Supervised Centrality Estimation and NeTwork Prediction (SuperCENT)

Equity Holding Network in China

Junhui Cai

Student Seminar
Sep 11, 2020

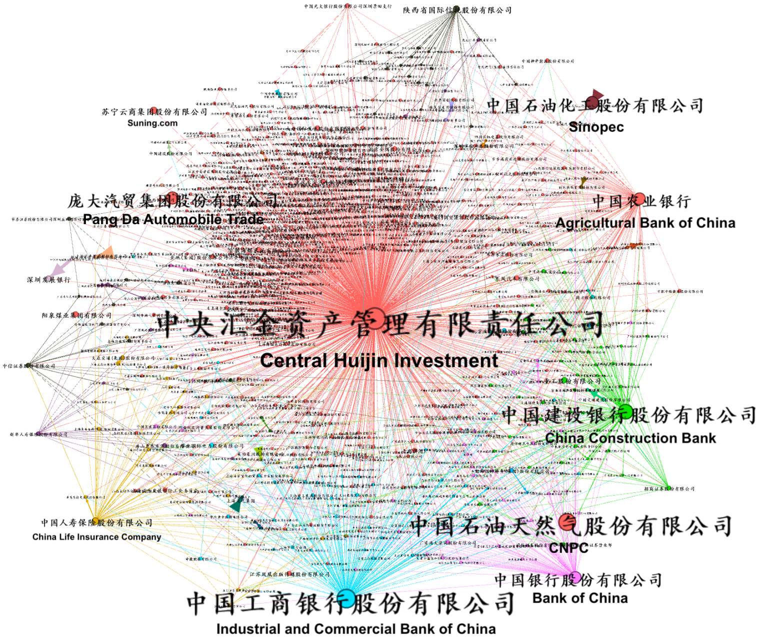
Joint work with Dan Yang, Linda Zhao and Wu Zhu

Outline of the talk

1. Introduction
   - Motivation: firm-to-firm equity holding network in China
   - Network centrality
   - Network centrality and predictive network models
2. Proposed methods
   - Two-stage: Prediction with naive centrality
   - SuperCENT: Simultaneous prediction and centrality estimation
   - Theoretical properties
   - Simulation
3. Summary and extension

Outline of the talk

1. Introduction
   - Motivation: firm-to-firm equity holding network in China
   - Network centrality
   - Network centrality and predictive network models
2. Proposed methods
   - Two-stage: Prediction with naive centrality
   - SuperCENT: Simultaneous prediction and centrality estimation
   - Theoretical properties
   - Simulation
3. Summary and extension


陕西省国际信托股份有限公司
苏宁云商集团股份有限公司
Suning.com
中国石油化工股份有限公司
Sinopec
庞大汽贸集团股份有限公司
Pang Da Automobile Trade
中国农业银行
Agricultural Bank of China
深圳发展银行
阳泉煤业集团有限公司
中央汇金资产管理有限责任公司
Central Huijin Investment
中国建设银行股份有限公司
China Construction Bank
中国人寿保险股份有限公司
China Life Insurance Company
中国石油天然气股份有限公司
CNPC
中国银行股份有限公司
Bank of China
中国工商银行股份有限公司
Industrial and Commercial Bank of China

Firm-to-firm equity holding network

As of 2019,

1. 50 million registered firms;
2. 108 million investments where 7 million are by firms;
3. 38 million changes in shareholders and/or investment amount(share) since 1949.

Our firm-to-firm equity holding network

1. consists of 7 million firms that invest to other firms and/or are invested by other firms;
2. is sparse with on average 1 investment per firm;
3. is directed and acyclic.

The data is huge (300G and keeps expanding) and noisy.

Firm-to-firm equity holding network

With this unprecedentedly comprehensive and unique data,
we are interested in

- structure of the network;
- formation of the network;
- impact of network position;
- ...

Questions of interest in this talk

1. A measure of importance of nodes: **centrality**
2. Estimate, infer, and predict centrality
3. Estimate and infer network effect and other parameters of interest

Firm-to-firm equity holding network

The firm-to-firm equity holding network $\mathcal{G} = (V, E)$ consists of

1. Nodes/vertices V: firms
2. Edges E: firm i invests to firm j ($i \to j$)

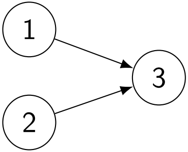
The network $\mathcal{G}$ can be represented as an adjacency matrix A where

$$A_{ij} = \begin{cases} s \geq 0 & \text{if firm } i \text{ invests } s\% \text{ of firm } j \\ 0 & \text{otherwise} \end{cases}$$

(s can also be the investment amount of firm i invests to firm j.)

