The Network of Firms Implied by the News

Gustavo Schwenkler
Santa Clara University

Hannan Zheng
Man Numeric

Open Source Quantitative Finance Research Conference
April 2025

Firm networks matter, but data are scarce

- Acemoglu et al. (2012): *"This paper argues that, in the presence of intersectoral input–output linkages, microeconomic idiosyncratic shocks may lead to aggregate fluctuations."*
- Herskovic (2018): *"Changes in the [production] network are sources of systematic risk reflected in equilibrium asset prices."*

However, existing databases are incomplete, sparse, or lagged

- Compustat segments, BEA input-output, variance decomposition, interbank links, alternative data (10-K similarity, internet co-searches)

$\rightarrow$ Lack of network data hinders accurate measurement of risks

Our solution: mine news reporting

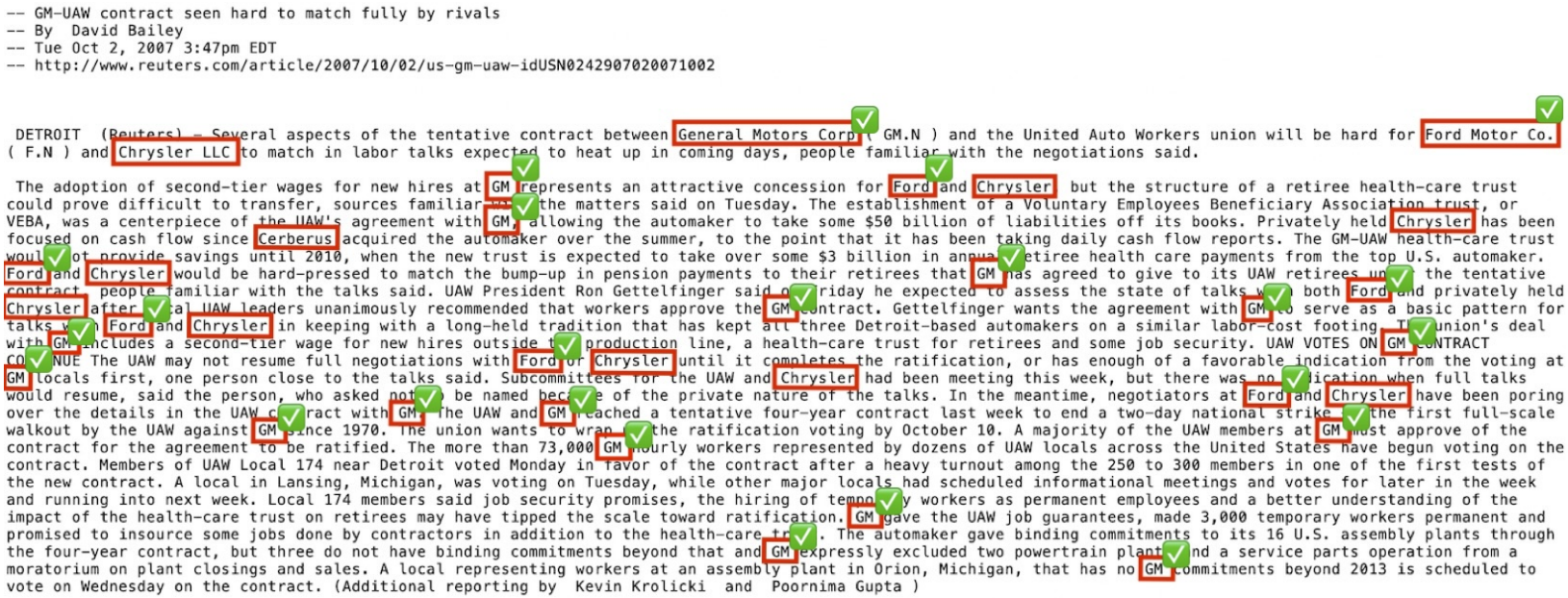
-- GM-UAW contract seen hard to match fully by rivals
-- By David Bailey
-- Tue Oct 2, 2007 3:47pm EDT
-- http://www.reuters.com/article/2007/10/02/us-gm-uaw-idUSN0242907020071002

DETROIT (Reuters) - Several aspects of the tentative contract between General Motors Corp (GM.N) and the United Auto Workers union will be hard for Ford Motor Co. (F.N) and Chrysler LLC to match in labor talks expected to heat up in coming days, people familiar with the negotiations said.

