Identification and Approximation of the Structure of Networks of Stochastic Processes

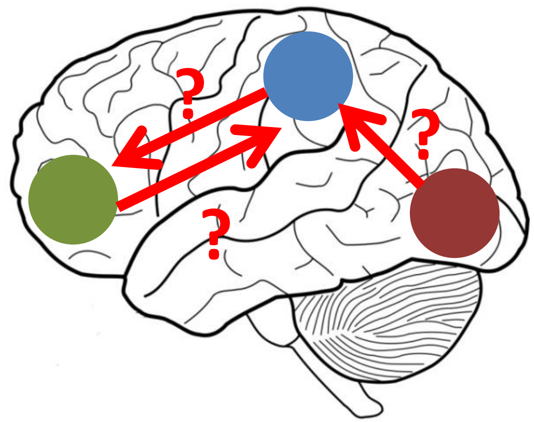
Chris Quinn

Challenge in Neuroscience

Which parts talk to which parts?

-- between brain regions

-- between cells

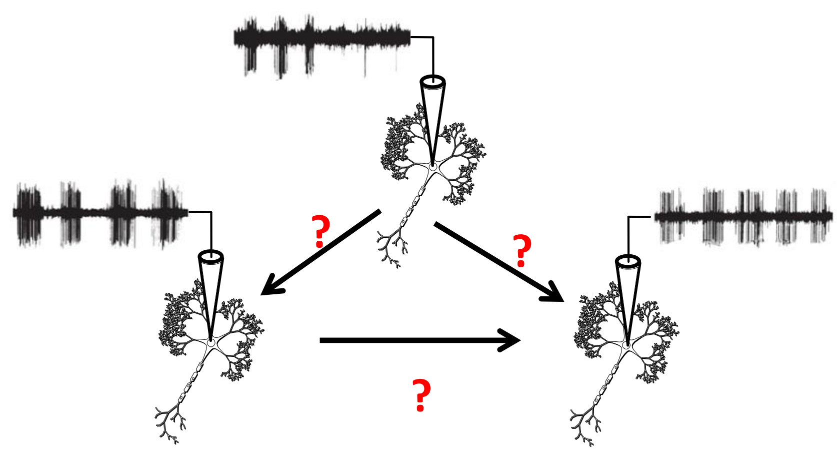


Motivation:

-- Understand cognition

-- Treat neurological diseases (e.g. epilepsy)

-- Enhance neuroprosthetic devices (e.g. arms; vision)