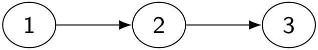
Centrality

Degree and eigenvector centrality



Citation network (vertices as papers and edges as citations)

- In-degree: number of citations
  - Problem: not all papers are equally influential
- Eigenvector centrality: more influential if cited by influential papers
  - $$\text{Iterate} \quad x_i = \sum_j A_{ji} x_j \quad \Longleftrightarrow \quad \mathbf{A}^\top \mathbf{x} = \lambda \mathbf{x}$$
  - Problems:
    - ⋆ directed acyclic graph has $\mathbf{x} = \mathbf{0}$



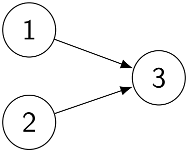
    - ⋆ only consider one direction

Centrality

Hub and authority scores (Kleinberg 99')

- Hub (u): important survey paper
- Authority (v): influential paper

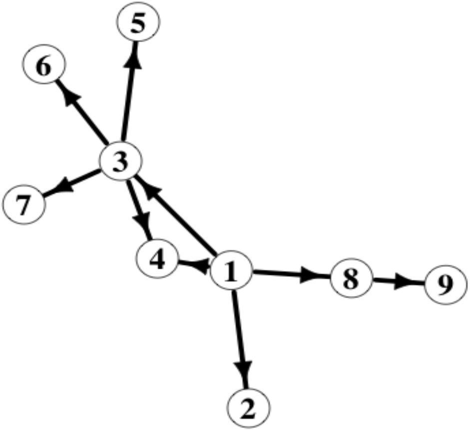
$$\text{Iterate} \begin{cases} \mathbf{A}^\top \boldsymbol{u} = \boldsymbol{v} \\ \mathbf{A}\boldsymbol{v} = \boldsymbol{u} \end{cases} \iff \begin{cases} \mathbf{A}\mathbf{A}^\top \boldsymbol{u} = \lambda' \boldsymbol{u} \\ \mathbf{A}^\top \mathbf{A}\boldsymbol{v} = \lambda' \boldsymbol{v} \end{cases}$$



$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \implies \boldsymbol{u} = \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix} \quad \boldsymbol{v} = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}$$

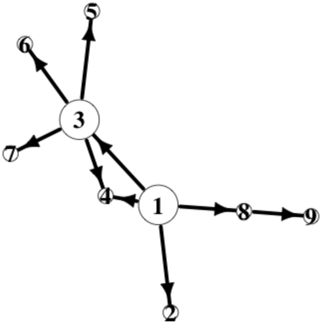
Centrality

Hub and authority scores (Kleinberg 99')

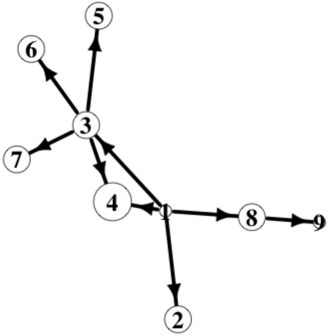


Centrality

Hub and authority scores (Kleinberg 99')



: Node size by hub score



: Node size by authority score

Centrality

Use cases

1. Information network
    - Internet: Google's PageRank
    - Citation network
2. Economic network
    - Industrial policy targets the most distorted industry in terms of resource misallocation in the input-output production network (Liu 19')
    - Venture capitals with better connection experience better performance (Hochberg, Ljungqvist, Lu 07')
    - Better firm position promotes firm growth (Allen, Cai, Gu, Qian, Zhu, Zhao 20')
    - ...

Models

We observe firm-to-firm holding network **A**, firm characteristics **x** and firm performance y.

- Network centrality model

$$\boldsymbol{A} = \boldsymbol{A}_0 + \boldsymbol{E} = d_0 \boldsymbol{u}_0 \boldsymbol{v}_0^\top + \boldsymbol{E} \tag{1}$$

- Predictive network model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}_x + \boldsymbol{u}_0 \beta_u + \boldsymbol{v}_0 \beta_v + \boldsymbol{\epsilon} \tag{2}$$

where $\mathbb{E}[E_{ij}] = 0$ and $\epsilon_i \sim N(0, \sigma_y^2)$.