The adoption of second-tier wages for new hires at GM represents an attractive concession for Ford and Chrysler, but the structure of a retiree health-care trust could prove difficult to transfer, sources familiar with the matters said on Tuesday. The establishment of a Voluntary Employees Beneficiary Association trust, or VEBA, was a centerpiece of the UAW's agreement with GM, allowing the automaker to take some $50 billion of liabilities off its books. Privately held Chrysler has been focused on cash flow since Cerberus acquired the automaker over the summer, to the point that it has been taking daily cash flow reports. The GM-UAW health-care trust would not provide savings until 2010, when the new trust is expected to take over some $3 billion in annual retiree health care payments from the top U.S. automaker. Ford and Chrysler would be hard-pressed to match the bump-up in pension payments to their retirees that GM has agreed to give to its UAW retirees under the tentative contract, people familiar with the talks said. UAW President Ron Gettelfinger said on Friday he expected to assess the state of talks with both Ford and privately held Chrysler after local UAW leaders unanimously recommended that workers approve the GM contract. Gettelfinger wants the agreement with GM to serve as a basic pattern for talks with Ford and Chrysler in keeping with a long-held tradition that has kept all three Detroit-based automakers on a similar labor-cost footing. The union's deal with GM includes a second-tier wage for new hires outside the production line, a health-care trust for retirees and some job security. UAW VOTES ON GM CONTRACT CONTINUE The UAW may not resume full negotiations with Ford or Chrysler until it completes the ratification, or has enough of a favorable indication from the voting at GM locals first, one person close to the talks said. Subcommittees for the UAW and Chrysler had been meeting this week, but there was no indication when full talks would resume, said the person, who asked not to be named because of the private nature of the talks. In the meantime, negotiators at Ford and Chrysler have been poring over the details in the UAW contract with GM. The UAW and GM reached a tentative four-year contract last week to end a two-day national strike -- the first full-scale walkout by the UAW against GM since 1970. The union wants to wrap up the ratification voting by October 10. A majority of the UAW members at GM must approve of the contract for the agreement to be ratified. The more than 73,000 GM hourly workers represented by dozens of UAW locals across the United States have begun voting on the contract. Members of UAW Local 174 near Detroit voted Monday in favor of the contract after a heavy turnout among the 250 to 300 members in one of the first tests of the new contract. A local in Lansing, Michigan, was voting on Tuesday, while other major locals had scheduled informational meetings and votes for later in the week and running into next week. Local 174 members said job security promises, the hiring of temporary workers as permanent employees and a better understanding of the impact of the health-care trust on retirees may have tipped the scale toward ratification. GM gave the UAW job guarantees, made 3,000 temporary workers permanent and promised to insource some jobs done by contractors in addition to the health-care trust. The automaker gave binding commitments to its 16 U.S. assembly plants through the four-year contract, but three do not have binding commitments beyond that and GM expressly excluded two powertrain plants and a service parts operation from a moratorium on plant closings and sales. A local representing workers at an assembly plant in Orion, Michigan, that has no GM commitments beyond 2013 is scheduled to vote on Wednesday on the contract. (Additional reporting by Kevin Krolicki and Poornima Gupta)

- News often reports about relationships between firms

Our methodology



-- GM-UAW contract seen hard to match fully by rivals
-- By David Bailey
-- Tue Oct 2, 2007 3:47pm EDT
-- http://www.reuters.com/article/2007/10/02/us-gm-uaw-idUSN0242907020071002

DETROIT (Reuters) - Several aspects of the tentative contract between General Motors Corp (GM.N) and the United Auto Workers union will be hard for Ford Motor Co. (F.N) and Chrysler LLC to match in labor talks expected to heat up in coming days, people familiar with the negotiations said.

