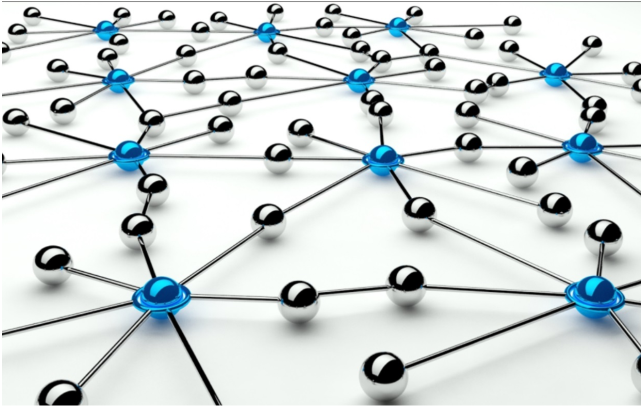
Text-Based Representations of Market Structures

Gerard Hoberg
University of Southern California
Marshall School of Business

Swissquote Conference [Nov 2018]

Concept of “industry” is undeniably important



* Which firms are industry rivals is most core to market structure.
* But world far more complex than “all or nothing” categorizations.
* What about close-rivals, partial-rivals, entry-threat rivals or complementary firms.

“Market Structure” » industry

Market structure much deeper:

- **Also:** Product differentiation.
- **Also:** Innovation and Change.
- **Also:** Supply chains .
- **Also:** Economies of scope & entry.

All are seminal to how markets work and evolve. Goal is to design a unified empirical framework incorporating all of the above.

Section I: Market Structure PAST

Section I: Market Structure PAST

Section II: Market Structure PRESENT

Section III: Impact and Uses

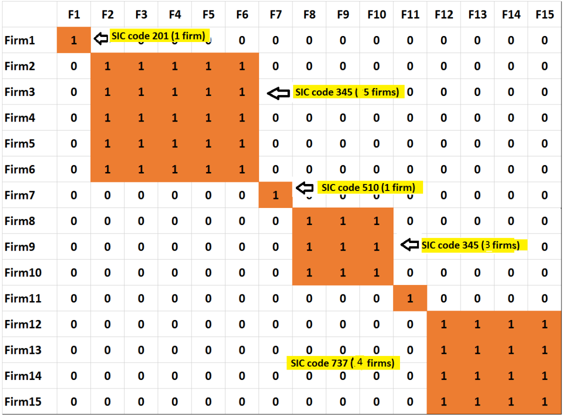
Section IV: Market Structure FUTURE

Section V: Using Text to Model Innovation

Old Way to Model Market Structure!

Govt employees gathering industry data for SIC codes.
This is how things were 50 years ago!

SIC code industries (a simple network!)



	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15
Firm1	1		0	0	0	0	0	0	0	0	0	0	0	0	0
Firm2	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0
Firm3	0	1	1	1	1	1					0	0	0	0	0
Firm4	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0
Firm5	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0
Firm6	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0
Firm7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Firm8	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0
Firm9	0	0	0	0	0	0	0	1	1	1					0
Firm10	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0
Firm11	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Firm12	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
Firm13	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
Firm14	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
Firm15	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1

SIC code 201 (1 firm)

SIC code 345 (5 firms)

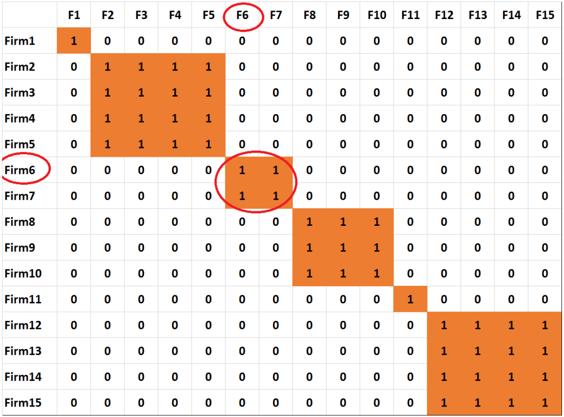
SIC code 510 (1 firm)

SIC code 345 (3 firms)

SIC code 737 (4 firms)

Researchers and practitioners typically identify industries using govt classifications such as SIC, NAICS.
These are constrained to be “transitive” & “binary”.

Limitation 1: Difficult to model change



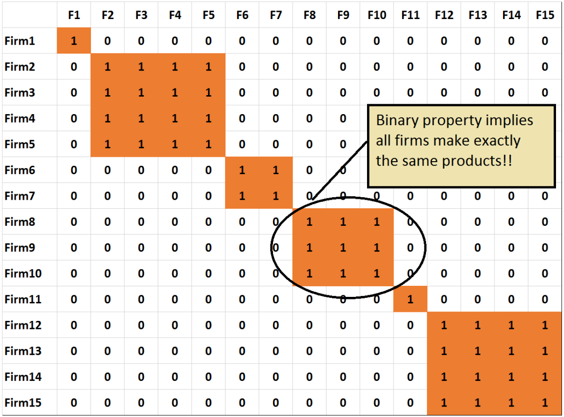
	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15
Firm1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Firm2	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0
Firm3	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0
Firm4	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0
Firm5	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0
Firm6	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
Firm7	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
Firm8	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0
Firm9	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0
Firm10	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0
Firm11	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Firm12	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
Firm13	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
Firm14	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
Firm15	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1

* How can a company model the aftermath of disruption?

* Can a firm detect a rival's gradual encroachment?