Goal: Estimate and inference for

1. centrality: $\boldsymbol{u}_0$ and $\boldsymbol{v}_0$
2. network effect: β_u, β_v and $\boldsymbol{\beta}_x$

Outline of the talk

1. Introduction
   - Motivation: firm-to-firm equity holding network in China
   - Network centrality
   - Network centrality and predictive network models
2. Proposed methods
   - Two-stage: Prediction with naive centrality
   - SuperCENT: Simultaneous prediction and centrality estimation
   - Theoretical properties
   - Simulation
3. Summary and extension

Two-stage estimation and inference

Two-stage approach

1. Obtain $\hat{\boldsymbol{u}}^{svd}, \hat{\boldsymbol{v}}^{svd}$ from singular value decomposition of $\mathbf{A}$
2. Regress $\boldsymbol{y}$ on $\boldsymbol{X}$ and $\hat{\boldsymbol{u}}^{svd}, \hat{\boldsymbol{v}}^{svd}$ and obtain $\hat{\boldsymbol{\beta}}_x, \hat{\beta}_u, \hat{\beta}_v$

Table A.2 from *Ownership network and firm growth: what do five million companies tell about Chinese economy? (Allen, Cai, Gu, Qian, Zhu, Zhao 20')*

Dependent variable: firm growth

ROA	Leverage	Firm age	Firm size	SOE	In-net	hub	authority
0.347***	0.0189***	0.00746***	-0.472***	0.00239	0.0408***	0.0193***	0.00161***
(0.00362)	(0.00256)	(0.00108)	(0.00152)	(0.00420)	(0.00234)	(0.00161)	(0.000415)

Low rank-based objective L_{LR}

- Network centrality model

$$\boldsymbol{A} = \boldsymbol{A}_0 + \boldsymbol{E} = d_0 \boldsymbol{u}_0 \boldsymbol{v}_0^\top + \boldsymbol{E}$$

- Predictive network model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}_x + \boldsymbol{u}_0 \beta_u + \boldsymbol{v}_0 \beta_v + \boldsymbol{\epsilon}$$

where $\mathbb{E}[E_{ij}] = 0$ and $\epsilon_i \sim N(0, \sigma_y^2)$.

Joint objective function

$$\min_{\boldsymbol{\beta}_x, \beta_u, \beta_v, d, \|\boldsymbol{u}\|=\|\boldsymbol{v}\|=1} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_x - \boldsymbol{u}\beta_u - \boldsymbol{v}\beta_v\|_2^2 + \lambda \|\boldsymbol{A} - d\boldsymbol{u}\boldsymbol{v}'\|_F^2 \qquad (3)$$

Low rank-based objective L_{LR}

$$\min_{\boldsymbol{\beta},d,\|\boldsymbol{u}\|=\|\boldsymbol{v}\|=1} L(\mathbf{y},\mathbf{X};\boldsymbol{\beta},d,\boldsymbol{u},\boldsymbol{v}) = \|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_x-\boldsymbol{u}\beta_u-\boldsymbol{v}\beta_v\|_2^2 + \lambda\|\mathbf{A}-d\boldsymbol{u}\boldsymbol{v}^\top\|_F^2$$

Result: β, $\boldsymbol{u}$ and $\boldsymbol{v}$

Input: $\mathbf{A},\mathbf{X},\mathbf{y},\lambda$;

Initiate $(d^{(0)},\boldsymbol{u}^{(0)},\boldsymbol{v}^{(0)}) = \arg\min_{d,\boldsymbol{u},\boldsymbol{v}} \|A-d\boldsymbol{u}\boldsymbol{v}^\top\|_F^2$;

for $t = 1, \ldots$ **do**

1. Given $\boldsymbol{u}^{(t-1)}$ and $\boldsymbol{v}^{(t-1)}$, $\boldsymbol{\beta}^{(t)} = (\mathbf{W}^{(t-1)\prime}\mathbf{W}^{(t-1)})^{-1}\mathbf{W}^{(t-1)\prime}\mathbf{Y}$
2. Given $\boldsymbol{u}^{(t-1)}$ and $\boldsymbol{v}^{(t-1)}$, $d^{(t)} = \boldsymbol{u}^{(t-1)\top}\mathbf{A}\boldsymbol{v}^{(t-1)}$
3. $\boldsymbol{u}^{(t)} = \left((\beta_u^{(t)})^2+\lambda(d^{(t)})^2\right)^{-1}\left[\beta_u^{(t)}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_x^{(t)}-\boldsymbol{v}^{(t-1)}\beta_v^{(t)})+\lambda d^{(t)}\mathbf{A}\boldsymbol{v}^{(t-1)}\right]$
4. Normalize $\boldsymbol{u}^{(t)}$
5. $\boldsymbol{v}^{(t)} = \left((\beta_v^{(t)})^2+\lambda(d^{(t)})^2\right)^{-1}\left[\beta_v^{(t)}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_x^{(t)}-\boldsymbol{u}^{(t)}\beta_u^{(t)})+\lambda d^{(t)}\mathbf{A}^\top\boldsymbol{u}^{(t)}\right]$
6. Normalize $\boldsymbol{v}^{(t)}$

