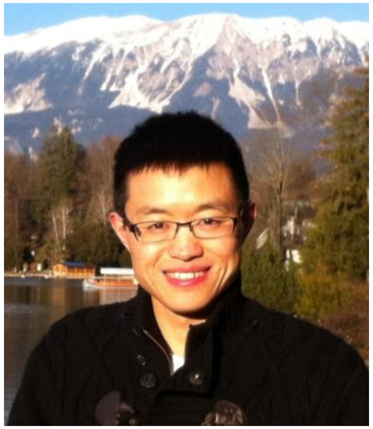
iPath: Forecasting the Pathway to Impact

Presenter: **Liangyue Li**

Joint work with

Hanghang Tong (ASU), **Jie Tang** (Tsinghua), **Wei Fan** (Baidu)



Foresee the Pathway to Impact

Geoffrey Hinton

Learning representations by back-propagating errors

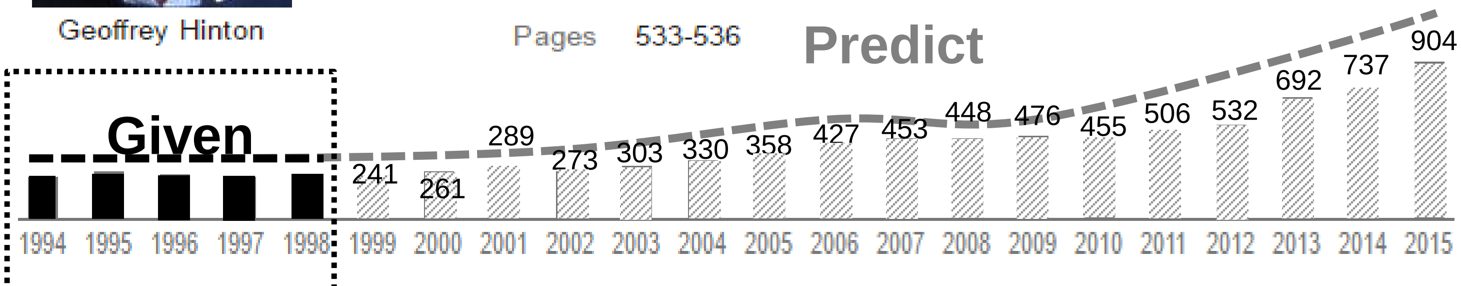
Authors David E Rumelhart, Geoffrey E Hinton, Ronald J Williams

Publication date 1986

Journal Nature

Volume 323

Pages 533-536



Implications of forecasting the pathway to impact

- Tracking research frontier
- Invoking early intervention

Question: how to foresee the impact pathway at the early stage?

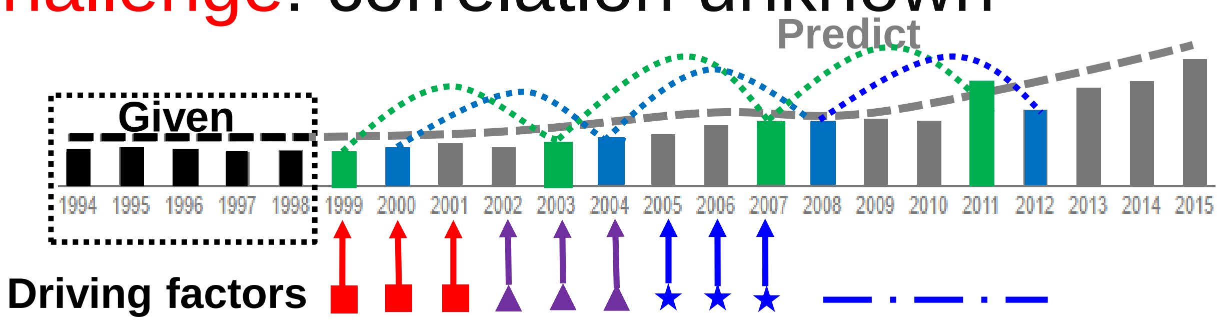
Modeling Scientific Impact

- **Effective scholarly feature design** [Yan+CIKM11]
- **Mechanistic model for the citation dynamics of individual papers** [Wang+Science13]
- ***iBall*- Joint Predictive Model for long-term impact prediction** [Li+KDD15]

