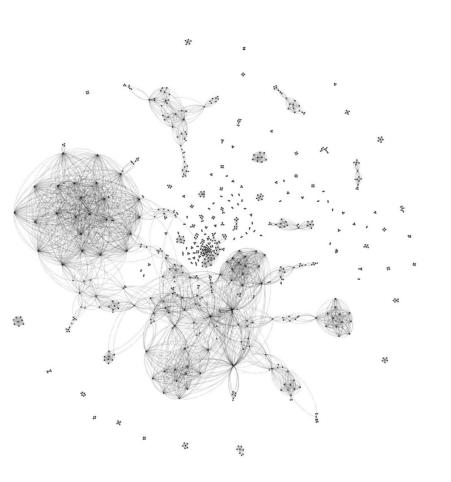
SmallWorld =  $(C_A/C_R)/(L_A/L_R)$ .

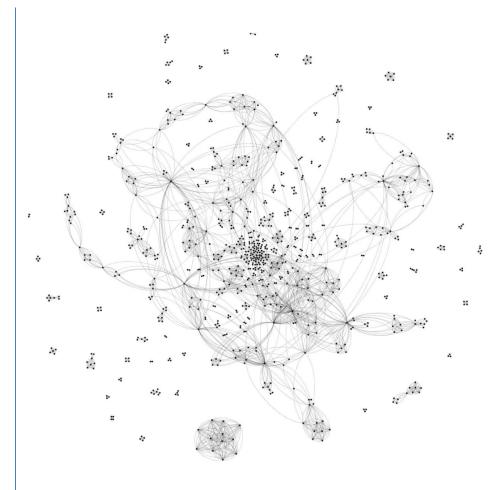
At a conceptual level:

- Take a given network
- Calculate the average path length between any two inventors [L<sub>A</sub>]
  - Higher L = information must bass through more people to reach someone
- Calculate the ratio of the total number of ties among all inventors vs. the total number of possible ties among all inventors [C<sub>A</sub>]
  - Higher C = pockets of dense clustering
- Generate a random network using the same number of nodes and ties as in the actual network
- Calculate the average path length [L<sub>R</sub>] and ratio of ties [C<sub>R</sub>] in the random network
- Calculate SmallWorld

Interpretation: Higher values of SmallWorld indicate networks with unusually high clustering and unusually low path lengths – classic nearly-decomposable networks.

### Whole Networks - examples from our data





- Co-invention network properties
  - Patent ownership and bibliometric dataset constructed by matching EPO's PATSTAT, USPTO, Bureau VanDjik's ORBIS database, Lee Fleming's Berkeley data project, and NBER dataset
- Innovative outcomes, e.g. patent citations
  - PATSTAT bibliometrics; examiner-added citations removed
- Characteristics of firms
  - COMPUSTAT

Two exciting twists: --Citations to applications --Patent families

### - Sample

- All 1,391 firms appearing in COMPUSTAT between 1986 and 2013 that had at least 250 patents during the sample period
- All patents for these firms with application dates between 1986 and 2013, aggregated at the patent family level: > 400K patent families
- Citations to these patents through 2019

# -

## What's a patent family, anyway?

- Silverman Widgets, Inc. files a U.S. patent application for a new transmogrifying demodulator
- The initial "seminal priority application" is very broad, and/or vague
- As the patent application progresses, Silverman Widgets revises the application, breaking it into separate applications for different parts
  - Or it adds new, clarifying claims (often based on subsequent research)
- Even after the first patent is granted, Silverman Widgets may file subsequent patents that the patent office notes are based on the same "seminal priority application," if they trace to those initial claims
- All patent applications linked to the seminal priority application are part of the same patent family (Rios 2020; Kuhn et al. 2020)

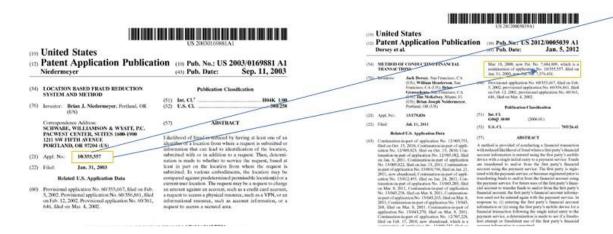


the inventor  $\rightarrow$ 

### Figure 1B: Domestic patent family.

Eight different granted patents covering the same technology, and claiming same priority date. Gauging the impact of this invention calls for aggregating at the domestic family level.