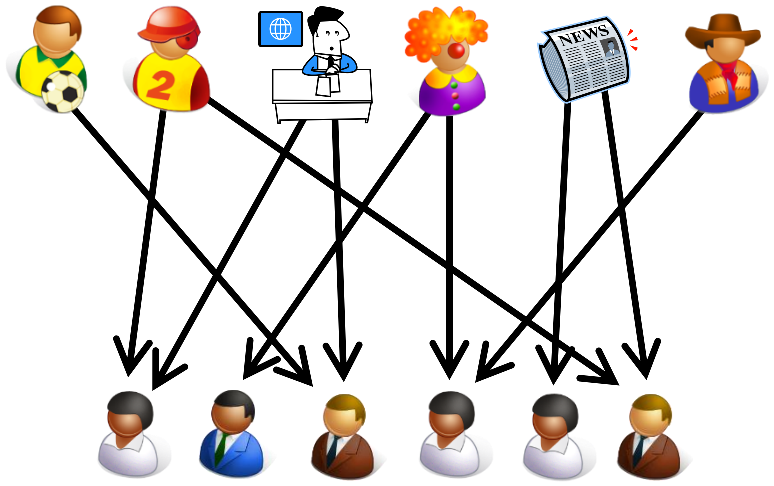
Challenge in Social Science

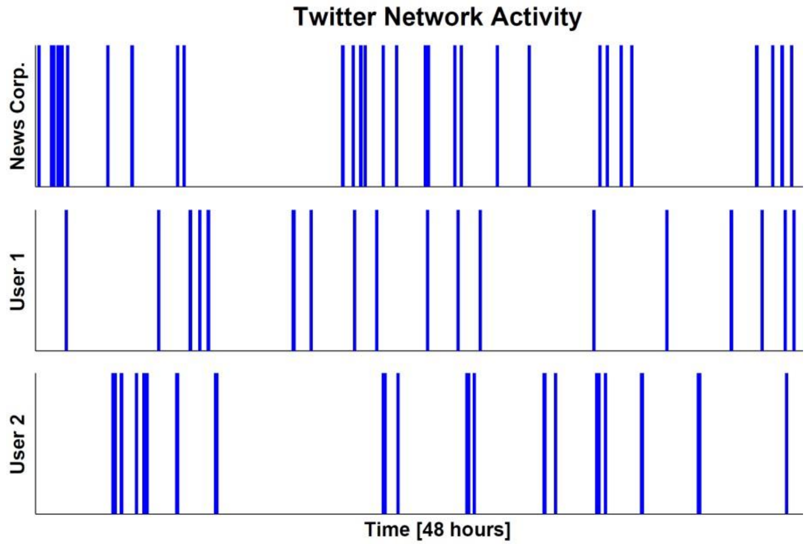
Online Social Networks

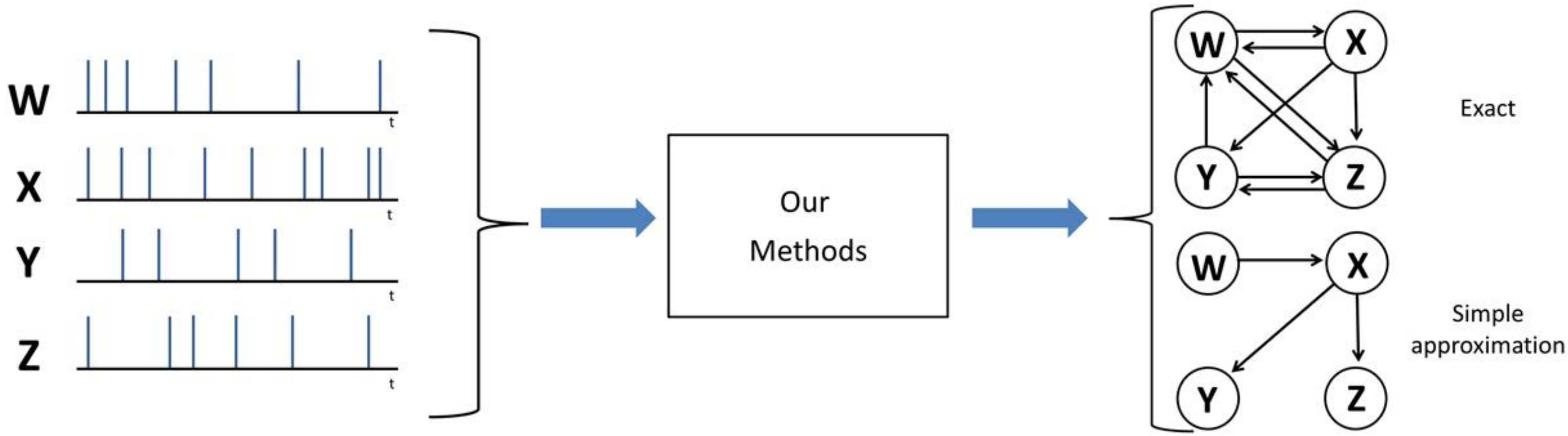
-- Learn group dynamics: behavior and communication

-- Advertisers care if user might influence friends (“like” this product)

-- Study information spreading dynamics (how and how fast)







Project:

-- Analyze causal influences between time series

-- Graphical models to succinctly represent influence structure

-- Algorithms to identify true topology and approximations

-- Estimation

Granger Causality

$$\mathbf{X} \xrightarrow{?} \mathbf{Y}$$

"We say that **X** is causing **Y** if we are better able to predict the future of **Y** using the all past knowledge than without the past of **X**."

Autoregressive models

$$Y_t = \sum_{\tau > 0} a_\tau Y_{t-\tau} + b_\tau X_{t-\tau} + c_\tau Z_{t-\tau} + E_t$$
$$Y_t = \sum_{\tau > 0} \tilde{a}_\tau Y_{t-\tau} + \tilde{c}_\tau Z_{t-\tau} + \tilde{E}_t.$$

Error terms (Gaussian): $E_t, \tilde{E}_t.$

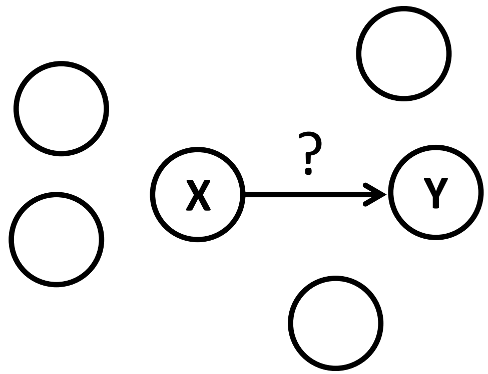
If $\mathrm{var}(E_t) = \mathrm{var}(\tilde{E}_t)$, $\mathbf{X} \not\to \mathbf{Y}$.