All for Point Prediction

Challenges

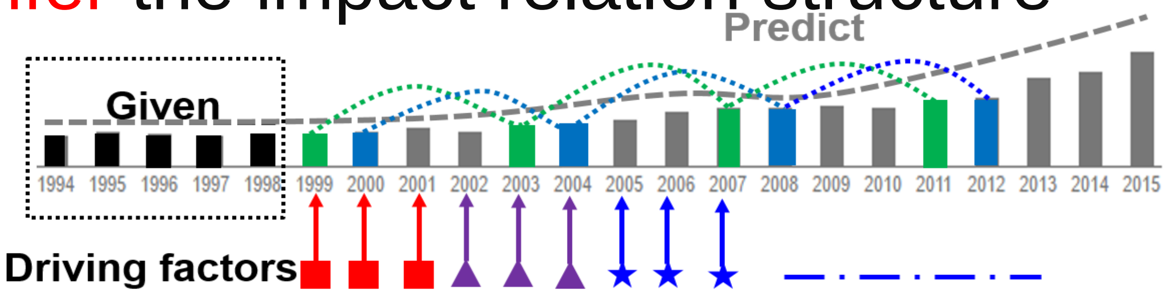
- C1: Output Space -- Correlation
  - Possible solution: multi-label/task learning
  - Challenge: correlation unknown



- C2: Parameter Space -- Smoothness
  - Possible solution: linear dynamic system
  - Challenge: transition process unknown

Design Objectives

- D1: Prediction Consistency (for C1)
  - Exploit the correlation in output space
  - Infer the impact relation structure

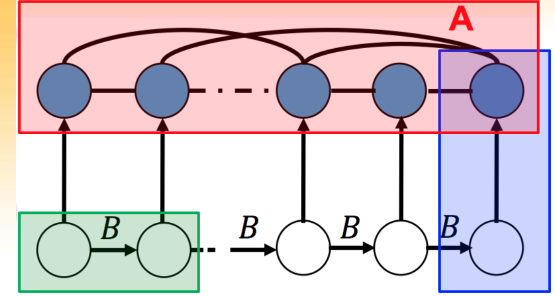


- D2: Parameter Smoothness (for C2)
  - Apply linear transition to adjacent parameters
  - Learn the linear transition process

Roadmap

- Motivations
- **Proposed Solutions: iPath**
- Experimental Results
- Conclusions

iPath -- Formulations



- Optimization Formulations

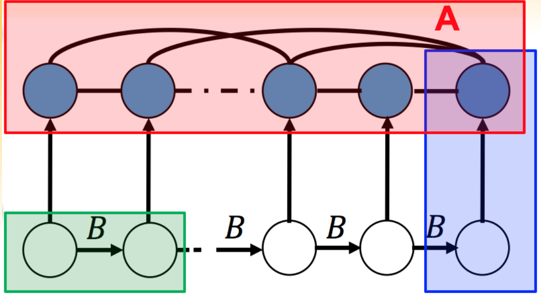
$$\min_{\mathbf{W},\mathbf{B},\mathbf{A}} \overset{\text{Empirical loss}}{\left[\mathcal{L}[f(\mathbf{X},\mathbf{W}),\mathbf{Y}]\right]} + \overset{\text{Prediction Consistency}}{\left(\alpha \sum_{i=1}^{l}\sum_{j=1}^{l} \mathbf{A}_{ij}\, g(\mathbf{w}_i,\mathbf{w}_j)\right)}$$

$$+ \left(\beta \sum_{t=2}^{l} \|\mathbf{w}_t - \mathbf{B}\mathbf{w}_{t-1}\|_2^2\right) \text{Parameter Smoothness}$$

$$+ \underbrace{\gamma\|\mathbf{B}-\mathbf{I}\|_F^2 + \delta\sum_{i=1}^{l}\Omega(\mathbf{w}_i) + \epsilon\|\mathbf{A}-\mathbf{A}_0\|_F^2}_{\text{Regularizations}}$$

- Remarks
  - **Prediction Consistency**: similar impacts have similar models
  - **Parameter Smoothness**: model parameters at adjacent time steps have linear transformation

iPath – linear formulation



- **Details**:

$$\min_{\mathbf{W},\mathbf{B},\mathbf{A}} \quad \|\mathbf{X}\mathbf{W}-\mathbf{Y}\|_F^2 + \alpha \sum_{i=1}^{l}\sum_{j=1}^{l} \boxed{\mathbf{A}_{ij}\|\mathbf{X}\mathbf{w}_i - \mathbf{X}\mathbf{w}_j\|_2^2}$$

$$+\beta \sum_{t=2}^{l} \|\mathbf{w}_t - \mathbf{B}\mathbf{w}_{t-1}\|_2^2 + \gamma\|\mathbf{B}-\mathbf{I}\|_F^2$$