We can see here the evolution of the Square payment ecosystem-seminal tech was a payment authentication method which enhanced original reader



### ... United States - Patent Application Publication - Pub. No.: US 2012/0118960 A Sather of al. Aug 21, 201 ATA MARINE 10--18

#### 

			110			
M)	PURCHAN	IF TRANSPECTIVE TON THOM A CARD BEADER OVERMETERS SPRENG TO A CVEX	1	A. 2002 - particular	and application.	HLMT Station Feb No. 2010(1344, Elat plantes, No. 2010);
(%)	inster.	Killer John Perrick Saller, Kar Freedom, CA (197), Anish Rahe		Patri	ative ( Section	-
		San Education 3. A (UN). Company Maples. Freedomi, C.A. (UN). Ault.	(11)	fac.03.	(2006.00	
		Burney, San Demisson, C.A.(URL, James M. McKaher, Massi, FL	(12)	6A.G.		11540
		110	455		SISTENT	
21)	App. No.:	13/298,814				- i muttale der hat some
22)	rist	Sec.15, 2011	- 144	sky a bound with	hough height a	reig: The coall leads all width (Internetion)
	B.M.	ed 110. Application Data	1.50	d head with a slow	a granteception	or, an empty path not enal book. The access
7	Markon Int. 2 Yes: 15:1762 contraction 38ed on Fell	-in-part of application No. 15/11214 7. 2011 which is a difference of application (No. 884) on FeL 111, 2010, which is the part of application No. 12/2017/2 177, 2020, networkstands which is of application No. 12/00/172, HeL 10 in our Put No. 12/00/172, HeL 10	11117	ing, with the host equilaterative of its and used host. A cost of anticertime. A cost re-strated on the its	uring length and good to the length o chil to the dot response signal agnetic signal	of to be officer in the 1 height origins being to not height origins of Dista in multiplet officer of to provide origination from the two insignation from the two insignation from

10-

#### 

pplication Publication	B on Pub. No.: US 2012/0126014 A1 on Pub. Date: May 24, 2012
ENTROPOLI MAD BLAD	anninuation of opplication No. 101001.007, 6540 on
Ellist John Patrick Audion: Too Francisco, CA (1915) Audio Bahn, San Francisco, CA (1915) Geograph Regles, Persteam, CA (1915) Jack	<ol> <li>Processed application No. 697152,007 (Ref on Feb. 9, 2002) percentenal application No. 60793,003, 80ed on Feb. 12, 2002 percentenal application No. 807161, 3485, 564169, Mar. 8, 2002</li> </ol>
Juney M. McKabey, Mani, Fl.	Publicative Classification
Aver .	ALL BALKS