end

Theoretical properties

Theorem

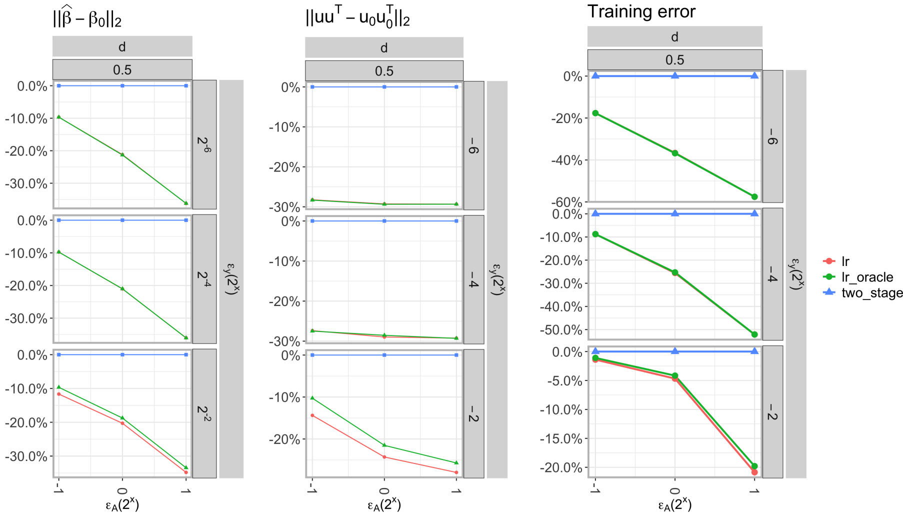
Under mild conditions, as $n \to \infty$,

$$n^{1/2} \begin{pmatrix} \hat{\boldsymbol{u}} - \boldsymbol{u}_0 \\ \hat{\boldsymbol{v}} - \boldsymbol{v}_0 \\ \hat{\beta}_u - \beta_u \\ \hat{\beta}_v - \beta_v \\ \hat{\boldsymbol{\beta}}_x - \boldsymbol{\beta}_x \end{pmatrix} \xrightarrow{\mathcal{D}} N\left(\boldsymbol{0}_{(2n+2+p)\times 1}, \boldsymbol{C} \begin{pmatrix} \sigma_y^2 \boldsymbol{I}_n & \boldsymbol{0}_{n \times n^2} \\ \boldsymbol{0}_{n^2 \times n} & \sigma_a^2 \boldsymbol{I}_{n^2} \end{pmatrix} \boldsymbol{C}^\top \right)$$

where $\boldsymbol{C} = \boldsymbol{C}(\lambda, d, \boldsymbol{u}, \boldsymbol{v}, \beta_u, \beta_v, \boldsymbol{\beta}_x)$

Simulation

- $\boldsymbol{u}, \boldsymbol{v} \sim N(0,1), d = .5$ and $\mathbf{X} \sim N(0, \mathbf{I}_3)$
- $\boldsymbol{\beta} = (5,3,1)$, $\beta_u = 5, \beta_v = 5$.



Outline of the talk

Summary and extensions

1. Firm-to-firm equity holding network
2. SuperCENT: simultaneous prediction and centrality estimation
3. Inference for both centrality and effect
4. Extension to longitudinal network

Related ongoing research

Led by: Linda Zhao

Franklin Allen	Xian Gu	Qian Jun	Dan Yang	Wu Zhu
Imperial College	Durham U	Fudan U	HKU	Penn Econ PhD

1. Ownership network and firm growth: what do five million companies tell about Chinese economy? Available on SSRN.
2. Ownership network in China: history, present and future.
3. Revisiting state-owned enterprises in China. Available on SSRN soon.
4. Government ownership in China.
5. The Covid-19 Economy in China.

Supported by: NSF, CREF, Wharton Global Initiatives, Mack Institute Fund

Thank You!