(Past of **X** does not help prediction.)

[Granger 1964, 1969]

Directed information graphs

"We say that **X** is causing **Y** if we are better able to predict the future of **Y** using the all past knowledge than without the past of **X**."



Draw an edge from **X** to **Y** if:

$$\mathrm{I}(\mathbf{X} \to \mathbf{Y} \| \underline{\mathbf{X}} \backslash \{\mathbf{X}, \mathbf{Y}\}) > 0$$

Theorem: Equivalent to minimum generative model graphs.

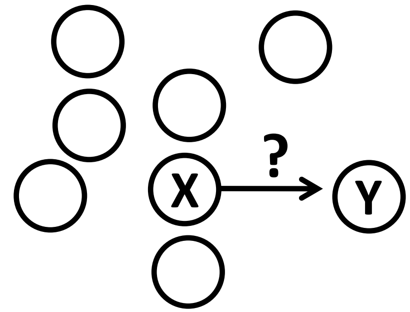
[CJQ, NK, TPC. ISIT 2011]

Exact Algorithms

- DI graph definition

  Draw an edge from **X** to **Y** if:

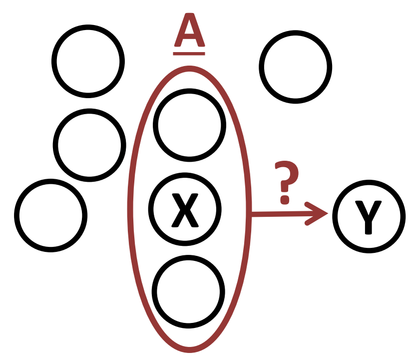
  $$\mathrm{I}(\mathbf{X} \to \mathbf{Y} \| \underline{\mathbf{X}} \backslash \{\mathbf{X}, \mathbf{Y}\}) > 0$$



- "causal Markov blanket"

  For each **Y**, can find parent set A by:

  $$\begin{aligned}&\text{Set } \underline{\mathrm{A}} \leftarrow \{1, \ldots, m\} \backslash \{i\}\\ &\text{For } j \in \underline{\mathrm{A}}\\ &\quad \text{If } \mathrm{I}(\mathbf{X}_j \to \mathbf{Y} \| \underline{\mathbf{X}}_{\underline{\mathrm{A}} \backslash \{j\}}) = 0\\ &\qquad \underline{\mathrm{A}} \leftarrow \underline{\mathrm{A}} \backslash \{j\}\end{aligned}$$



- Adaptive

[CJQ, NK, TPC. TIT submitted]

Exact Algorithms

Bounded in-degree

Case: know upper bound K on in-degree

Let A index true parents of Y

Let $\underline{\mathrm{A}} \subset \{1,2,3\},\ |\underline{\mathrm{A}}| \leq K = 2$

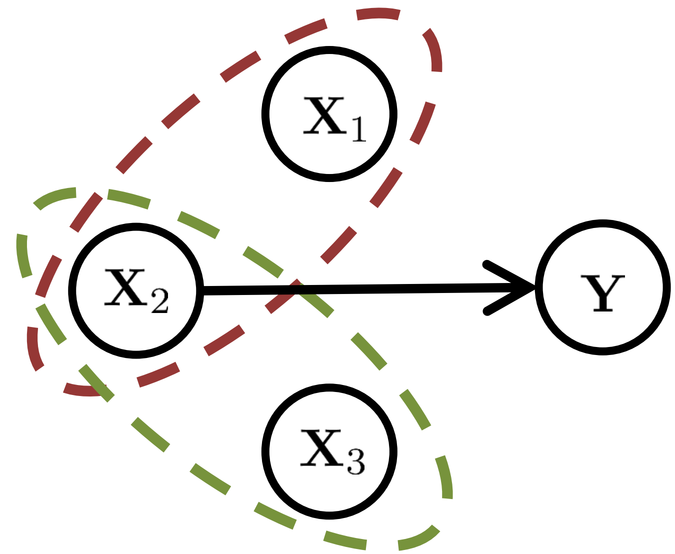
$$\mathrm{I}(\mathbf{X}_1, \mathbf{X}_2 \to \mathbf{Y}) = 3$$