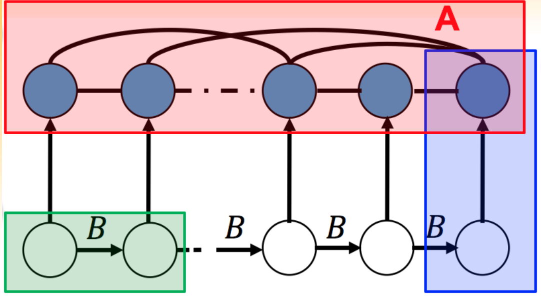
$$+\delta \sum_{i=1}^{l} \|\mathbf{w}_i\|_2^2 + \epsilon\|\mathbf{A}-\mathbf{A}_0\|_F^2$$

- **Intuition**:

Similar impacts (large $\mathbf{A}_{ij}$)

➡ Similar Predictions (small $\|\mathbf{X}\mathbf{w}_i - \mathbf{X}\mathbf{w}_j\|_2^2$)

iPath – non-linear formulation



- **Details**:

$$\min_{\mathbf{W},\mathbf{B},\mathbf{A}} \quad \|\mathbf{KW}-\mathbf{Y}\|_F^2 + \alpha \sum_{i=1}^{l}\sum_{j=1}^{l} \boxed{\mathbf{A}_{ij}\|\mathbf{Kw}_i - \mathbf{Kw}_j\|_2^2}$$

$$+\beta \sum_{t=2}^{l} \|\mathbf{w}_t - \mathbf{Bw}_{t-1}\|_2^2 + \gamma\|\mathbf{B}-\mathbf{I}\|_F^2$$

$$+\delta \sum_{i=1}^{l} \mathbf{w}_i'\mathbf{Kw}_i + \epsilon\|\mathbf{A}-\mathbf{A}_0\|_F^2$$

- **Intuition**:

Similar Impacts (large $\mathbf{A}_{ij}$)

Similar Predictions (small $\|\mathbf{Kw}_i - \mathbf{Kw}_j\|_2^2$)

iPath – Optimization Solutions

- Alternating Optimization Strategy

Optimize for W

$$\min_{\mathbf{W}} \ \|\mathbf{X}\mathbf{W}-\mathbf{Y}\|_F^2+\alpha\sum_{i=1}^{l}\sum_{j=1}^{l}\mathbf{A}_{ij}\|\mathbf{X}\mathbf{w}_i-\mathbf{X}\mathbf{w}_j\|_2^2$$
$$+\beta\sum_{t=2}^{l}\|\mathbf{w}_t-\mathbf{B}\mathbf{w}_{t-1}\|_2^2+\delta\sum_{i=1}^{l}\|\mathbf{w}_i\|_2^2$$