# No. 17.30

--- United States

.... Patent Ap

Sather et al. OU CARDINAN

#### 

(17)	United Patent Sather et	Application Publicati	on	- Pub. N	a: US 2012/01	26013 A1 y 24, 2013
ite	METROD FOR TRAVAMITING PRODUCTOR FROM ACADO READER TO A MODEL OF BUT WITH THE CARD READER OF THE RAD BEAD OFFICE			Mar 13, 200	d application No. 120 6. anti- Par. No. 120 d application No. 1933 anti-Pat. No. 1935	ULAR MAR
	FROM A CENTRE OF AN ASSOCIATED - BOUNDAG		(10)	Provisional application No. 48(201,667, 884) on Ful 3, 2002, pervisional application No. 40(716,811, 85)		
(29)	Dooster Ellist Adva Safet, Sol Transien, CAUDO, Salah Baba, Son		ver Fall, 52, 28 state, filled we b	C. providenced spation	mine No. 407347	
		Francisco, C.L.(170), Gaugary Wagden, Presidente, C.A.(270), Back		P.M	to arise Chevillening	
		Burney, San Francisco, CA (UR), James M. McRathers, Microll F1.	(71)	Bail CL.	(2000.011)	
		(inter-	(12)	18.0		220.94
(11)	100.00	11096,504	dft		ABITEM T	
(III)	1944	Nov. 17, 2011	10	Field of Streements	tring tail termsticing for 3 and if a formering. The second for	mit Ares and
	No. ba	od C.S. Application Date			fa angintic only then with a holometer the si	
	theirs is 1 No. 19170	August of application No. 191922147, 7. 2019, which is endowment application One And on Art. 11, 2011, which is a support of application No. 122997228, 17, 2010, new abandward, which is a	1111	a don them in who signal is pr who arrays the a processed for	and from the storgest robusted automatics of its over magnetic based sig- ingl signed that the over act signed is some to the	ic strips A re- als scool on the perf is contacted mechanicalities on

10-

	application No. 12000 anin Pas No. 12000 application No. 12000 m Pas, No. 12000			CONTERP CONTRACTOR	66 (B) (
17 percente 17 percente 18 percente	Annual No. of Status Status No. of Status A. (201)	Phys. Bill, High			012340
	(2090.01)				8
-		120.049	- 000	Appl No.	
-	ABOTEM'T ginliessingle Jaul	NB-APVEY MAR	(0)		
	brooking. The stud law				-

United States

Patent Application Pub

#### 

licati	on	- Pub. No.	: US 2012/01 e: Maj	26010 A1 24, 2012	(1 <sup>1</sup> )	United Patent Dursey of	App
		ine. 10, 2000, p	applements No. 197		(96)	CARD REAL	
05	1	3, 200C, perchase	hamps, No. 09301,0 mail appleosing No. 6 (	spirits, 84.1, 82-4	09	investory.	40.20
1.17.0		Public	when Classification				in the
A	(12)	NALES CANA THE EX.CL	(Jeek m)	215.649			10.4
	- 65		ABSTRACT		100	Apr. 10-	-
	24	a real head scaling	and it have appled to	a sector de las		Title .	-
	100	The multihead po-	who deter out they using he signed indication of	pattor strips and I data wontil car		-	
NULIAE Annies William William Milliam Allana	111111	the comparison energy. Denses absorbers on provided with an ending line track and a concretence finite. The analysis of adjust there and a unequilate to a proceeding chemical with the externation without. The analysis of digist if them and therefore the externation finite of grant and assessments in sites a processed of giptal angle that the maximum endines and anonyme. The maximum endine grant- dence or aligned that any park is adopted to be serviced in a service of the maximum devices and deduces are sensible and and of the devices and deduces are compared and and of the maximum endines.			100	Factor of the State	1, (01) 100, 384 10, 284 10, 28
1.00	de :	enie anne				10-	

(A. 201 201 200 20006 A.)	

N BELA			anethin prior	IR som Per Ma 14	state which is
-	And Davier, San Francisco, CA (101), Nathan P. McCaster, San Francisco, CA (103), Justin Philip Campilian, Nutritical Statistics, CA (103), Daga Maning, San Francisco, CA		A 2002 per martine 12.1	epitonics No. 6615 Sectored epitodese N 2012, presented and Mar. 4, 2012	to MATCHING BOT HEAD
	clift; Offree Named Carrier Oniples: You Vesterant C.4 (201)		Publication Chevillustion		
	Animer M. Melkahery, Kilanin, 11. (170)	(71)	Sal-US. Grad Test Grad Test	(348,45)	
÷.,	11/246,077	-	LACK	0000.003	Disast Driver
	Nex 11,200	12		ABILITY	
-	of C.A. Applicative Deta	1.1	-	hearing, want hand at	A Real Property lies and the second se
notion in part of application No. 13/112/14 <sup>1</sup> , 302 (7) (2014) adjustion after some of applications (19/10/00). Biol on Adj. 12, 2014. Models is a contrast singutt of application No. 12/10/1228, a Tab. 17, 2006, new adjustioned wheelp is a minim of application. No. 12000/122, Bod on		22213	president a com- part of an and a second sec	equal lines the read i spirit eigent that the controller with course digited capacit lines to sympt and signal se- spirat and signal se- art. As contain park to	and and connects to recommendar can not legal structure a Sector observation recognition of the recognition of the recognition of the Sector