$$\mathrm{I}(\mathbf{X}_2, \mathbf{X}_3 \to \mathbf{Y}) = 3$$

$$\mathrm{I}(\mathbf{X}_1, \mathbf{X}_3 \to \mathbf{Y}) = 2$$

$$\underline{\mathrm{A}} = \{1,2\} \bigcap \{2,3\} = 2$$

$$\underline{\mathrm{A}} = \bigcap \arg\max_{\underline{\mathrm{A}}'}\ \mathrm{I}(\mathbf{X}_{\underline{\mathrm{A}}'} \to \mathbf{Y})$$



Repeat for each node Y

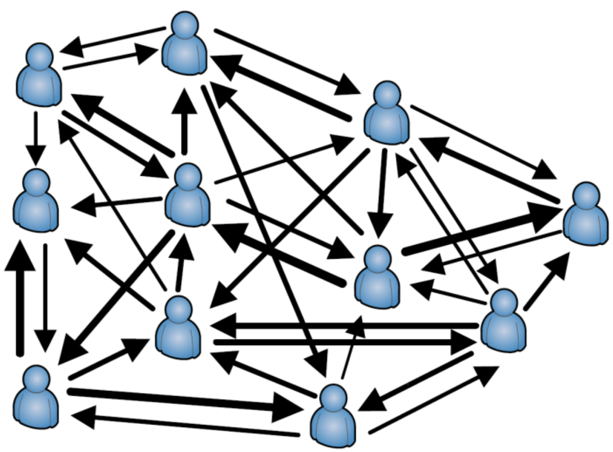
Thm. *Recovers exact structure*

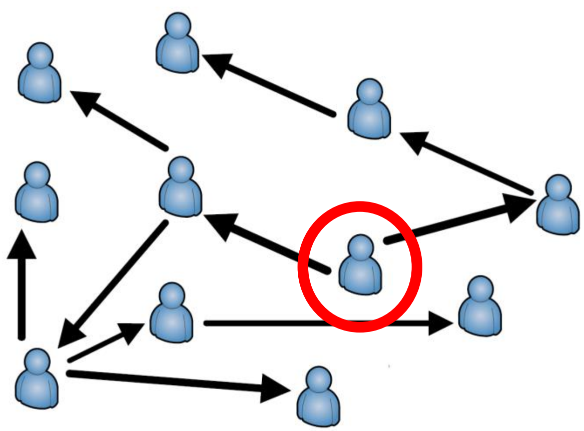
Pf. Markov blanket

[CJQ, NK, TPC. TIT submitted]

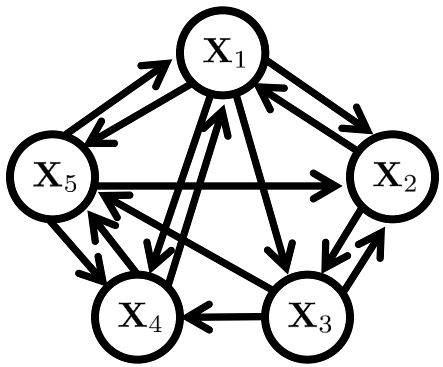
The need for approximations

> 1 billion active users (Sept. 2012)

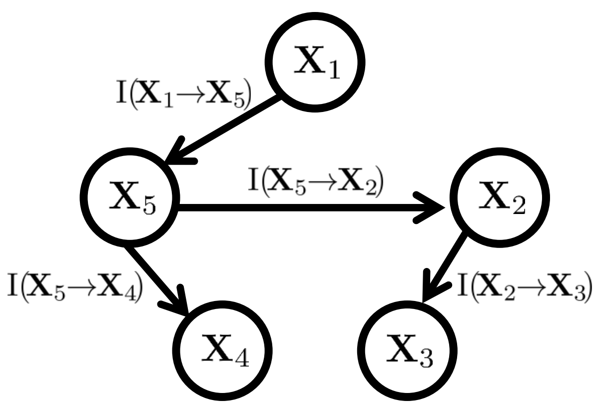




Directed Tree Approximations



$$P_{\underline{\mathbf{X}}}(\underline{\mathbf{x}}) = \prod_{i=1}^{m} P_{\mathbf{X}_i \| \underline{\mathbf{X}}_{[m]\setminus\{i\}}}(\mathbf{x}_i \parallel \underline{\mathbf{x}}_{[m]\setminus\{i\}}).$$