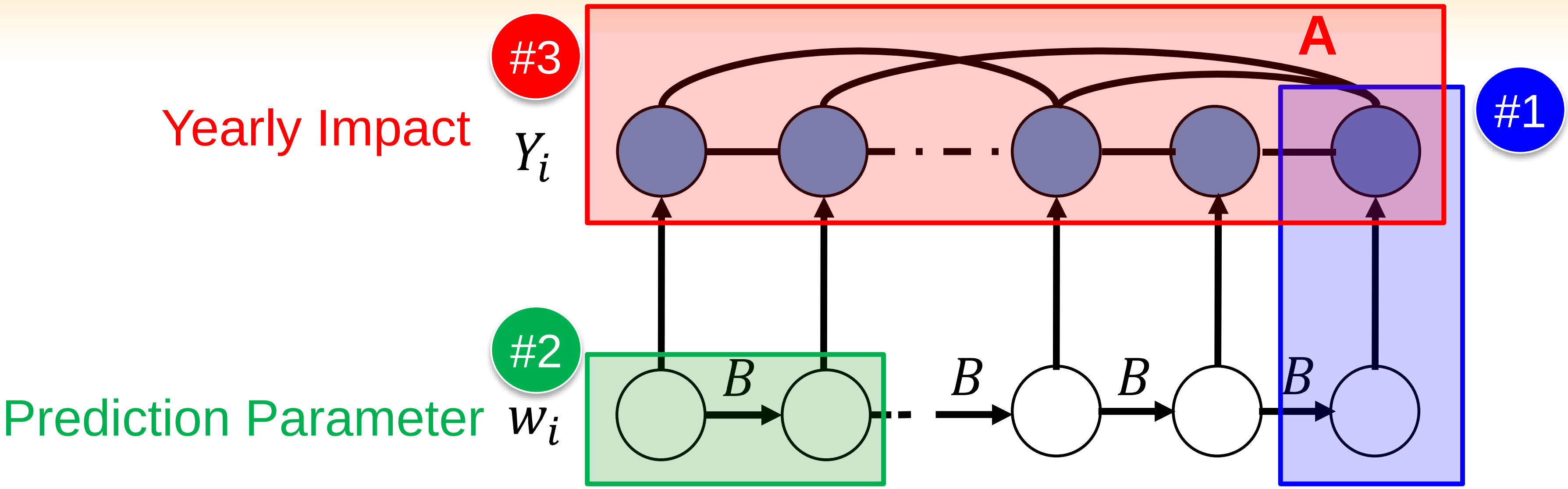
Optimize for B

$$\min_{\mathbf{B}} \ \beta\sum_{t=2}^{l}\|\mathbf{w}_t-\mathbf{B}\mathbf{w}_{t-1}\|_2^2+\gamma\|\mathbf{B}-\mathbf{I}\|_F^2$$

Optimize for A

$$\min_{\mathbf{A}} \ \alpha\sum_{i=1}^{l}\sum_{j=1}^{l}\mathbf{A}_{ij}\|\mathbf{X}\mathbf{w}_i-\mathbf{X}\mathbf{w}_j\|_2^2+\epsilon\|\mathbf{A}-\mathbf{A}_0\|_F^2$$

Probabilistic Interpretation



Impact prob cond. on parameter	#1 $\mathbf{Y}_i \mid \mathbf{w}_i \sim \mathcal{N}(\mathbf{X}\mathbf{w}_i, \sigma_y^2 \mathbf{I})$
Transition prob	#2 $\mathbf{w}_t \mid \mathbf{w}_{t-1} \sim \mathcal{N}(\mathbf{B}\mathbf{w}_{t-1}, \sigma_w^2 \mathbf{I})$
MRF of impacts	$p(\mathbf{Y}) = \frac{1}{Z}\exp(-E(\mathbf{Y}))$, where $E(\mathbf{Y}) = \sum_{c \in \mathcal{C}} \Phi_c(\mathbf{Y}_c)$ #3 $\Phi_{e=(\mathbf{Y}_i, \mathbf{Y}_j)} = \mathbf{A}_{ij}\lVert\mathbf{Y}_i - \mathbf{Y}_j\rVert_2^2$ $= \mathbf{A}_{ij}\lVert\mathbf{X}\mathbf{w}_i - \mathbf{X}\mathbf{w}_j\rVert_2^2$
Joint dist.	$\arg\max_{\mathbf{Y},\mathbf{X},\mathbf{W}} = p(\mathbf{w}_1)\prod_{t=2}^{l} p(\mathbf{w}_t \mid \mathbf{w}_{t-1}) \prod_{i=1}^{l} p(\mathbf{Y}_i \mid \mathbf{w}_i) p(\mathbf{Y})$