* Only a 100% transformation can justify reclassification (RARE).

Limitation 2: Cannot model product heterogeneity



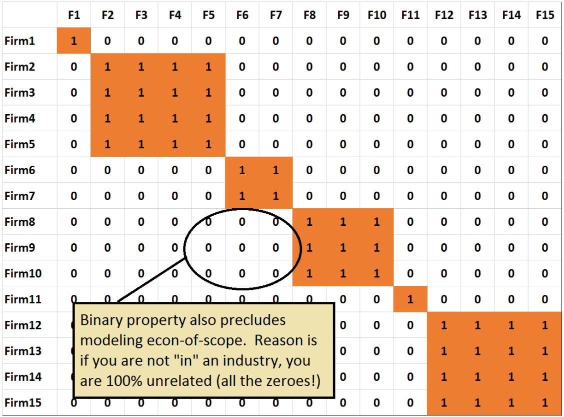
	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15
Firm1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Firm2	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0
Firm3	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0
Firm4	0	1	1	1	1	0	0	0	0						
Firm5	0	1	1	1	1	0	0	0	0						
Firm6	0	0	0	0	0	1	1	0	0						
Firm7	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
Firm8	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0
Firm9	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0
Firm10	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0
Firm11	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Firm12	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
Firm13	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
Firm14	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
Firm15	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1

Binary property implies all firms make exactly the same products!!

*SIC models all products in an industry as 100% identical.

*iPhones and Androids are exactly the same, right?

Limitation 3: Cannot model economies of scope



	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15
Firm1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Firm2	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0
Firm3	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0
Firm4	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0
Firm5	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0
Firm6	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
Firm7	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
Firm8	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0
Firm9	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0
Firm10	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0
Firm11	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Firm12	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0	0	1	1	1	1
Firm13	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0	0	1	1	1	1
Firm14	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0	0	1	1	1	1
Firm15	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1

Binary property also precludes modeling econ-of-scope. Reason is if you are not "in" an industry, you are 100% unrelated (all the zeroes!)

*Can firms identify markets with synergies or “low cost” entry.

*Cannot model entry threats either. No information in SIC codes!

Section II: Market Structure PRESENT

Section I: Market Structure PAST

Section II: Market Structure PRESENT

Section III: Impact and Uses

Section IV: Market Structure FUTURE

Section V: Using Text to Model Innovation

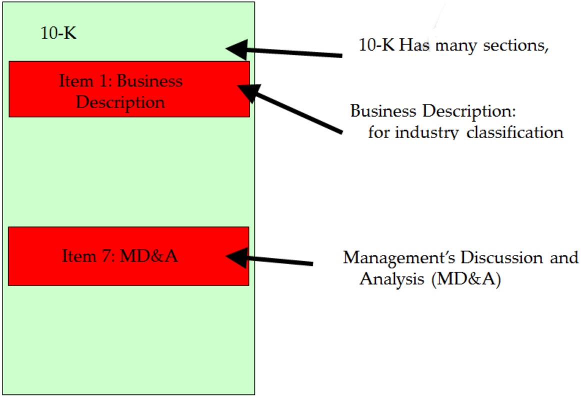
New Way to Model Market Structure!

Computers are the most loyal work force ever. They only want electricity! Use scaleable NLP to create generalized concept of industry.

Coming slides relate to Textual Network Industry (TNIC) Research

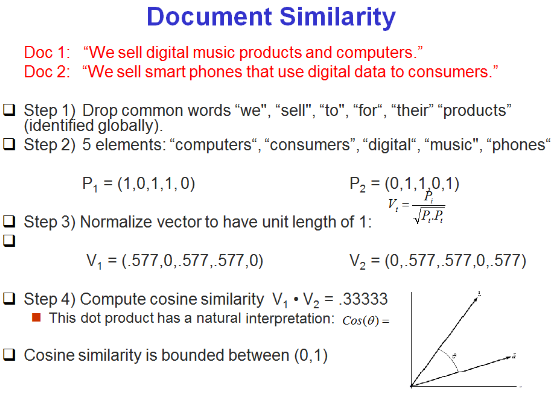
- Hoberg and Phillips (2016JPE): TNIC industries, endogenous industry boundaries.
- Hoberg and Phillips (2010RFS): Mergers, new product synergies.
- Hoberg, Phillips, and Prabhala (2014JF): Product market fluidity and competitive threat.
- Fresard, Hoberg and Phillips (2016WP): Vertical relatedness in TNIC space and links to mergers and innovation.
- See website www.marshall.usc.edu/industrydata (linked from my homepage at USC).

Parse 10-K Filings from the SEC


10-K
Item 1: Business Description
Item 7: MD&A
10-K Has many sections,
Business Description: for industry classification
Management's Discussion and Analysis (MD&A)

Extract Item 1 (Business Description) from every 10-K in each year.

Use Cosine Similarities



Document Similarity

Doc 1: “We sell digital music products and computers.”
Doc 2: “We sell smart phones that use digital data to consumers.”