### What's a patent family, anyway? (aside)

- Silverman Widgets, Inc. decides to also file "equivalent" patent applications for the same technology in Canada, Japan, etc.
- Thanks to harmonization of patent data through the EPO, PATSTAT links these applications in different jurisdictions to the same seminal priority application
- All patent applications across jurisdictions that are linked to the seminal priority application are also part of the same patent family (Rios 2020; Kuhn et al. 2020)

### **Key Variables**

SelfRatio <sub>jk</sub>	(# of citations by firm k to patents or patent applications in firm k's patent family j through 2019) / (# of all citations to patents or patent applications in firm k's patent family j through 2019). [for H1] for H2, include a time-clock variable = # of months since the date of the patent application; interact with SelfRatio. Key variable is SelfRatio*Months
Giant <sub>kt</sub>	(# of inventors in the largest component of firm k's inventor network in year t) / (# of all inventors in firm k's inventor network in year t) [Higher Giant → more integrated network]
Entropy <sub>kt</sub>	A variant of 1 - herfindahl index of the sizes of components of firm k's inventor network in year t. See equations (1) and (2) in the text for precise definition. [Lower Entropy → more integrated network]

### **Patent Family Control Variables**

Generality <sub>jk</sub>	1 - herfindahl index of the technology classes of patent applications that cite firm k's patent family j, per Hall et al. (2001)
Originality <sub>jk</sub>	1 - herfindahl index of the technology classes of patent applications that are cited by firm k's patent family j, per Hall et al. (2001)
<b>Cites To Publications</b> <sub>jk</sub>	Count of the number of citations to scientific publications that appear on patent applications in firm k's patent family j
LnFamilyAge <sub>jk</sub>	Number of years between application year of seminal application and application year for the most recent application in firm k's patent family j, as of 2013
Patent complexity <sub>jk</sub>	The Fleming/Sorenson complexity measure averaged for all applications in firm k's patent family j



# Firm-year Control Variables

LnR&D <sub>kt-1</sub>	Natural log of (1 + firm k's R&D expenditure) in year t-1
LnAssets <sub>kt-1</sub>	Natural log of (1 + firm k's assets) in year t-1
LnEmployees <sub>kt-1</sub>	Natural log of (1 + firm k's employees) in year t-1
LnSales <sub>kt-1</sub>	Natural log of (1 + firm k's revenue) in year t-1
LnPatents <sub>kt-1</sub>	Natural log of (1 + the number of patent applications submitted by firm k) in year t-1
LnComponents <sub>kt-1</sub>	Natural log of (1 + the number of distinct components in firm k's inventor network) "Round Up The Usual Suspects!" (CASABLANCA)



### Measures of Patent Value (as a check on the validity of SelfRatio)

YrsRenewed <sub>jk</sub>	Average number of years that the granted patents in firm k's patent family j are renewed, through 2019
Jurisdictions <sub>jk</sub>	Count of the number of countries in which at least one application in firm k's patent family j is submitted
Triad <sub>jk</sub>	Set equal to 1 if at least one application in firm k's patent family j is granted in all three of these jurisdictions: US, Japan, and Europe; set equal to 0 otherwise
Patent family size <sub>jk</sub>	Count of the total number of patent applications in firm k's patent family j as of 2013