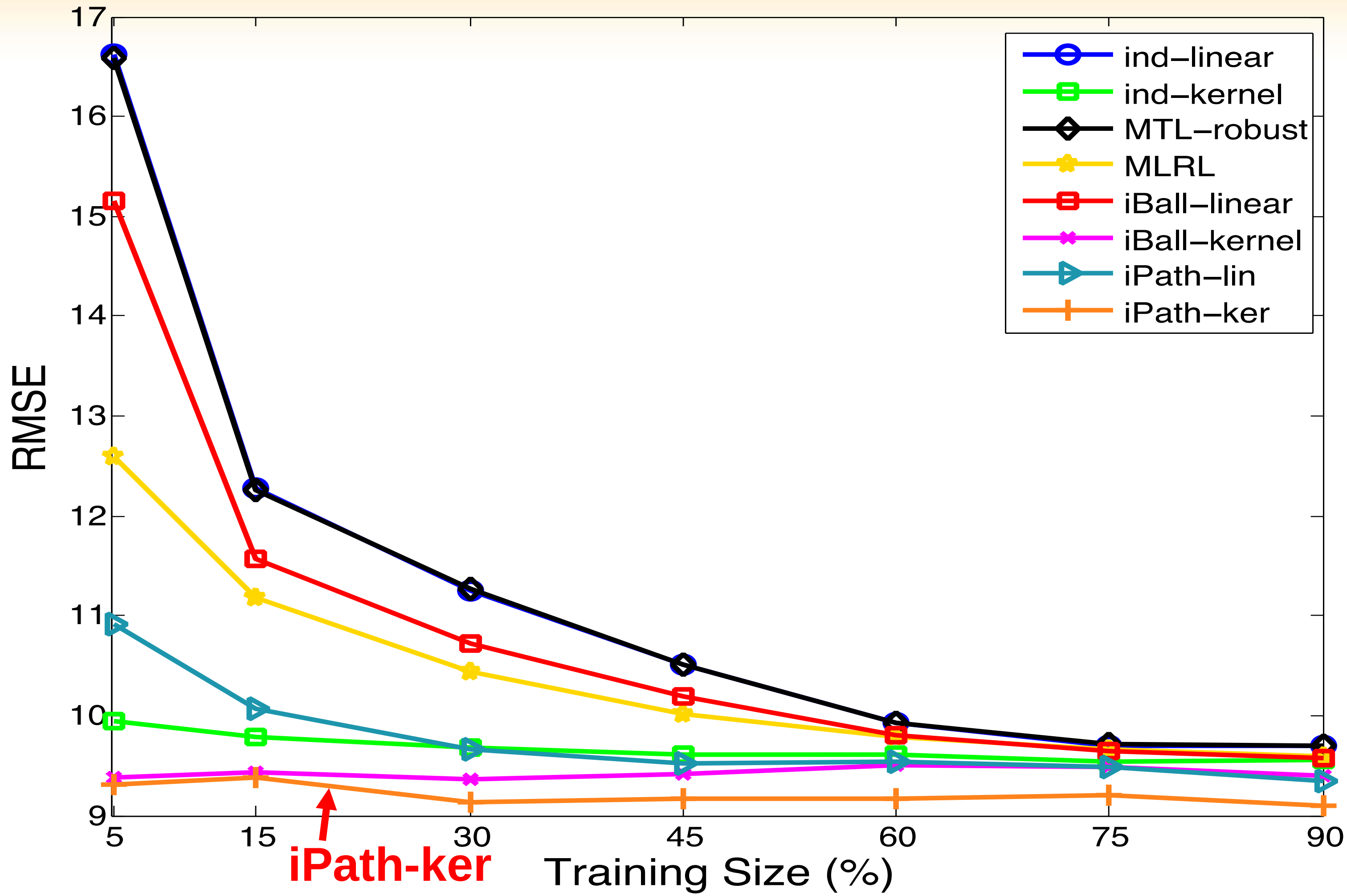
Roadmap

- Motivations
- Proposed Solutions: iPath
- **Experimental Results**
- Conclusions

Experiment Setup

- **Datasets**: *AMiner* (2,243,976 papers, 1,274,360 authors, 8,882 venues)
- **Task**: Observing the first 5 years' citations, predict yearly citations from year 6 – 15
- **Evaluation Metric**: Root Mean Squared Error (RMSE)