- ❑ Step 1) Drop common words “we", “sell", “to", “for“, “their” “products” (identified globally).
- ❑ Step 2) 5 elements: “computers“, “consumers”, “digital“, “music", “phones“

$P_1 = (1,0,1,1,0)$ $P_2 = (0,1,1,0,1)$

- ❑ Step 3) Normalize vector to have unit length of 1: $V_i = \frac{P_i}{\sqrt{P_i . P_i}}$
- ❑

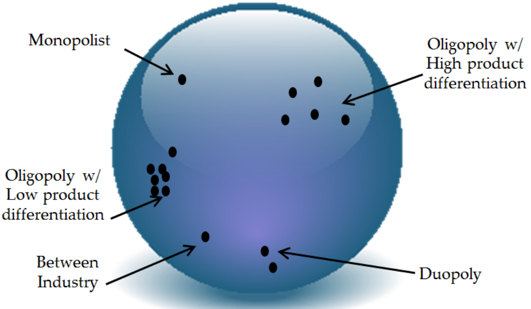
$V_1 = (.577,0,.577,.577,0)$ $V_2 = (0,.577,.577,0,.577)$

- ❑ Step 4) Compute cosine similarity $V_1 \bullet V_2 = .33333$
  - ■ This dot product has a natural interpretation: $Cos(\theta) =$
- ❑ Cosine similarity is bounded between (0,1)

Key: map firm text to vectors of length one in Euclidean space! Firms have product market “locations” on a unit sphere.

Product Market as a Sphere

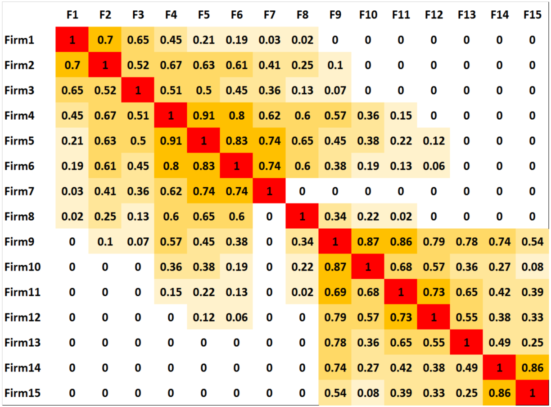
Product Market Space is a high dimensional Unit Sphere



Monopolist
Oligopoly w/ High product differentiation
Oligopoly w/ Low product differentiation
Between Industry
Duopoly

Every firm has a location on this sphere. So it has 5000+ public firms and is 80,000 dimensional. Firms residing in dense areas face high levels of competition. Firms in isolated parts of the space are differentiated and effective monopolies.

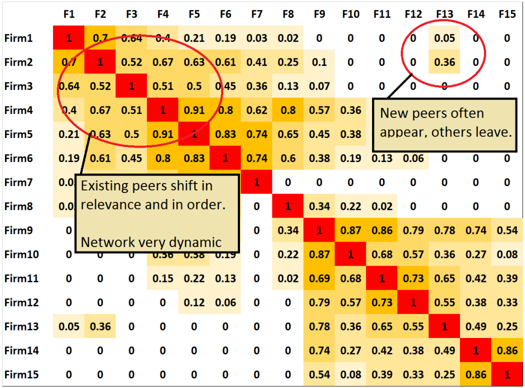
TNIC Industry Network Structure



	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15
Firm1	1	0.7	0.65	0.45	0.21	0.19	0.03	0.02	0	0	0	0	0	0	0
Firm2	0.7	1	0.52	0.67	0.63	0.61	0.41	0.25	0.1	0	0	0	0	0	0
Firm3	0.65	0.52	1	0.51	0.5	0.45	0.36	0.13	0.07	0	0	0	0	0	0
Firm4	0.45	0.67	0.51	1	0.91	0.8	0.62	0.6	0.57	0.36	0.15	0	0	0	0
Firm5	0.21	0.63	0.5	0.91	1	0.83	0.74	0.65	0.45	0.38	0.22	0.12	0	0	0
Firm6	0.19	0.61	0.45	0.8	0.83	1	0.74	0.6	0.38	0.19	0.13	0.06	0	0	0
Firm7	0.03	0.41	0.36	0.62	0.74	0.74	1	0	0	0	0	0	0	0	0
Firm8	0.02	0.25	0.13	0.6	0.65	0.6	0	1	0.34	0.22	0.02	0	0	0	0
Firm9	0	0.1	0.07	0.57	0.45	0.38	0	0.34	1	0.87	0.86	0.79	0.78	0.74	0.54
Firm10	0	0	0	0.36	0.38	0.19	0	0.22	0.87	1	0.68	0.57	0.36	0.27	0.08
Firm11	0	0	0	0.15	0.22	0.13	0	0.02	0.69	0.68	1	0.73	0.65	0.42	0.39
Firm12	0	0	0	0	0.12	0.06	0	0	0.79	0.57	0.73	1	0.55	0.38	0.33
Firm13	0	0	0	0	0	0	0	0	0.78	0.36	0.65	0.55	1	0.49	0.25
Firm14	0	0	0	0	0	0	0	0	0.74	0.27	0.42	0.38	0.49	1	0.86
Firm15	0	0	0	0	0	0	0	0	0.54	0.08	0.39	0.33	0.25	0.86	1

Network nearly unconstrained: it is intransitive and non-binary.
All values in $[0, 1]$ (partial product overlaps) are possible.

Consequence 1: **Can** model changing industries!