The adoption of second-tier wages for new hires at GM represents an attractive concession for Ford and Chrysler but the structure of a retiree health-care trust could prove difficult to transfer, sources familiar with the matters said on Tuesday. The establishment of a Voluntary Employees Beneficiary Association trust, or VEBA, was a centerpiece of the UAW's agreement with GM, allowing the automaker to take some $50 billion of liabilities off its books. Privately held Chrysler has been focused on cash flow since Cerberus acquired the automaker over the summer, to the point that it has been taking daily cash flow reports. The GM-UAW health-care trust would not provide savings until 2010, when the new trust is expected to take over some $3 billion in annual retiree health care payments from the top U.S. automaker. Ford and Chrysler would be hard-pressed to match the bump-up in pension payments to their retirees that GM has agreed to give to its UAW retirees under the tentative contract, people familiar with the talks said. UAW President Ron Gettelfinger said on Friday he expected to assess the state of talks with both Ford and privately held Chrysler after local UAW leaders unanimously recommended that workers approve the GM contract. Gettelfinger wants the agreement with GM to serve as a basic pattern for talks with Ford and Chrysler in keeping with a long-held tradition that has kept all three Detroit-based automakers on a similar labor-cost footing. The union's deal with GM includes a second-tier wage for new hires outside the production line, a health-care trust for retirees and some job security. UAW VOTES ON GM CONTRACT CONTINUE The UAW may not resume full negotiations with Ford or Chrysler until it completes the ratification, or has enough of a favorable indication from the voting at GM locals first, one person close to the talks said. Subcommittees for the UAW and Chrysler had been meeting this week, but there was no indication when full talks would resume, said the person, who asked not to be named because of the private nature of the talks. In the meantime, negotiators at Ford and Chrysler have been poring over the details in the UAW contract with GM. The UAW and GM reached a tentative four-year contract last week to end a two-day national strike, the first full-scale walkout by the UAW against GM since 1970. The union wants to wrap the ratification voting by October 10. A majority of the UAW members at GM must approve of the contract for the agreement to be ratified. The more than 73,000 GM hourly workers represented by dozens of UAW locals across the United States have begun voting on the contract. Members of UAW Local 174 near Detroit voted Monday in favor of the contract after a heavy turnout among the 250 to 300 members in one of the first tests of the new contract. A local in Lansing, Michigan, was voting on Tuesday, while other major locals had scheduled informational meetings and votes for later in the week and running into next week. Local 174 members said job security promises, the hiring of temporary workers as permanent employees and a better understanding of the impact of the health-care trust on retirees may have tipped the scale toward ratification. GM gave the UAW job guarantees, made 3,000 temporary workers permanent and promised to insource some jobs done by contractors in addition to the health-care trust. The automaker gave binding commitments to its 16 U.S. assembly plants through the four-year contract, but three do not have binding commitments beyond that and GM expressly excluded two powertrain plants and a service parts operation from a moratorium on plant closings and sales. A local representing workers at an assembly plant in Orion, Michigan, that has no GM commitments beyond 2013 is scheduled to vote on Wednesday on the contract. (Additional reporting by Kevin Krolicki and Poornima Gupta)

- Identify mentions of firms in text data and establish link when two firms are mentioned in same sentence

Firm & link identification approach

1. Named Entity Recognition (spaCy in R) to identify mentions of *entities*: date, duration, location, money, number, ordinal, organization, percent, person, time, and other.
2. Keep organizations, filter out non-firms, and cluster same firms with unique name stem
3. Match with known firms from CRSP database
4. Establish link when firms are mentioned in same sentence
   - Can be relaxed to article co-mentions, with similar results