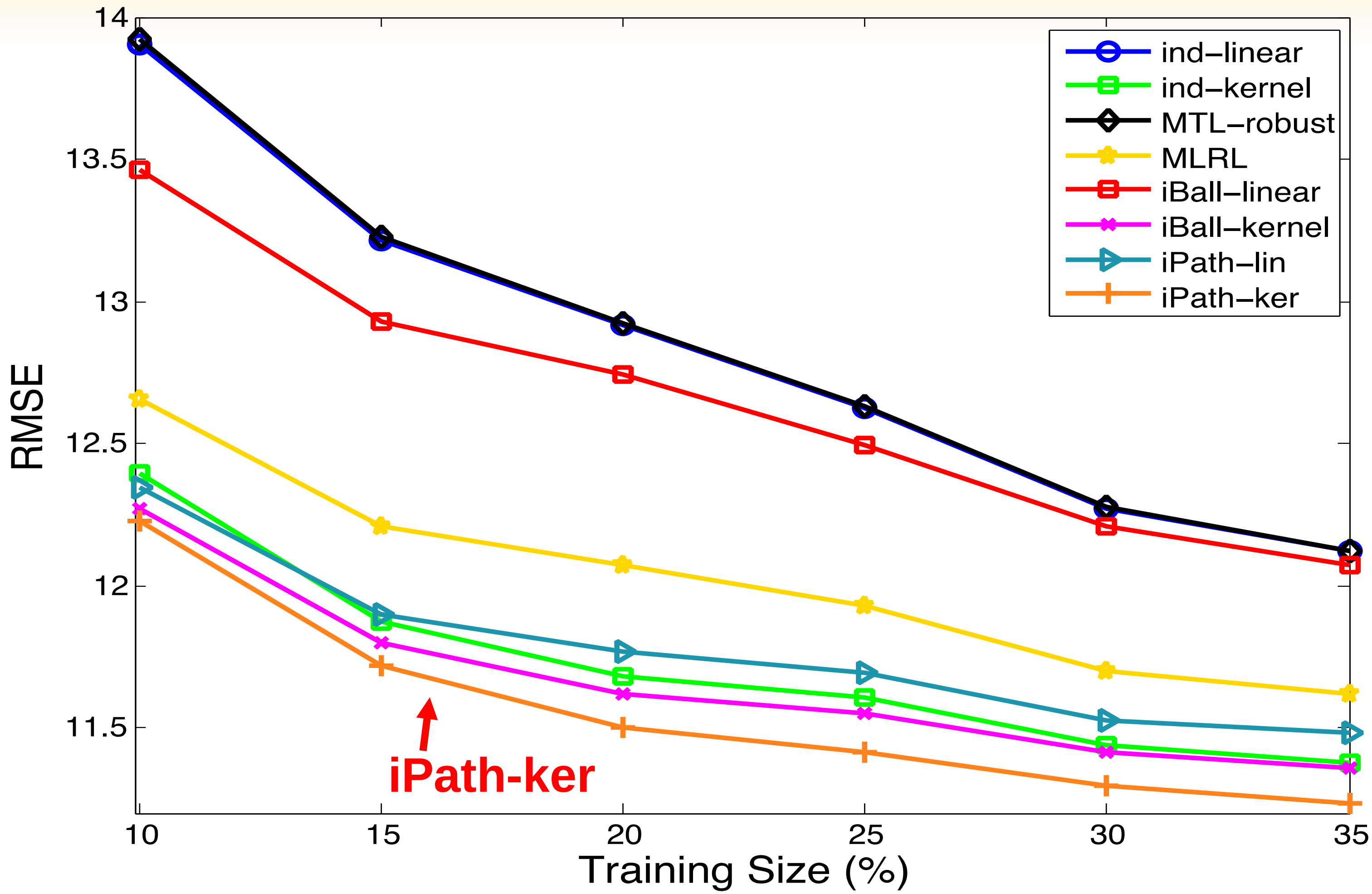
Paper Impact Pathway Prediction



Obs: iPath-ker performs the best among all the competitors

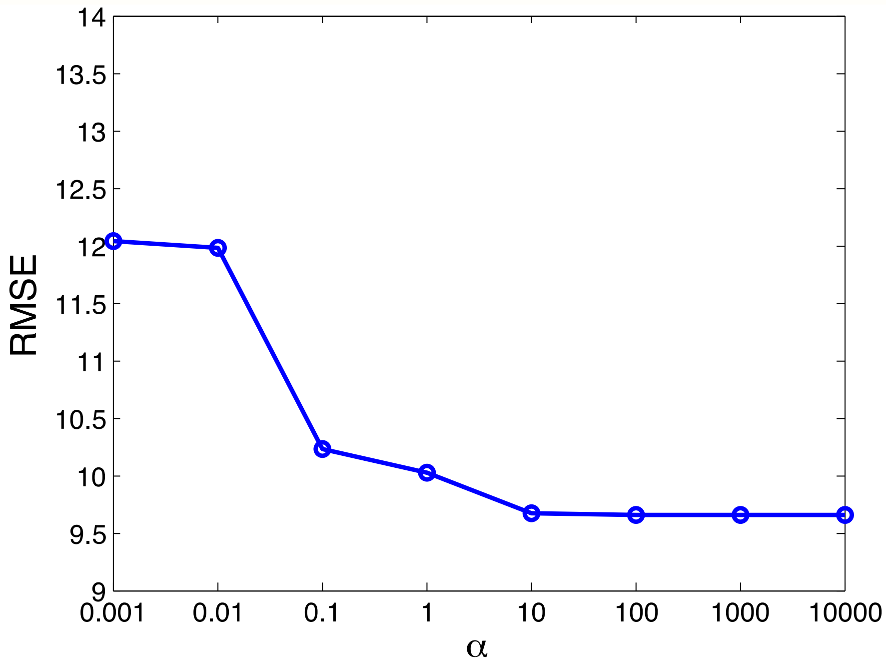
DATA Lab

Author Impact Pathway Prediction

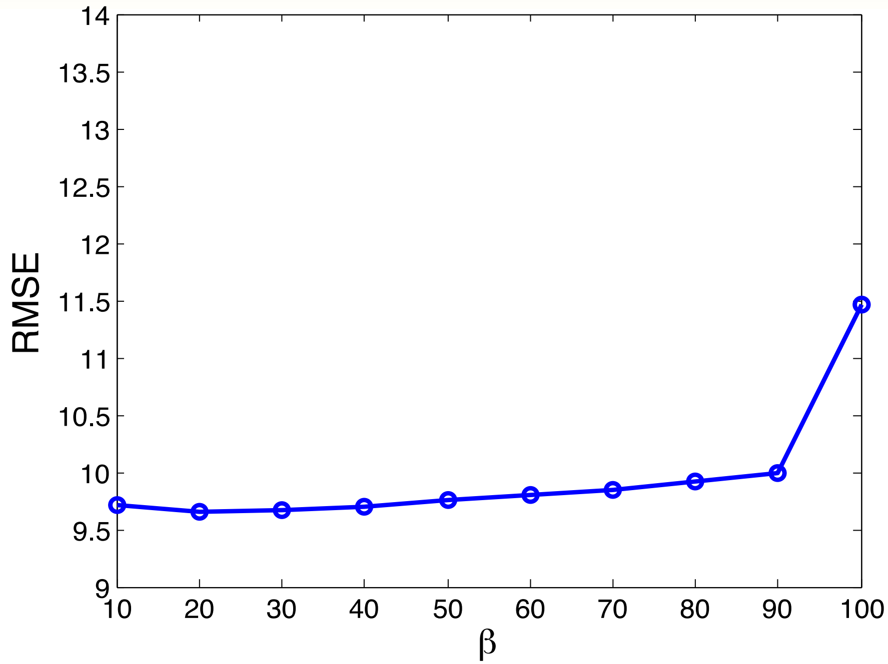


Obs: iPath-ker performs the best among all the competitors