	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15
Firm1	1	0.7	0.64	0.4	0.21	0.19	0.03	0.02	0	0	0	0	0.05	0	0
Firm2	0.7	1	0.52	0.67	0.63	0.61	0.41	0.25	0.1	0	0	0	0.36	0	0
Firm3	0.64	0.52	1	0.51	0.5	0.45	0.36	0.13	0.07	0	0	0	0	0	0
Firm4	0.4	0.67	0.51	1	0.91	0.8	0.62	0.8	0.57	0.36	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Firm5	0.21	0.63	0.5	0.91	1	0.83	0.74	0.65	0.45	0.38	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Firm6	0.19	0.61	0.45	0.8	0.83	1	0.74	0.6	0.38	0.19	0.13	0.06	0	0	0
Firm7	0.0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	0	0	0	0	0	0	0	0
Firm8	0.0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	1	0.34	0.22	0.02	0	0	0	0
Firm9	0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0.34	1	0.87	0.86	0.79	0.78	0.74	0.54
Firm10	0	0	0	0.36	0.38	0.19	0	0.22	0.87	1	0.68	0.57	0.36	0.27	0.08
Firm11	0	0	0	0.15	0.22	0.13	0	0.02	0.69	0.68	1	0.73	0.65	0.42	0.39
Firm12	0	0	0	0	0.12	0.06	0	0	0.79	0.57	0.73	1	0.55	0.38	0.33
Firm13	0.05	0.36	0	0	0	0	0	0	0.78	0.36	0.65	0.55	1	0.49	0.25
Firm14	0	0	0	0	0	0	0	0	0.74	0.27	0.42	0.38	0.49	1	0.86
Firm15	0	0	0	0	0	0	0	0	0.54	0.08	0.39	0.33	0.25	0.86	1

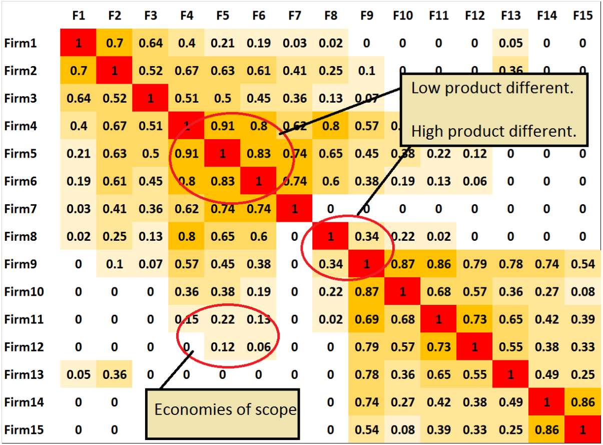
New peers often appear, others leave.

Existing peers shift in relevance and in order.

Network very dynamic

Network fully redrawn each year. Small and large changes observable. We find massive annual change. 30% of peers this year are not next year! SIC-code peers are nearly 100% persistent.

Consequence 2: **Can** model product diff. & entry threats



	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15
Firm1	1	0.7	0.64	0.4	0.21	0.19	0.03	0.02	0	0	0	0	0.05	0	0
Firm2	0.7	1	0.52	0.67	0.63	0.61	0.41	0.25	0.1	0	0	0	0.36	0	0
Firm3	0.64	0.52	1	0.51	0.5	0.45	0.36	0.13	0.07	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Firm4	0.4	0.67	0.51	1	0.91	0.8	0.62	0.8	0.57	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Firm5	0.21	0.63	0.5	0.91	1	0.83	0.74	0.65	0.45	0.38	0.22	0.12	0	0	0
Firm6	0.19	0.61	0.45	0.8	0.83	1	0.74	0.6	0.38	0.19	0.13	0.06	0	0	0
Firm7	0.03	0.41	0.36	0.62	0.74	0.74	1	0	0	0	0	0	0	0	0
Firm8	0.02	0.25	0.13	0.8	0.65	0.6	0	1	0.34	0.22	0.02	0	0	0	0
Firm9	0	0.1	0.07	0.57	0.45	0.38	0	0.34	1	0.87	0.86	0.79	0.78	0.74	0.54
Firm10	0	0	0	0.36	0.38	0.19	0	0.22	0.87	1	0.68	0.57	0.36	0.27	0.08
Firm11	0	0	0	0.15	0.22	0.13	0	0.02	0.69	0.68	1	0.73	0.65	0.42	0.39
Firm12	0	0	0	0	0.12	0.06	0	0	0.79	0.57	0.73	1	0.55	0.38	0.33
Firm13	0.05	0.36	0	0	0	0	0	0	0.78	0.36	0.65	0.55	1	0.49	0.25
Firm14	0	0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0.74	0.27	0.42	0.38	0.49	1	0.86
Firm15	0	0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0.54	0.08	0.39	0.33	0.25	0.86	1

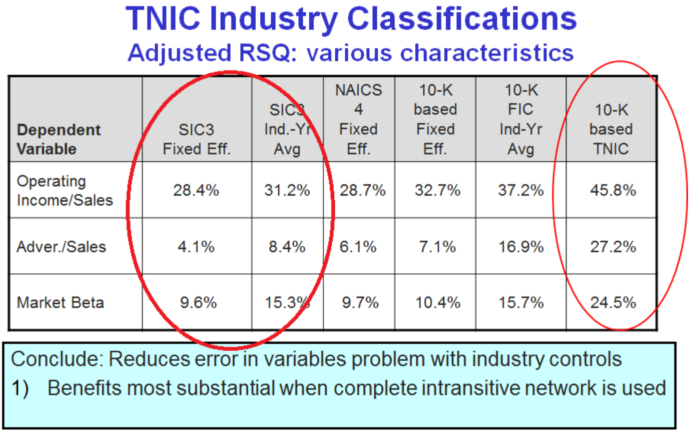
Low product different.

High product different.