### **Summary Statistics**

	Ν	Mean	SD	Min	Max
SelfRatio	430,061	0.166	0.278	0.000	1.000
Giant	430,061	0.503	0.200	0.025	1.000
Entropy	430,061	0.407	0.233	0.012	1.000
SmallWorld	429,461	16.123	16.579	0.169	60.73
Months	415,491	85.782	63.836	0.000	312
Jurisdictions	430,061	3.029	3.027	1.000	45.0
Family Size	430,061	3.991	4.687	1.000	297
Triadic	430,061	0.411	0.492	0.000	1.000
YearsRenewed	430,061	5.712	5.954	0.000	25.00
Generality	409,596	0.468	0.259	0.000	0.975
Originality	421,840	0.732	0.197	-0.652	0.989
Cites to Publications	425,484	2.687	7.908	0.000	220
ln(Family Age)	430,061	0.381	0.623	0.000	3.367
Complexity	429,966	3.143	1.978	0.151	68.000
ln Patents/year t-1	430,061	6.064	1.719	0.693	9.235
ln(components)	430,061	4.660	1.233	0.693	6.382
ln(R&D	430,061	6.722	1.760	0.000	9.437
ln(Assets) t-1	430,061	9.609	1.878	0.678	13.22
ln(Emp)t-1	430,061	3.938	1.435	0.005	6.777
ln(Sales) t-1	430,061	9.355	1.984	0.000	12.449

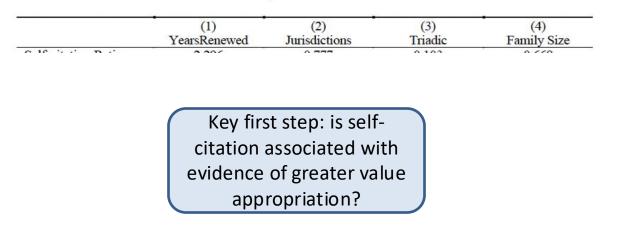


Table 5: Patent-family value as a function of self-citation ratio

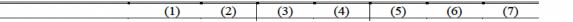
 $Value_{j} = \alpha + \beta_{1}SelfRatio_{j} + \gamma X_{kt-1} + \theta Z_{j} + \delta Firm_{k} + \omega Year_{t} + \mu Tech_{j} + \epsilon_{j}$ 

	(1)	(2)	(3)	(4)
	YearsRenewed	Jurisdictions	Triadic	Family Size
Self-citation Ratio	2.296	0.777	0.103	0.668
	(0.135)	(0.132)	(0.022)	(0.159)
	[0.000]	[0.000]	[0.000]	[0.000]
Patents/year <sub>t-1</sub>	-0.006	-0.297	-0.044	-0.367
	(0.196)	(0.106)	(0.020)	(0.143)
	[0.976]	[0.005]	[0.029]	[0.010]
ln(components)	-0.047	0.286	0.053	0.215
	(0.216)	(0.092)	(0.019)	(0.134)
	[0.829]	[0.002]	[0.005]	[0.109]
ln(R&D) <sub>t-1</sub>	0.181	0.030	0.006	0.165
	(0.144)	(0.073)	(0.013)	(0.099)
	[0.209]	[0.683]	[0.641]	[0.094]
ln(Assets) <sub>t-1</sub>	-0.098	0.073	0.005	-0.140
	(0.095)	(0.053)	(0.010)	(0.086)
	[0.302]	[0.163]	[0.596]	[0.103]
ln(Sales) <sub>t-1</sub>	-0.051	0.035	-0.007	0.085
	(0.036)	(0.026)	(0.006)	(0.043)
	[0.157]	[0.179]	[0.247]	[0.046]
ln(Emp) <sub>t-1</sub>	0.106	-0.126	-0.013	-0.356
	(0.227)	(0.174)	(0.029)	(0.262)
	[0.641]	[0.468]	[0.659]	[0.174]
Originality	-0.175	0.202	0.045	0.403
	(0.152)	(0.115)	(0.017)	(0.177)
	[0.249]	[0.080]	[0.007]	[0.023]
Generality	0.312	0.257	0.047	0.327
	(0.127)	(0.071)	(0.014)	(0.080)
	[0.014]	[0.000]	[0.001]	[0.000]
Cites to Publications	0.039	0.002	0.002	0.011
	(0.006)	(0.003)	(0.000)	(0.005)
	[0.000]	[0.590]	[0.000]	[0.019]
ln(Family Age)	-1.551	1.316	0.207	2.631
	(0.084)	(0.108)	(0.011)	(0.173)
	[0.000]	[0.000]	[0.000]	[0.000]
Complexity	-0.126	-0.157	0.009	0.213
	(0.016)	(0.027)	(0.003)	(0.038)
	[0.000]	[0.000]	[0.000]	[0.000]
Constant	7.516	1.368	0.340	1.055
	(0.653)	(0.545)	(0.074)	(0.884)
	[0.000]	[0.012]	[0.000]	[0.233]
Observations	412,839	412,839	412,839	412,839
Adjusted <i>R</i> <sup>2</sup>	0.149	0.251	0.181	0.305
Firm, Tech, & Year FE	Yes	Yes	Yes	Yes