Sensitivity Analysis



RMSE vs. α



RMSE vs. β

Obs: iPath is stable in a large range of parameter spaces

Performance Gain Analysis

RMSE	Paper Impact	Author Impact
1	9.602	11.608
1 + 2	9.507	11.548
1 + 2 + 3	9.335	11.489
1 + 2 + 3 + 4	9.171	11.391

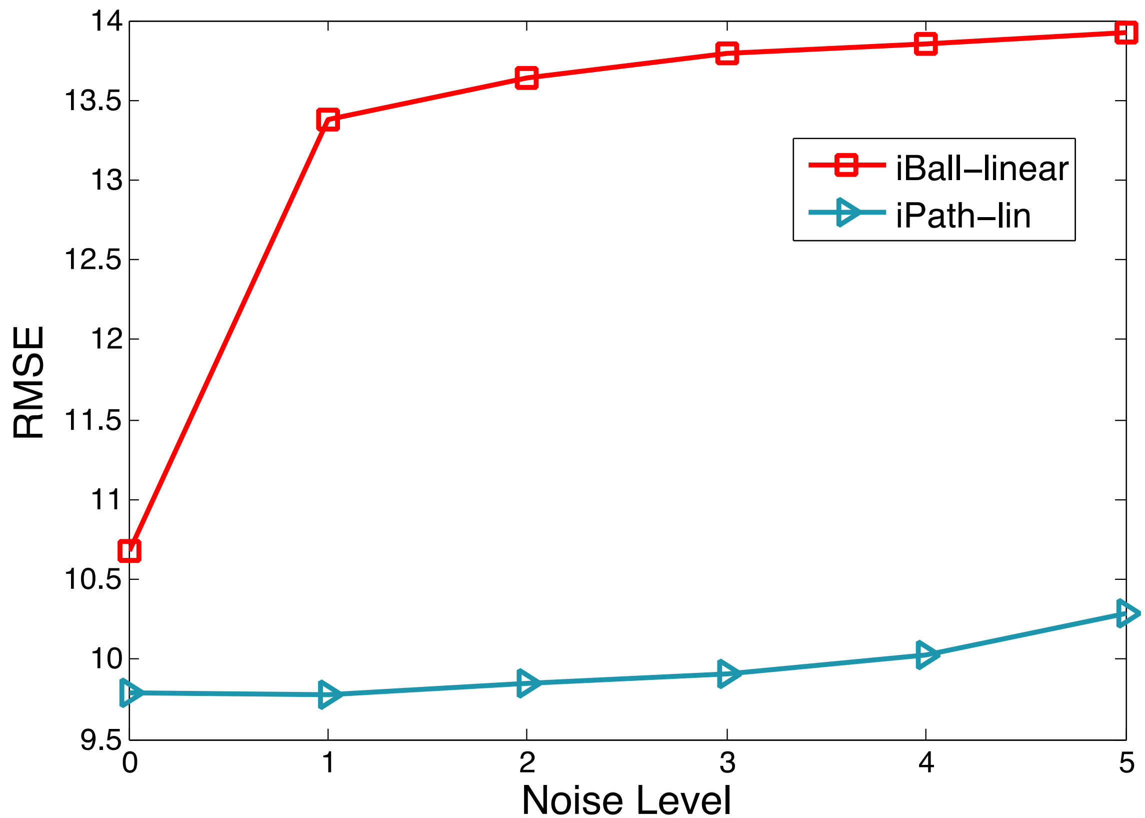
1 Basic form; 2 relation; 3 transition; 4 inferring