Economies of scope

Complete information about product differentiation.
Positive but lower similarities indicate scope for entry & synergies.

Validation is Strong. Superior signal on profitability and stock returns



TNIC Industry Classifications

Adjusted RSQ: various characteristics

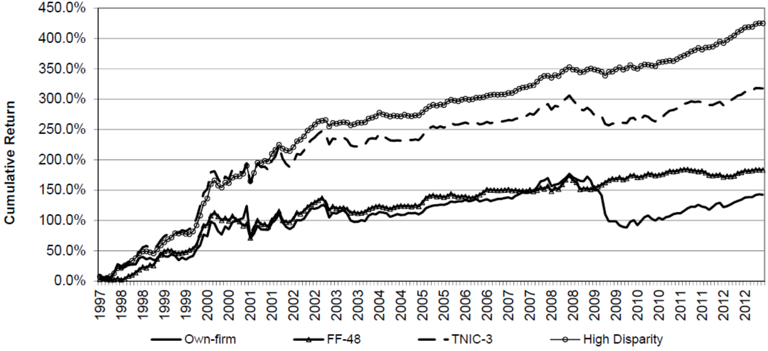
Dependent Variable	SIC3 Fixed Eff.	SIC3 Ind.-Yr Avg	NAICS 4 Fixed Eff.	10-K based Fixed Eff.	10-K FIC Ind-Yr Avg	10-K based TNIC
Operating Income/Sales	28.4%	31.2%	28.7%	32.7%	37.2%	45.8%
Adver./Sales	4.1%	8.4%	6.1%	7.1%	16.9%	27.2%
Market Beta	9.6%	15.3%	9.7%	10.4%	15.7%	24.5%

Conclude: Reduces error in variables problem with industry controls

1) Benefits most substantial when complete intransitive network is used

Not only is TNIC more general in modeling of market structure. It is 50% more informative than SIC! Validation is critical.

Much Stronger Momentum Stock Returns



*SIC-code momentum weak out after Moskowitz and Grinblatt (1999).
*Even basic TNIC momentum strategies reliably profitable.
*See Hoberg and Phillips JFQA (forthcoming).

Section III: Market Structure Research Impact and Uses

Section I: Market Structure PAST

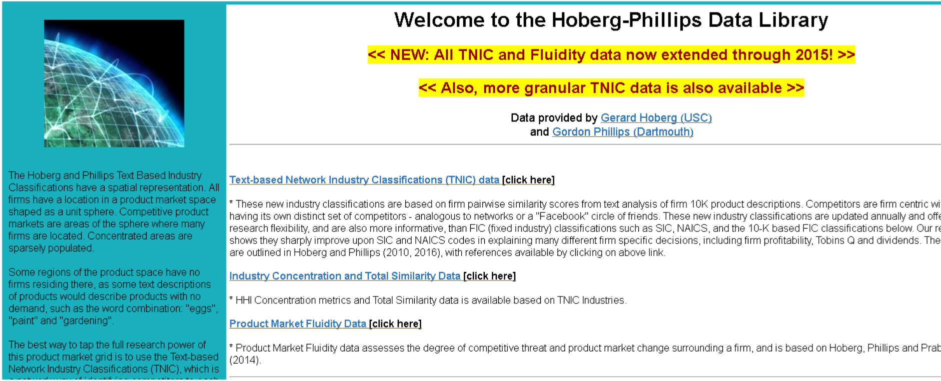
Section II: Market Structure PRESENT

Section III: Impact and Uses

Section IV: Market Structure FUTURE

Section V: Using Text to Model Innovation

TNIC data available on the web since 2010



The Hoberg and Phillips Text Based Industry Classifications have a spatial representation. All firms have a location in a product market space shaped as a unit sphere. Competitive product markets are areas of the sphere where many firms are located. Concentrated areas are sparsely populated.

Some regions of the product space have no firms residing there, as some text descriptions of products would describe products with no demand, such as the word combination: "eggs", "paint" and "gardening".

The best way to tap the full research power of this product market grid is to use the Text-based Network Industry Classifications (TNIC), which is

Welcome to the Hoberg-Phillips Data Library

<< NEW: All TNIC and Fluidity data now extended through 2015! >>

<< Also, more granular TNIC data is also available >>

Data provided by Gerard Hoberg (USC)
and Gordon Phillips (Dartmouth)

Text-based Network Industry Classifications (TNIC) data [click here]

* These new industry classifications are based on firm pairwise similarity scores from text analysis of firm 10K product descriptions. Competitors are firm centric wi having its own distinct set of competitors - analogous to networks or a "Facebook" circle of friends. These new industry classifications are updated annually and offe research flexibility, and are also more informative, than FIC (fixed industry) classifications such as SIC, NAICS, and the 10-K based FIC classifications below. Our re shows they sharply improve upon SIC and NAICS codes in explaining many different firm specific decisions, including firm profitability, Tobins Q and dividends. The are outlined in Hoberg and Phillips (2010, 2016), with references available by clicking on above link.

