

TOWARD A BETTER MEASURE OF BUSINESS PROXIMITY: TOPIC MODELING FOR INDUSTRY INTELLIGENCE

Zhan (Michael) Shi

Department of Information Systems, W. P. Carey School of Business, Arizona State University,
Tempe, AZ 85287-4606 U.S.A. {zmshi@asu.edu}

Gene Moo Lee

Department of Information Systems and Operations Management, College of Business, The University of Texas at Arlington,
Arlington, TX 76019 U.S.A. {gene.lee@uta.edu}

Andrew B. Whinston

Department of Information, Risk, and Operations Management, McCombs School of Business, The University of Texas at Austin,
Austin, TX 78712 U.S.A. {abw@uts.cc.utexas.edu}

Appendix

Additional Tables

Table A. ERGM Notations	
Network Graph	
Y, Y_{ij}	a random network graph matrix, its ij element
Y_{-ij}	all elements except ij
y	the set of all possible graphs for a fixed set of notes y, y_{ij}
y, y_{ij}	a realization of the random network graph and its ij element
$z_i(y)$	a statistic of network graph y
Network Statistics	
t	total number of edges
d_2	number of nodes which have at least 2 edges
h_s^{sta}	number of edges within state s
h_c^{cat}	number of edges within category c
p_g	sum of geographic proximity over all edges
p_s	sum of social proximity over all edges
p_i	sum of investor proximity over all edges
p_b	sum of business proximity over all edges
Nodal Characteristics:	
s_i	state where i 's headquarters is located
c_i	category to which i belongs
Dyadic Characteristics	
$p_{g,ij}$	geographic proximity of i and j
$p_{s,ij}$	social proximity of i and j
$p_{i,ij}$	investor proximity of i and j
$p_{b,ij}$	business proximity of i and j

Table A2. Model Coefficients from Sample 1							
	Coeff.	S.E.	p-value		Coeff.	S.E.	p-value
Geographic	-0.2699	0.3440	0.4326	NY	-	-	-
Social	0.0532	0.0108	0.0000	OH	-	-	-
Investor	0.0270	0.0522	0.6049	OK	-	-	-
Business	0.4635	0.1378	0.0008	OR	-	-	-
Edges	-12.5625	3.7908	0.0009	PA	-	-	-
Degree > 2	2.4820	0.6438	0.0001	RI	-	-	-
				SC	-	-	-
				SD	-	-	-
State				TN	-	-	-
AL	-	-	-	TX	-	-	-
AR	-	-	-	UT	-	-	-
AZ	-	-	-	VA	-	-	-
CA	2.3899	0.8178	0.0035	VT	-	-	-
CO	-	-	-	WA	-	-	-
CT	-	-	-	WI	-	-	-
DC	-	-	-	WV	-	-	-
DE	-	-	-	WY	-	-	-
FL	-	-	-				
GA	-	-	-				
HI	-	-	-	Category			
IA	-	-	-	advertising	-	-	-
ID	-	-	-	biotech	-	-	-
IL	-	-	-	cleantech	-	-	-
IN	-	-	-	consulting	-	-	-
KS	-	-	-	ecommerce	-	-	-
KY	-	-	-	education	-	-	-
LA	-	-	-	enterprise	2.9201	0.8882	0.0010
MA	4.6361	1.1201	0.0000	games video	3.0284	1.0953	0.0057
MD	-	-	-	hardware	3.7045	1.7912	0.0386
MN	-	-	-	legal	-	-	-
MO	-	-	-	mobile	1.8611	1.2047	0.1223
MS	-	-	-	network hosting	-	-	-
MT	-	-	-	other	-	-	-
NC	-	-	-	public relations	-	-	-
NE	-	-	-	search	-	-	-
NH	9.7899	1.5931	0.0000	security	-	-	-
NJ	5.6899	1.6428	0.0005	semiconductor	-	-	-
NM	-	-	-	software	-	-	-
NV	-	-	-	web	-0.9020	2.1375	0.6721

Table A3. Category-Based Selective Mixing Coefficients (100 Samples): Equation (10) Excluding $\theta_b p_b$

	Number of Samples with Coefficient	Number of Samples Coefficient > 0	Number of Samples p -value < 1.0%		Number of Samples with Coefficient	Number of Samples Coefficient > 0	Number of Samples p -value < 1.0%
advertising	28	38	14	mobile	27	27	16
biotech	37	37	32	net hosting	8	8	6
cleantech	12	12	10	other	0	–	–
consulting	12	12	9	pub rel	10	10	6
ecommerce	12	12	6	search	0	–	–
education	0	–	–	security	0	–	–
enterprise	22	22	20	semiconductor	17	17	14
games video	28	28	16	software	90	85	55
hardware	31	31	29	web	78	70	22
legal	0	–	–				

Table A4. LDA Results of CrunchBase Data

Topic	Dimension	Top 5 Words
1	Product	video, music, digital, entertainment, artists
2	Product	news, site, blog, articles, publishing
3	Product	job, jobs, search employers, career
4	Product	people, community, members, share, friends
5	Product	facebook, friends, share, twitter, photos
6	Product	energy, power, solar, systems, water
7	Product	systems, design, applications, devices, semiconductor
8	Product	consulting, clients, support, systems experience
9	Product	event, sports, events, fans, tickets
10	Product	insurance, financial, credit, tax mortgage
11	Product	deals, shopping, consumers, local, retailers
12	Product	health, care, medical, healthcare, patient
13	Product	students, learning, education, college, school
14	Product	food, restaurants, fitness, restaurant, pet
15	Product	investment, financial, investors, capital, trading
16	Product	advertising, publishers, advertisers, brands, digital
17	Product	manage, project, documents, document, tools
18	Product	treatment, medical, research, clinical, diseases
19	Product	games, game, gaming, virtual, entertainment
20	Product	security, compliance, secure, protection, access
21	Product	search, engine, website, seo, optimization
22	Product	search, user, engine, results, relevant
23	Product	fashion, art, brands, custom, design
24	Product	equipment, repair, car, home, accessories
25	Product	law, legal, government, public, federal
26	Product	analytics, research, analysis, intelligence, performance
27	Product	travel, travelers, vacation, hotel, hotels
28	Product	real, estate, home, buyers, property
29	Product	payment, card, cards, credit, payments
30	Technology/Product	phone, email, text, voice, messaging
31	Technology/Product	wireless, networks, communications, internet, providers
32	Technology/Product	cloud, storage, hosting, server, servers

Table A4. LDA Results of CrunchBase Data		
Topic	Dimension	Top 5 Words
33	Technology/Product	app, apps, iphone, android, applications
34	Technology/Product	design, applications, application, custom, website
35	Technology/Product	site, website, free, allows, user
36	Technology/Product	testing, test, monitoring, tracking, performance
37	Market/Technology	digital, clients, brand, agency, design
38	Market	sales, customer, lead, email, leads
39	Market	solution, cost, costs, applications, enterprise
40	Market	organization, community, support, organization, businesses
41	Market	make, people, time, just, way
42	Market	quality, customer, needs, clients, provide
43	Market	systems, operates, headquartered, subsidiary, serves
44	Market	united, states, offices, america, europe
45	Market	san, york, city, california, francisco
46	Market	award, magazine, awards, bst, world
47	Market	million, world, leading, largest, global
48	Market/Team	team, experience, industry, world, market
49	Team	partners, ventures, capital, including, san
50	Team	launched, million, product, ceo, acquired