$$\min_{\mathbf{W},\mathbf{B},\mathbf{A}} \left(\|\mathbf{KW}-\mathbf{Y}\|_F^2 \right) + \left(\alpha \sum_{i=1}^{l}\sum_{j=1}^{l} \mathbf{A}_{ij}\|\mathbf{Kw}_i-\mathbf{Kw}_j\|_2^2 \right) + \left(\beta \sum_{t=2}^{l} \|\mathbf{w}_t-\mathbf{Bw}_{t-1}\|_2^2 \right) + \gamma\|\mathbf{B}-\mathbf{I}\|_F^2 + \delta\sum_{i=1}^{l}\mathbf{w}_i'\mathbf{Kw}_i + \left(\epsilon\|\mathbf{A}-\mathbf{A}_0\|_F^2 \right)$$

Obs: relation, transition and inferring are all beneficial in improving the prediction

Robustness to Noise in A_0

$$\min_{\mathbf{W},\mathbf{B},\mathbf{A}} \|\mathbf{XW}-\mathbf{Y}\|_F^2 + \alpha \sum_{i=1}^{l}\sum_{j=1}^{l} \mathbf{A}_{ij}\|\mathbf{Xw}_i - \mathbf{Xw}_j\|_2^2$$
$$+\beta\sum_{t=2}^{l}\|\mathbf{w}_t-\mathbf{Bw}_{t-1}\|_2^2 + \gamma\|\mathbf{B}-\mathbf{I}\|_F^2$$
$$+\delta\sum_{i=1}^{l}\|\mathbf{w}_i\|_2^2 + \epsilon\|\mathbf{A}-\mathbf{A}_0\|_F^2$$

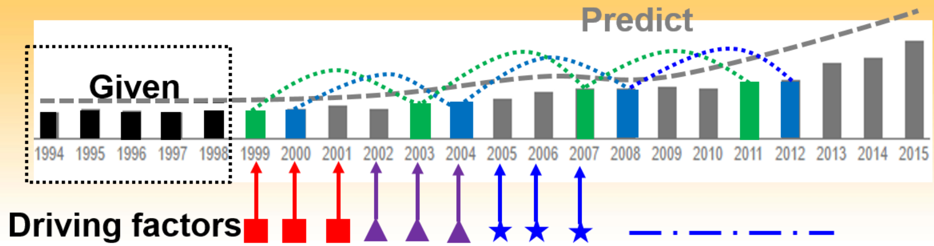


Obs: iPath degenerates gradually with the noise level

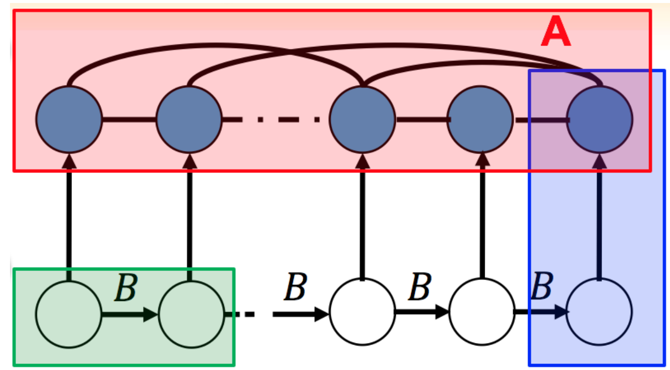
Roadmap

- Motivations
- Proposed Solutions: iPath
- Experimental Results
- **Conclusions**

Conclusions



- **Goals**: predict the pathway to impact
- **Solutions**: *i*Path prediction model
  - Design objectives:
    - Prediction Consistency
    - Parameter Smoothness



  - Results:
    - Lower error than competitors
    - Robust to noise in impact relations