Industry Concentration and Total Similarity Data [click here]

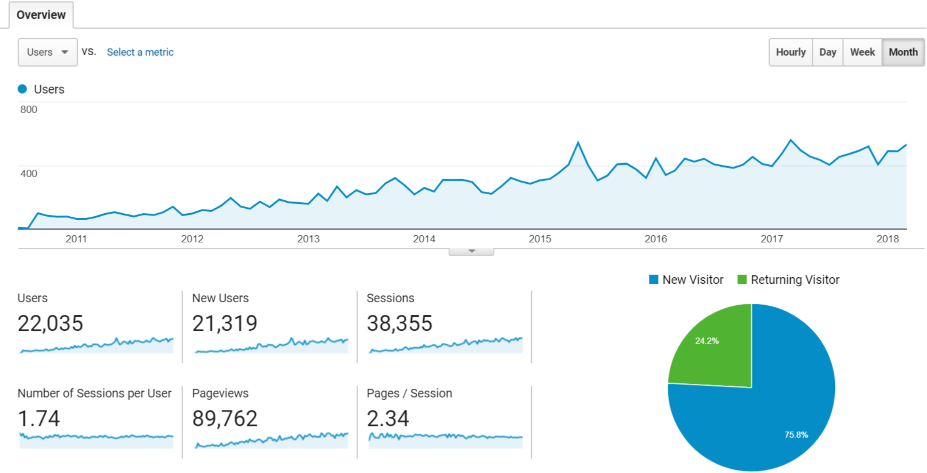
* HHI Concentration metrics and Total Similarity data is available based on TNIC Industries.

Product Market Fluidity Data [click here]

* Product Market Fluidity data assesses the degree of competitive threat and product market change surrounding a firm, and is based on Hoberg, Phillips and Prab (2014).

Now at: www.marshall.usc.edu/industrydata

Monthly volume steadily increasing



Users ▾ vs. Select a metric

Users
22,035

New Users
21,319

Sessions
38,355

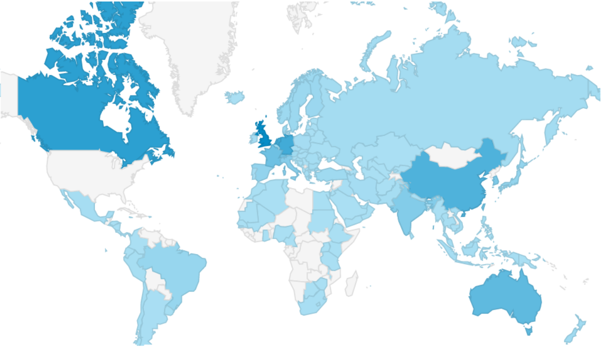
Number of Sessions per User
1.74

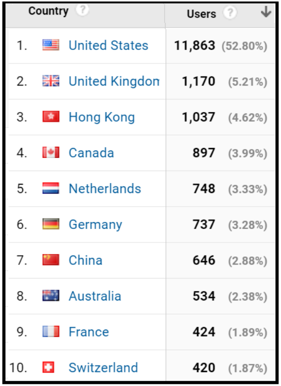
Pageviews
89,762

Pages / Session
2.34

Now roughly 500 visitors per month. 22,000 total since inception.

Users are global!

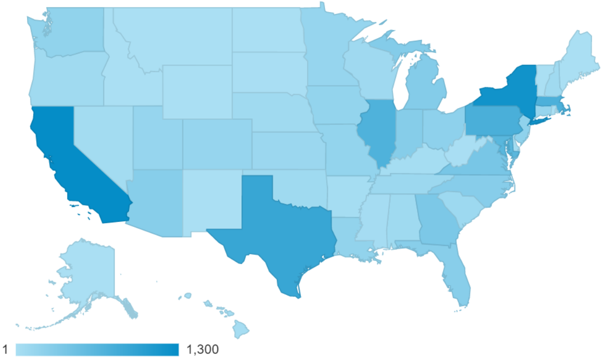




	Country	Users	
1.	United States	11,863	(52.80%)
2.	United Kingdon	1,170	(5.21%)
3.	Hong Kong	1,037	(4.62%)
4.	Canada	897	(3.99%)
5.	Netherlands	748	(3.33%)
6.	Germany	737	(3.28%)
7.	China	646	(2.88%)
8.	Australia	534	(2.38%)
9.	France	424	(1.89%)
10.	Switzerland	420	(1.87%)

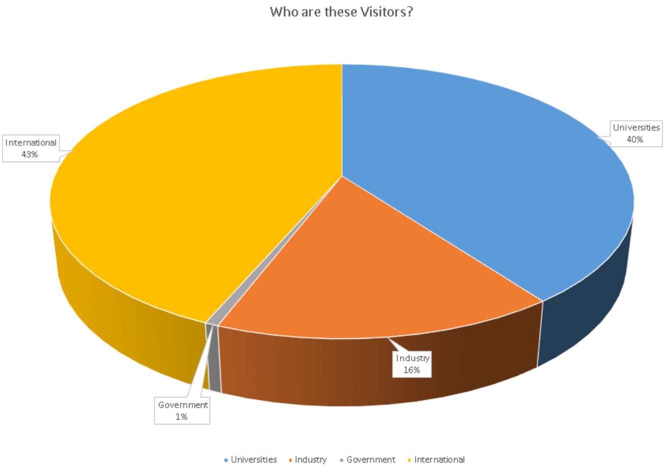
*Switzerland is ranked 99th in population but ranks 10th on TNIC website visits! That beats Japan, Italy, and India!

Visitors hail from all 50 states



Yes, even North Dakota!

Who are these users?



* Not just for academic research!

* Industry users mostly I-banks, hedge funds, & tech firms.