$$\widehat{P}_{\underline{\mathbf{X}}}(\underline{\mathbf{x}}) = \prod_{i=1}^{m} P_{\mathbf{X}_i \| \mathbf{X}_{a(i)}}(\mathbf{x}_i \parallel \mathbf{x}_{a(i)})$$

Theorem [QKC TSP 2013]:

$$\arg\min_{\widehat{P}_{\underline{\mathbf{X}}}} \mathrm{D}(P_{\underline{\mathbf{X}}} \parallel \widehat{P}_{\underline{\mathbf{X}}}) = \arg\max_{\widehat{P}_{\underline{\mathbf{X}}}} \sum_{i=1}^{m} \mathrm{I}(\mathbf{X}_{a(i)} \to \mathbf{X}_i).$$

Algorithm

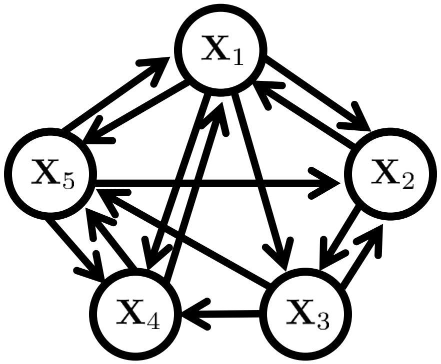
- For edge X→Y, set I(X→Y) as the edge weight
- Run a maximum weight directed spanning tree algorithm (Edmunds)

Properties

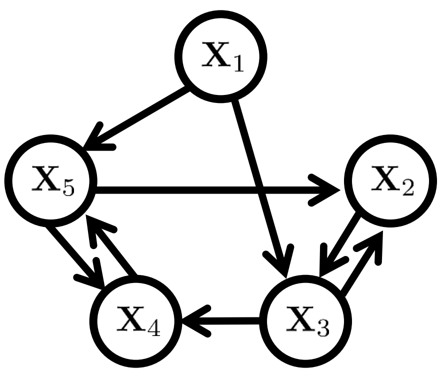
- Only pairwise statistics needed
- Analogous to Chow and Liu (1968)

[CJQ, NK, TPC. TSP 2013]

Bounded In-degree Approximations



$P_{\underline{\mathbf{X}}}(\underline{\mathbf{x}})$



$\widehat{P}_{\underline{\mathbf{X}}}(\underline{\mathbf{x}})$

<u>Theorem</u>:

$$\arg\min_{\widehat{P}_{\underline{\mathbf{X}}}} \mathrm{D}(P_{\underline{\mathbf{X}}} \,\|\, \widehat{P}_{\underline{\mathbf{X}}}) = \arg\max_{\widehat{P}_{\underline{\mathbf{X}}}} \sum_{i=1}^{m} \mathrm{I}(\mathbf{X}_{\underline{\mathrm{A}}(i)} \to \mathbf{X}_i).$$

<u>Properties</u>

- Bounded in-degree K
- Root node
- Graph contains directed spanning tree

[CJQ, NK, AP. ISIT 2013]

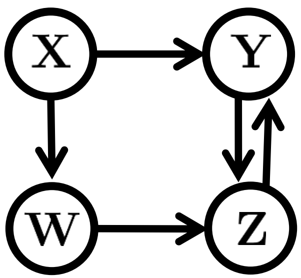
Other Approximations

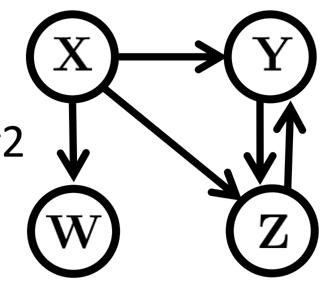
- <u>Greedy Search</u> $\mathbf{X}_1 \leftarrow \arg\max \mathrm{I}(\mathbf{Z} \to \mathbf{Y}), \quad \mathbf{X}_2 \leftarrow \arg\max \mathrm{I}(\mathbf{Z} \to \mathbf{Y} \| \mathbf{X}_1), \ \ldots$