### Table 5: Patent-family value as a function of self-citation ratio

	(1)	(2)	(3)	(4)
	YearsRenewed	Jurisdictions	Triadic	Family Size
Self-citation Ratio	2.296	0.777	0.103	0.668
	(0.135)	(0.132)	(0.022)	(0.159)
	[0.000]	[0.000]	[0.000]	[0.000]
Patents/year <sub>t-1</sub>	-0006	-0.297	-0.044	-0.367
	(0. 6)	(0.106)	(0.020)	(0.143)
	[0.9	[0.005]	.029]	[0.010]
ln(components)	-0.04	0.286	53	0.215
	(0.21	(0.092)	( 9)	(0.134)
	[0.829]	[0.002]		[0.109]
$\ln(R\&D)_{t-1}$	0.181	0.030	0.	0.165
	(2			(0.099)
	1 std dev	[0.094]		
ln(Assets) <sub>t-1</sub>	9% increase in YrsRenewed			-0.140
()	(0.095)	(0.053)	(0.01	(0.086)
	[0.302]	[0.163]	1 std dovi	increase in SelfRatio $\rightarrow$
ln(Sales) <sub>t-1</sub>	-0.051	0.035		
	(0.036)	(0.026)	5%-10%	increase in other DVs
	[0.157]	[0.179]	[0.247]	[0.046]
ln(Emp) <sub>t-1</sub>	0.106	-0.126	-0.013	-0.356
	(0.227)	(0.174)	(0.029)	(0.262)
	[0.641]	[0.468]	[0.659]	[0.174]
Originality	-0.175	0.202	0.045	0.403
6	(0.152)	(0.115)	(0.017)	(0.177)
	[0.249]	[0.080]	[0.007]	[0.023]
Generality	0.312	0.257	0.047	0.327
5	(0.127)	(0.071)	(0.014)	(0.080)
	[0.014]	[0.000]	[0.001]	[0.000]
Cites to Publications	0.039	0.002	0.002	0.011
	(0.006)	(0.003)	(0.000)	(0.005)
	[0.000]	[0.590]	[0.000]	[0.019]

Table 3: Self-citation ratio (SelfRatio) as a function of inventor network structure



Testing H1: Is a more connected inventor network structure associated with a higher selfcitation rate?

 $SelfRatio_{j} = \alpha + \beta_{1}Networkmeasure_{kt-1} + \gamma X_{kt-1} + \theta Z_{j} + \delta Firm_{k} + \theta Z_{kt-1}$ 