All codes available at `www.news-networks.net`

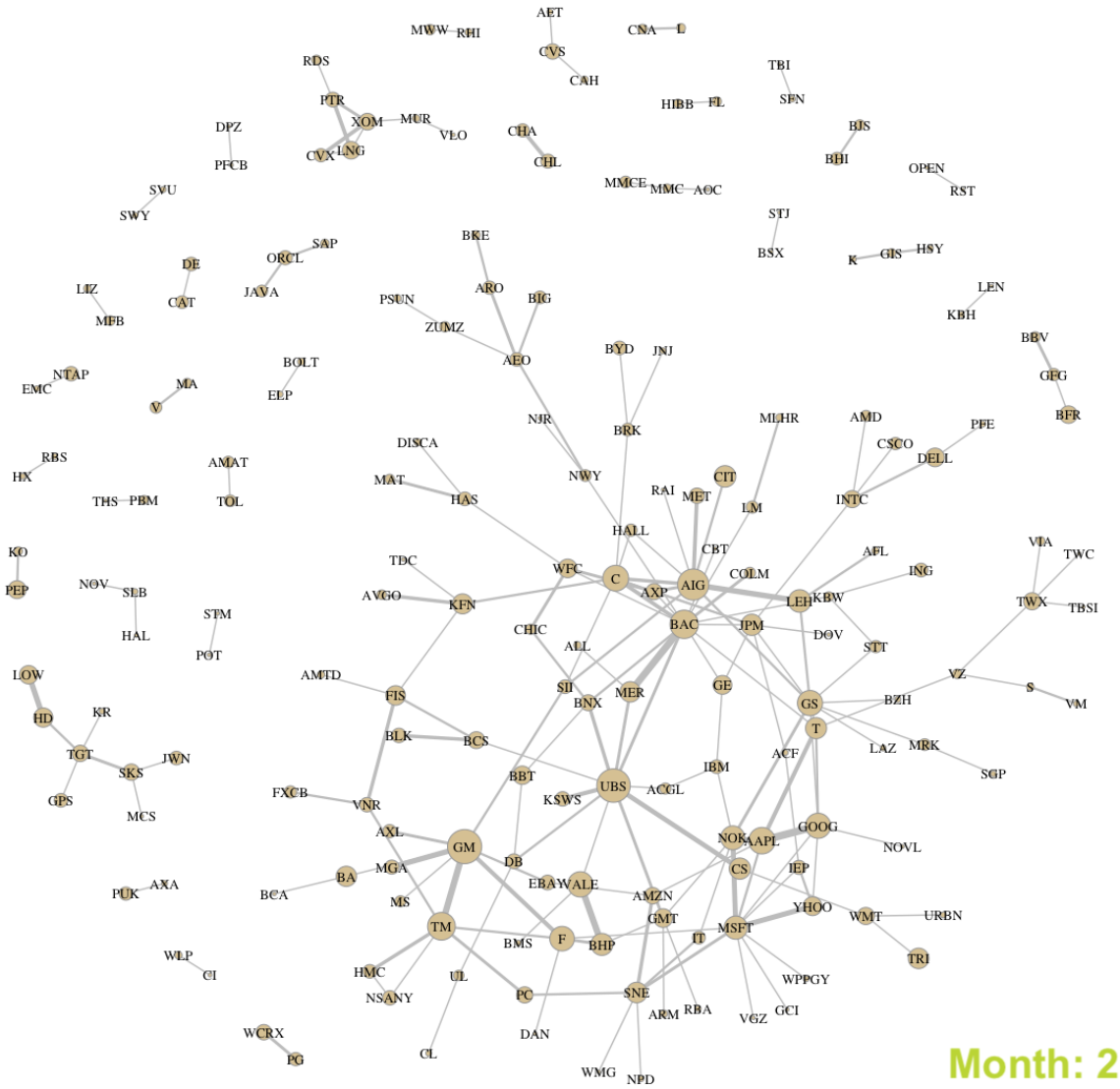
Data and output

- We apply methodology to two news sample:
  - 106,521 articles from Reuters (2006-10 through 2013-11)
  - 430,770 articles from NYT (1981-01 through 2023-12)
- We identify 40,000 links among 13,500 firms in NYT sample
  - All data available at `www.news-networks.net`