Section IV: Market Structure FUTURE

Section I: Market Structure PAST

Section II: Market Structure PRESENT

Section III: Impact and Uses

Section IV: Market Structure FUTURE

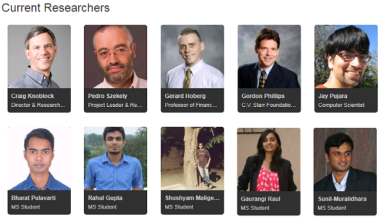
Section V: Using Text to Model Innovation

Maximizing TNIC's Signal w/ Recent NLP Advances

- Semantic relatedness can improve power beyond cosine similarities.
- Embedding technologies showing strong promise in pilot study.
- Latent Dirichlet Allocation not recommended for economic reasons.

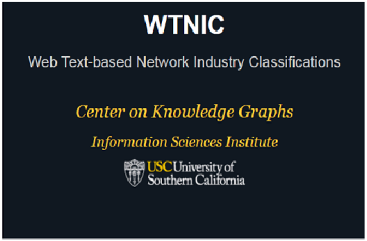
* New improved TNIC models coming soon. Stay tuned!

Major Expansion to Private Firms



2 PI's from B-School
2 PI's from Comp Sci (Viterbi)
1 Newly added Viterbi PhD

5 Masters Students in Comp Sci
1 Undergraduate RA



WTNIC

Web Text-based Network Industry Classifications

Center on Knowledge Graphs

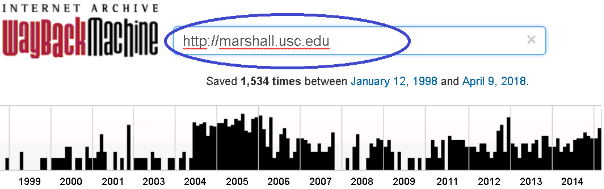
Information Sciences Institute

USC University of Southern California

We thank the National Science Foundation!
$500,000 Grant
Joint B-School & USC's Viterbi School (ISI)

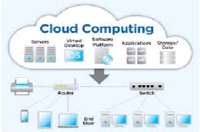
Joint NSF-funded research b/t B-School and Comp Sci Researchers.

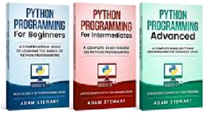
Get Annual Snapshots of 800,000 Private Firm Websites



http://marshall.usc.edu

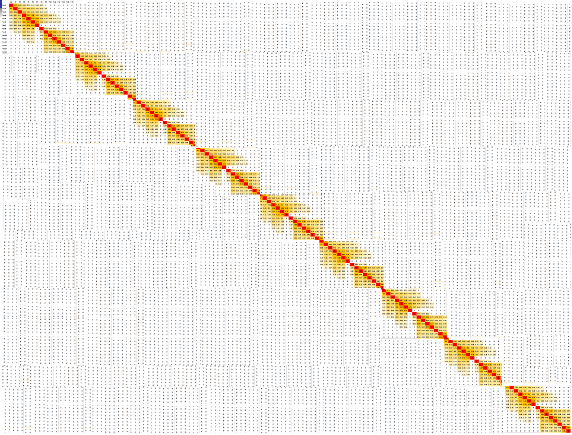
Saved **1,534 times** between January 12, 1998 and April 9, 2018.




PYTHON PROGRAMMING For Beginners
PYTHON PROGRAMMING For Intermediates
PYTHON PROGRAMMING Advanced
ADAM STEWART

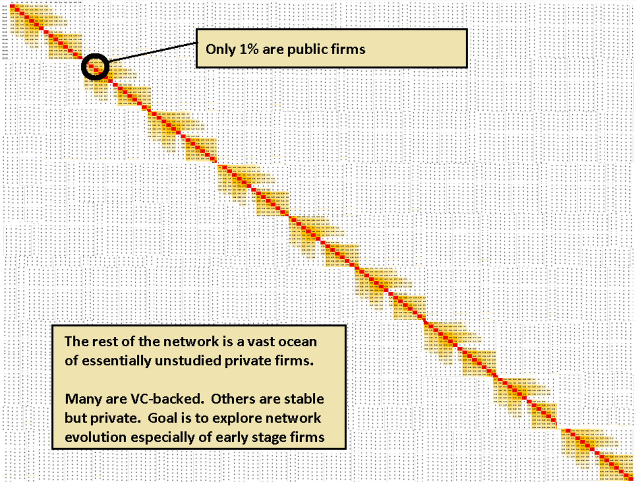
(1) Gather and process web pages, (2) extract product market text, (3) store on cloud, and (4) code up scaleable NLP to compute network.

Very large network



* 18 years and 800,000 firms. Time varying and intransitive non-binary network.
* Use of standard similarity methods not feasible.
* Use scaleable sparse technologies and dimensionality reductions (embeddings).

Private firms are just very numerous



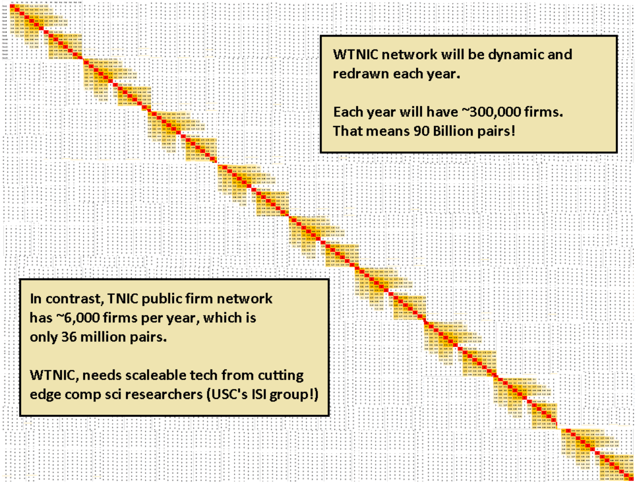
Only 1% are public firms

The rest of the network is a vast ocean of essentially unstudied private firms.