 $\omega Year_t + \mu Tech_j + \epsilon_j$ 

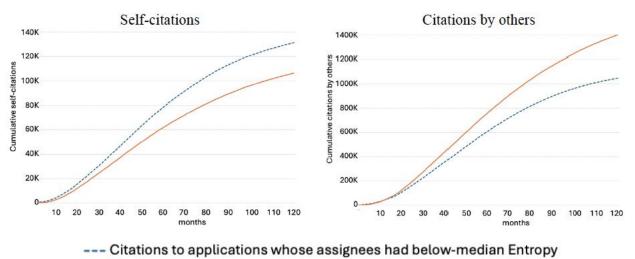
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Entropy	-0.112	-0.113				-0.165	
	(0.039) [0.004]	(0.041)				(0.050)	
Ciant	[0.004]	[0.005]	0.101	0.100		[0.001]	0 1 4 2
Giant			0.101 (0.026)	0.100 (0.027)			0.143 (0.035)
			[0.000]	[0.000]			[0.000]
SmallWorld				L	-0.000	0.001	-0.018
Shahwork					(0.001)	(0.001)	(0.004)
					0.808	[0.231]	[0.000]
Entropy*SmallWorld						-0.018	
						(0.004)	
						[0.000]	
Giant*SmallWorld							0.019
							(0.004)
	0.005		0.000		0.020	0.011	[0.000]
Patents/year <sub>t-1</sub>	0.005 (0.015)	0.002 (0.014)	0.003 (0.014)	0.000 (0.013)	0.020 (0.013)	-0.011 (0.012)	-0.014 (0.011)
	[0.745]	[0.902]	[0.828]	[0.991]	[0.114]	[0.369]	[0.210]
ln(components)	-0.035	-0.033	-0.034	-0.032	-0.051	0.003	0.005
in(components)	(0.012)	(0.011)	(0.011)	(0.010)	(0.011)	(0.010)	(0.010)
	[0.003]	0.003	[0.002]	[0.002]	0.000	0.808	[0.581]
ln(R&D) <sub>t-1</sub>	0.022	0.022	0.021	0.022	0.023	0.024	0.023
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.009)	(0.009)
	[0.030]	[0.022]	[0.033]	[0.025]	[0.021]	[0.013]	[0.015]
In(Assets) <sub>t-1</sub>	0.017	0.017	0.017	0.017	0.017	0.018	0.018
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)
	[0.000]	[0.001]	[0.000]	[0.001]	[0.000]	[0.004]	[0.004]
ln(Sales) <sub>t-1</sub>	0.012	0.012	0.012	0.012 (0.005)	0.012	0.012	0.012
	(0.005) [0.011]	(0.005) [0.011]	(0.005) [0.011]	[0.011]	(0.005) [0.015]	(0.005) [0.016]	(0.005) [0.016]
ln(Emp) <sub>t-1</sub>	-0.033	-0.031	-0.033	-0.030	-0.027	-0.031	-0.030
m(rsmp)t-1	(0.011)	(0.011)	(0.012)	(0.011)	(0.010)	(0.010)	(0.010)
	[0.004]	[0.004]	[0.005]	[0.005]	[0.008]	[0.002]	[0.003]
Originality		-0.058		-0.058	-0.056	-0.060	-0.060
		(0.016)		(0.016)	(0.015)	(0.015)	(0.014)
		[0.000]		[0.000]	[0.000]	[0.000]	[0.000]
Generality		-0.060		-0.060	-0.057	-0.058	-0.058
		(0.021)		(0.021)	(0.022)	(0.021)	(0.021)
		[0.005]		[0.005]	[0.011]	[0.005]	[0.005]
Cites to Publications		0.002		0.002	0.002	0.002	0.002
		(0.000) [0.000]		(0.000) [0.000]	(0.000) [0.000]	(0.000) [0.000]	(0.000) [0.000]
In (Energian Ange)							
ln(Family Age)		0.029 (0.005)		0.029 (0.005)	0.030 (0.005)	0.026 (0.005)	0.027 (0.005)
		[0.003]		[0.003]	[0.003]	[0.003]	[0.003]
Complexity		-0.004		-0.004	-0.004	-0.004	-0.004
Complexity		(0.001)		(0.001)	(0.001)	(0.001)	(0.001)
		[0.000]		0.000	<u>[0.000</u> ]	<u>[0.000</u> ]	<u>[</u> 0.000]
Observations	430,037	406,577	430,037	406,577	406,030	406,030	406,030
Adjusted R <sup>2</sup>	0.141	0.154	0.142	0.154	0.151	0.160	0.162
Firm <del>,</del> Tech-& Year FE	Yes						