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	permco	mentions	firm_name	permno	ticker	stems	stems_short	mentioned_names											
2	20799	604	GENERAL MO	12079	GM	_general_mo	_gm_	_general_motors_|_carmaker_general_motors_|_general_motors_company_|_general_motors_gmac_										
3	37504	529	U B S AG	88239	UBS	_ubs_	_ubs_	_ubs_|_ubs_securities_|_global_swiss_bank_ubs_|_ubs_wealth_management_americas_										
4	137	342	AMERICAN IN	66800	AIG	_american_ir	_aig_	_american_international_group_|_aig_|_aig_financial_products_|_aig_financial_advisors_|_financials_shine_aig_										
5	3151	203	BANK OF AMI	59408	BAC	_bank_of_am	_bofa_	_bank_of_america_|_bank_of_america_chief_										
6	4521	178	TOYOTA MOT	76655	TM	_toyota_	_toyota_	_toyota_motor_|_toyota_|_brademann_toyota_										
7	7	143	APPLE INC	14593	AAPL	_apple_	_apple_	_apple_|_google_apple_|_apple_win_smartphone_share_|_apple_iphone_|_apple_research_motion_										
8	20483	123	CITIGROUP II	70519	C	_citi_	_citi_	_citi_|_citi_futures_perspective_|_citi_american_international_group_|_citi_private_equity_|_citi_bank_america_										
9	37665	115	COMPANHIA	89330	RIO	_rio_	_rio_|_vale_	_rio_|_rio_tinto_|_chinalcos_rio_tinto_										
10	35048	112	GOLDMAN S/	86868	GS	_goldman_sa	_goldman_sa	_goldman_sachs_|_goldman_sachs_group_|_goldman_sachs_group_chief_executive_										
11	45483	110	GOOGLE INC	90319	GOOG	_google_	_google_|_go	_google_|_google_voice_|_google_voice_iphone_|_google_desktop_|_google_chrome_|_google_chrome_operating_system_|_google_maps_|_google_owned_										
12	20750	99	FORD MOTO	25785	F	_ford_	_ford_	_ford_motor_|_ford_|_ford_explorer_|_texas_motors_ford_|_ford_fusion_|_ford_escape_|_ford_focus_										
13	30453	88	NOKIA CORP	87128	NOK	_nokia_	_nokia_	_nokia_|_nokia_executive_|_nokia_booklet_3_|_nokia_booklet_|_nokia_world_|_nokia_money_|_non_nokia_										
14	8048	77	MICROSOFT	10107	MSFT	_microsoft_	_microsoft_	_microsoft_										
15	21606	68	LEHMAN BRO	80599	LEH	_lehman_bro	_lehman_bro	_lehman_brothers_|_lehman_brothers_holdings_										
16	653	61	B H P BILLITO	75039	BHP	_bhp_	_bhp_	_bhp_|_bhp_billiton_										
17	43340	56	C I T GROUP I	89463	CIT	_cit_	_cit_	_cit_group_|_cit_										
18	20436	53	JPMORGAN C	47896	JPM	_jpmorgan_	_jpmorgan_	_jpmorgan_|_jpmorgan_chase_|_jpmorgan_securities_|_jpmorgan_chase_citigroup_|_jpmorgan_chase_bank_										
19	42125	53	CREDIT SUIS	89154	CS	_credit_suiss	_credit_suiss	_credit_|_credit_suisse_group_										
20	21645	51	AT & T INC	66093	T	_at&t_	_at&t_	_at&t_										
21	21633	47	SONY CORP	51131	SNE	_sony_	_sony_	_sony_|_sony_ericsson_|_sony_music_entertainment_|_sony_pictures_										

Month: 2009-08

Monthly networks going back to 1981



Accuracy of methodology

RIC mentions:	77,326
Unique firms associated with RICs:	1,861
Sentences with RICs:	60,615
RICs correctly matched by our algorithm (*true positives* or TP):	36,965
RICs incorrectly matched by our algorithm (*false positives* or FP):	1,347
RICs not matched by our algorithm (*false negatives* or FN):	38,571
Precision (TP / (TP + FP)):	96.48%
Recall (TP / (TP + FN)):	48.94%
F1 Score (2 × Precision × Recall / (Precision + Recall)):	64.94%