Many are VC-backed. Others are stable but private. Goal is to explore network evolution especially of early stage firms

Standard TNIC will be dwarfed by WTNIC.
Understanding private firm entry and evolution invaluable to corporate managers, investors, and researchers.

Excellent Progress on WTNIC Execution



WTNIC network will be dynamic and redrawn each year.

Each year will have ~300,000 firms. That means 90 Billion pairs!

In contrast, TNIC public firm network has ~6,000 firms per year, which is only 36 million pairs.

WTNIC, needs scaleable tech from cutting edge comp sci researchers (USC's ISI group!)

Alpha version of network only 3-6 months away!

Section V: Using Text to Model Innovation

Section I: Market Structure PAST

Section II: Market Structure PRESENT

Section III: Impact and Uses

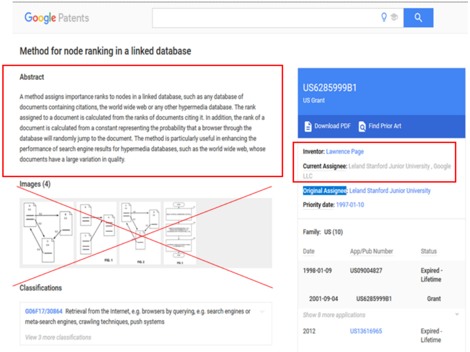
Section IV: Market Structure FUTURE

Section V: Using Text to Model Innovation

Innovation: Patent Markets

Bowen, Fresard and Hoberg (2018)

- > 8 million patents since 1900. Text structure: abstract – description – claim



Method for node ranking in a linked database

Abstract

A method assigns importance ranks to nodes in a linked database, such as any database of documents containing citations, the world wide web or any other hypermedia database. The rank assigned to a document is calculated from the ranks of documents citing it. In addition, the rank of a document is calculated from a constant representing the probability that a browser through the database will randomly jump to the document. The method is particularly useful in enhancing the performance of search engine results for hypermedia databases, such as the world wide web, whose documents have a large variation in quality.

Images (4)

Classifications

G06F17/30864 Retrieval from the Internet, e.g. browsers by querying, e.g. search engines or meta-search engines, crawling techniques, push systems

View 3 more classifications

US6285999B1
US Grant

Inventor: Lawrence Page

Current Assignee: Leland Stanford Junior University , Google LLC

Original Assignee: Leland Stanford Junior University

Priority date: 1997-01-10

Family: US (10)

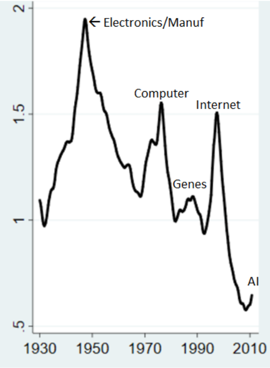
Date	App/Pub Number	Status
1998-01-09	US09004827	Expired - Lifetime
2001-09-04	US6285999B1	Grant
Show 8 more applications		
2012	US13616965	Expired - Lifetime

*Firms compete to develop technology. Rich textual form indicates a spherical mapping similar to TNIC.
*Highly volatile data structure, but crucial to our economic future.

Technological Disruption Declining since WW II?

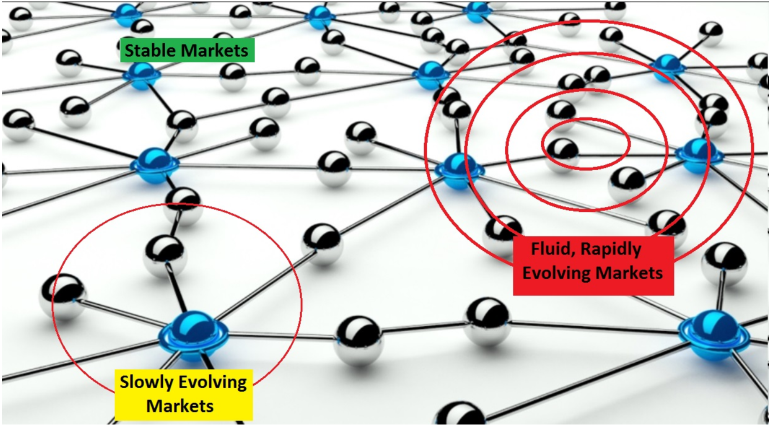
Breakthrough innovations Declining?

Technological Disruptiveness (rapidly growing vocab) vs Time



- Many temporary spikes
- They don’t persist!
- Pervasive low frequency decline since WW II
- BFH (2018): this explains part of decline in IPOs

Technology Development vs Product Development



*HPP (2014JF) examines rate of change of vocabulary in firm 10-Ks (product market fluidity).
*Increased fluidity indicates more agile rivals and direct threats.
*Firms respond by saving liquidity (more cash, fewer dividends).

Conclusions

- Textual Network Industries (TNIC) was born on a PC in 2006.
- Stage 1 model of U.S. public firms completed in late 2008.
- Much completed research done following stage 1.
- Stage 2 expansion to U.S. private firms in 3-6 months.
- Work under way using same technologies on innovation data.
- Implications for optimizing corporate strategy, investors, researchers, regulators and more.