Table 3: Self-citation ratio (SelfRatio) as a function of inventor network structure

 Firm, Tech-& Year FE
 Yes
 Yes

ase in Entropy → e in SelfRatio	tio ( <i>Self</i>	R <i>atio</i> ) as	a func	1 std dev increase in Giant → 8% increase in SelfRatio
Entropy	(1) -0.112 (0.039) [0.004]	(2) -0.113 (0.041) [0.005]	(3)	
Giant			0.101 (0.026) [0.000]	0.100 (0.027) [0.000]
SmallWorld Entropy*SmallWorld				
Giant*SmallWorld				
 Originality		-0.058		-0.058
Generality		(0.016) [0.000] -0.060		(0.016) [0.000] -0.060
Cites to Publications		(0.021) [0.005] 0.002 (0.000) [0.000]		(0.021) [0.005] 0.002 (0.000) [0.000]

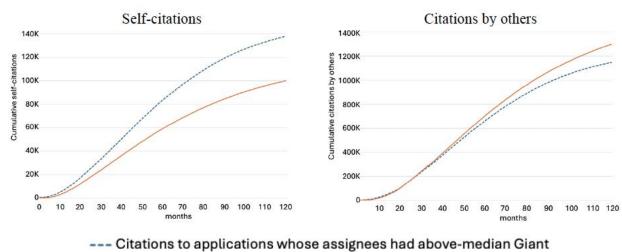
### The extent and timing of citations: Low- vs. high-integration networks



PANEL A: Integration proxied by Entropy (lower Entropy = higher integration)

— Citations to applications whose assignees had above-median Entropy

PANEL B: Integration proxied by Giant (higher Giant = higher integration)



---- Citations to applications whose assignees had below-median Giant

	(1)	(2)	(3)	(4)
Months	-0.001	-0.001	-0.001	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
	[0.000]	[0.000]	[0.000]	[0.023]
Entropy	-0.109	-0.218		
	(0.041)	(0.051)		
	[0.007]	[0.000]		
Entropy* Months		0.002		
		(0.000)		
		[0.000]		
Giant			0.096	0.189
			(0.027)	(0.035)
			[0.000]	[0.000]
Giant* Months				-0.001
				(0.000)
				[0.000]
SmallWorld				
SmallWorld* Months				
Entropy*SmallWorld				
Entropy*Months*				
SmallWorld				
Giant* SmallWorld				
Giant* Months*				
SmallWorld				

Table 4: Self-citation ratio as a function of inventor network structure and patent age

### Alternative specifications and robustness checks

- Fractional response models (Wooldridge 2010; Villadsen & Wulff 2019a, b)
- Firm-level estimation instead of patent-family level
- Patent-level estimation instead of patent-family level
- Sample = granted patents instead of patent applications
- DV: count of self-citations instead of self-citation ratio
- Duration models time-to-citation as a function of Giant or Entropy

### Summary

- Greater integration of intra-firm inventor network is associated with greater rates of self-citation
  - Our interpretation: evidence of greater ability to appropriate knowledge, a.k.a. generative appropriability
- This relationship is most salient in the crucial early years of a patent family's life – integration is associated with the relative speed with which a firm builds on its innovations
- Not a causal study, but we rule out some alternative explanations
  - E.g., that this is simply driven by different types of innovation generated by different inventor network configurations

### Inspirational/aspirational closing

- Contributes to the recent upsurge in work around whole-networks and innovation, particularly around small worlds and connectedness
- Extends our understanding to knowledge appropriation as well as knowledge creation
- Potential future research
  - Costs of integration?
    - Belderbos, Park & Carree (2021) finding
    - does the periphery generate more radical breakthroughs?
    - Inventor resistance [influence on inventor departures?]
  - Complementarities between inventor network structure and other mechanisms of knowledge appropriation