Total firm mentions identified by our algorithm:	514,015
Unique firms identified by our algorithm:	2,930
Sentences in which our algorithms identifies a firm:	410,120

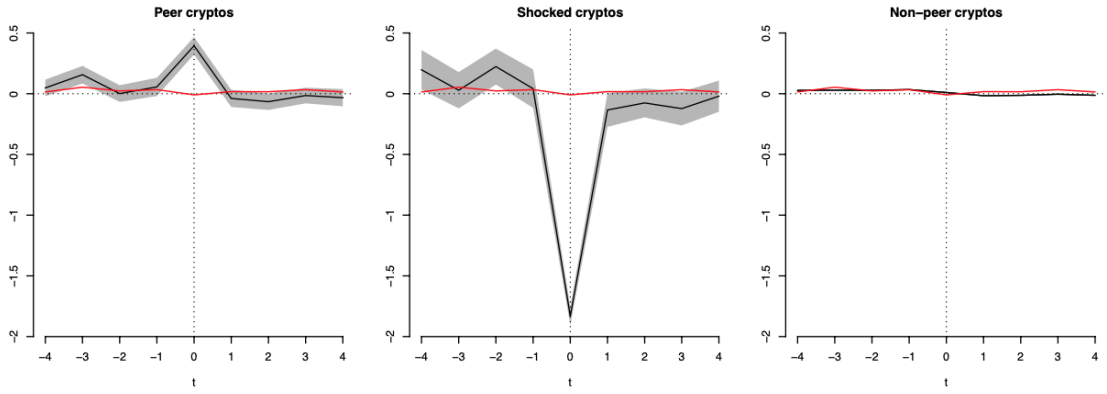
Application 1: predict aggregate outcomes

		Forecast horizon T											
		1	2	3	4	5	6	7	8	9	10	11	12
Log consumption†	Density	−0.061	◇ −0.144	−0.135	* −0.377	◇ −0.608	* −0.716	* −0.685	* −0.817	* −0.893	** −0.811	** −0.851	** −0.967
		(−1.510)	(−1.725)	(−0.963)	(−2.050)	(−1.879)	(−1.995)	(−2.265)	(−2.397)	(−2.449)	(−2.781)	(−2.777)	(−2.678)
	Centrality	0.042	0.084	−0.025	0.140	0.283	0.307	0.245	0.288	0.207	0.033	0.131	0.210
		(0.963)	(1.087)	(−0.376)	(0.830)	(0.923)	(0.894)	(0.745)	(0.762)	(0.592)	(0.144)	(0.503)	(0.678)
	Interconnectivity	−0.099	−0.203	−0.157	−0.357	−0.400	−0.277	−0.132	−0.076	−0.033	0.117	0.072	−0.021
		(−1.362)	(−1.401)	(−0.960)	(−1.174)	(−1.101)	(−1.010)	(−0.725)	(−0.413)	(−0.207)	(0.603)	(0.366)	(−0.106)
	Adj. R^2	0.009	0.065	0.091	0.122	0.158	0.153	0.117	0.133	0.158	0.133	0.122	0.159
Log defaults†	Density	−0.113	−0.119	−0.150	−0.140	−0.191	−0.146	−0.151	−0.145	−0.166	−0.110	−0.143	−0.017
		(−1.561)	(−0.867)	(−0.674)	(−0.542)	(−0.588)	(−0.425)	(−0.416)	(−0.343)	(−0.383)	(−0.249)	(−0.277)	(−0.036)
	Centrality	* 0.170	0.201	0.284	0.341	0.432	0.534	0.626	0.761	0.984	◇ 1.076	◇ 1.219	◇ 1.231
		(2.039)	(1.376)	(1.278)	(1.269)	(1.363)	(1.584)	(1.581)	(1.462)	(1.523)	(1.675)	(1.825)	(1.760)
	Interconnectivity	◇ −0.141	−0.148	−0.247	−0.341	−0.417	−0.521	−0.607	−0.728	−0.972	−1.096	◇ −1.257	◇ −1.306
		(−1.864)	(−1.035)	(−0.890)	(−1.043)	(−0.936)	(−1.104)	(−1.189)	(−1.223)	(−1.307)	(−1.530)	(−1.664)	(−1.909)
	Adj. R^2	0.583	0.660	0.700	0.726	0.736	0.747	0.751	0.756	0.761	0.760	0.763	0.769
Log ind. prod.†	Density	−0.026	−0.134	−0.164	−0.316	−0.559	−0.782	−0.789	◇ −0.898	◇ −1.006	◇ −1.014	* −1.085	* −1.209
		(−0.486)	(−1.366)	(−0.982)	(−1.349)	(−1.415)	(−1.521)	(−1.577)	(−1.701)	(−1.701)	(−1.862)	(−1.974)	(−2.031)
	Centrality	−0.003	0.072	0.013	0.001	0.136	0.182	0.052	0.004	−0.121	−0.387	−0.377	−0.254
		(−0.046)	(0.573)	(0.076)	(0.002)	(0.327)	(0.341)	(0.085)	(0.005)	(−0.152)	(−0.522)	(−0.471)	(−0.300)
	Interconnectivity	−0.005	−0.128	−0.108	−0.250	−0.363	−0.399	−0.243	−0.143	−0.062	0.097	0.101	0.026
		(−0.081)	(−0.767)	(−0.556)	(−0.803)	(−0.820)	(−0.952)	(−0.750)	(−0.484)	(−0.198)	(0.326)	(0.305)	(0.071)
	Adj. R^2	0.091	0.073	0.098	0.130	0.145	0.165	0.160	0.168	0.187	0.196	0.186	0.194

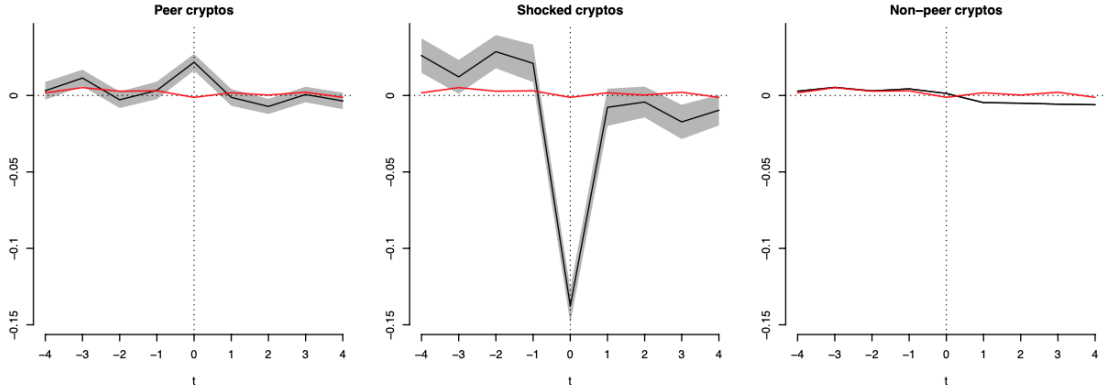
Table 11: *Predictive regression estimates (Part 1).* This table presents the estimates of the coefficient β_T of Regression (1) for the monthly density, centrality, and interconnectivity of the news-implied network extracted from the New York Times sample. The dependent variables are cumulative differences over T months of log consumption, log defaults, and log industrial production; see Section 5.1 for a description of these variables. We standardize all regressors and apply a Newey-West standard error adjustment in which the lag is equal to the sample size to the power one-over-four. The values in parentheses give t-statistics. ***, **, *, and ◇ denote significance on the 99.9%, 99%, 95%, and 90% confidence levels, respectively. † indicates that the coefficients have been scaled by a factor of 100.

Also other variables: credit spreads, S&P 500 returns, VIX

Application 2: crypto peer trading



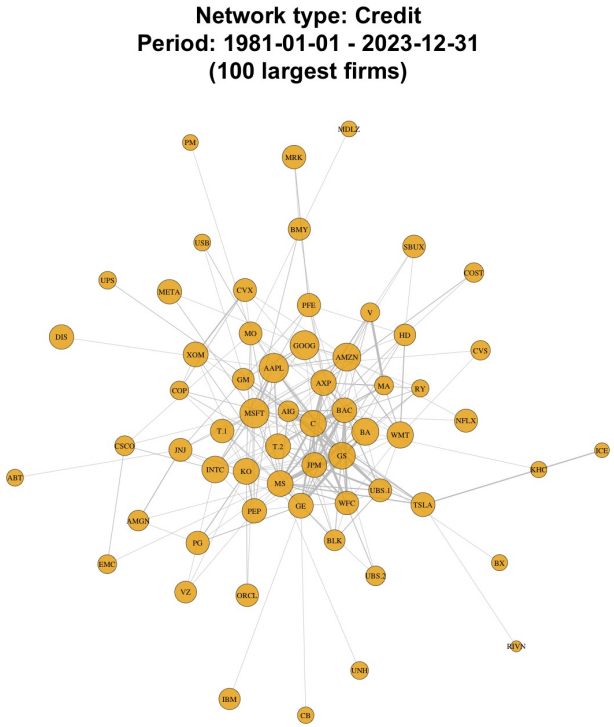
(a) Standardized abnormal return in week $e + t$, where e is the event week.

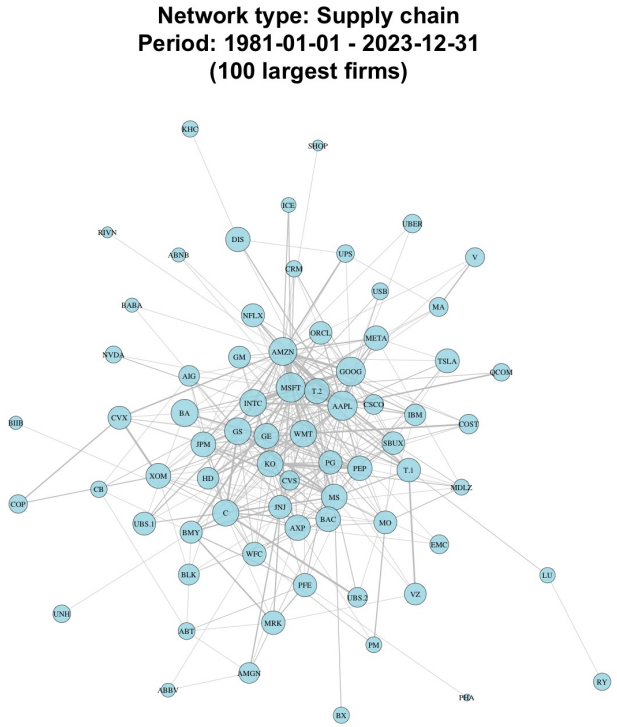


(b) Raw abnormal return in week $e + t$, where e is the event week.

6.5% annualized alpha, SR $= 1.4$

Application 3: network classification





Also: peer & equity links (other links also possible)

Conclusion

- We propose a methodology to extract firm links from text data
- We apply it to Reuters & NYT samples to identify 40,000 links among 13,500 firms
- Results enable several applications:
  - Predict aggregate outcomes
  - Predict returns among peers
  - Construct credit, supply chain, and other types of networks
- This is only the first step
  - Other asset classes, other text databases, ...
  - Data repository: `www.news-networks.net`

Thank you!

References

Acemoglu, Daron, Vasco M. Carvalho, Asuman Ozdaglar and Alireza Tahbaz-Salehi (2012), 'The network origins of aggregate fluctuations', *Econometrica* **80**(5), 1977–2016.

Herskovic, Bernard (2018), 'Networks in production: Asset pricing implications', *The Journal of Finance* **73**(4), 1785–1818.