  Theorem: If $\mathrm{I}(\mathbf{X}_{i+1} \to \mathbf{Y} \| \mathbf{X}_1, \ldots, \mathbf{X}_i) \leq \alpha \, \mathrm{I}(\mathbf{X}_i \to \mathbf{Y} \| \mathbf{X}_1, \ldots, \mathbf{X}_{i-1})$

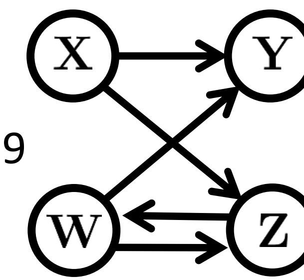
$$\underbrace{\sum_{i=1}^{m} \mathrm{I}(\underline{\mathbf{X}}_{\underline{\mathrm{B}}(i)} \to \mathbf{X}_i)}_{\text{Greedy}} \geq \left(1 - \exp\left(\frac{-K}{\sum_{i=0}^{K-1} \alpha^i}\right)\right) \underbrace{\sum_{i=1}^{m} \mathrm{I}(\underline{\mathbf{X}}_{\underline{\mathrm{A}}(i)} \to \mathbf{X}_i)}_{\text{Optimal}}.$$

- <u>Top Approximations</u>

#1


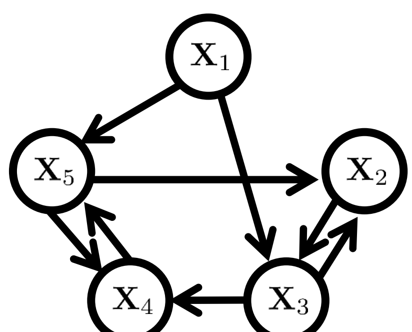
#2


. . .

#19


- <u>Robust Approximations</u>

$$\mathrm{I}(\mathbf{X} \to \mathbf{Y}) \in (0.7, 1.4)$$



$$\sum_{i=1}^{m} \mathrm{I}(\mathbf{X}_{a(i)} \to \mathbf{X}_i) \ \in (1.8, 5.4)$$

[CJQ, NK, AP. ISIT 2013 and others.]

Parametric Estimation

- Parametric distribution

$$\widehat{P}(y_t = 1|y^{t-1}, x^{t-1}, z^{t-1}) = 1/(1 + \exp\{-(\widehat{\alpha}_0 + \sum_{i=1}^{l} \widehat{\alpha}_i y_{t-i} + \widehat{\beta}_i x_{t-i} + \widehat{\gamma}_i z_{t-i})\})$$

- Entropy

$$\widehat{\mathrm{H}}(\mathbf{Y}\|\mathbf{X}, \mathbf{Z}) \triangleq -\frac{1}{n} \log_2 \widehat{P}(\mathbf{Y}\|\mathbf{X}, \mathbf{Z})$$

- Directed Information

$$\widehat{\mathrm{I}}(\mathbf{X} \to \mathbf{Y}\|\mathbf{Z}) \triangleq \widehat{\mathrm{H}}(\mathbf{Y}\|\mathbf{Z}) - \widehat{\mathrm{H}}(\mathbf{Y}\|\mathbf{X}, \mathbf{Z})$$

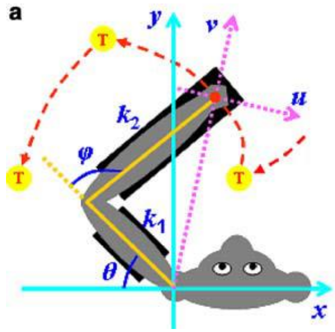
- Minimum description length penalty

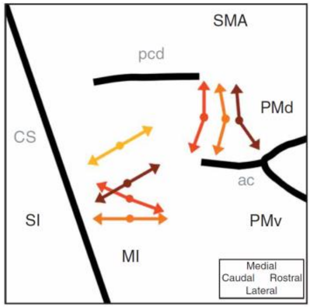
$$\widehat{\mathrm{I}}(\mathbf{X} \to \mathbf{Y}\|\mathbf{Z}) \geq \frac{l \log_2 n}{2n}$$

- Confidence intervals and sample complexity for plug-in empirical, parametric

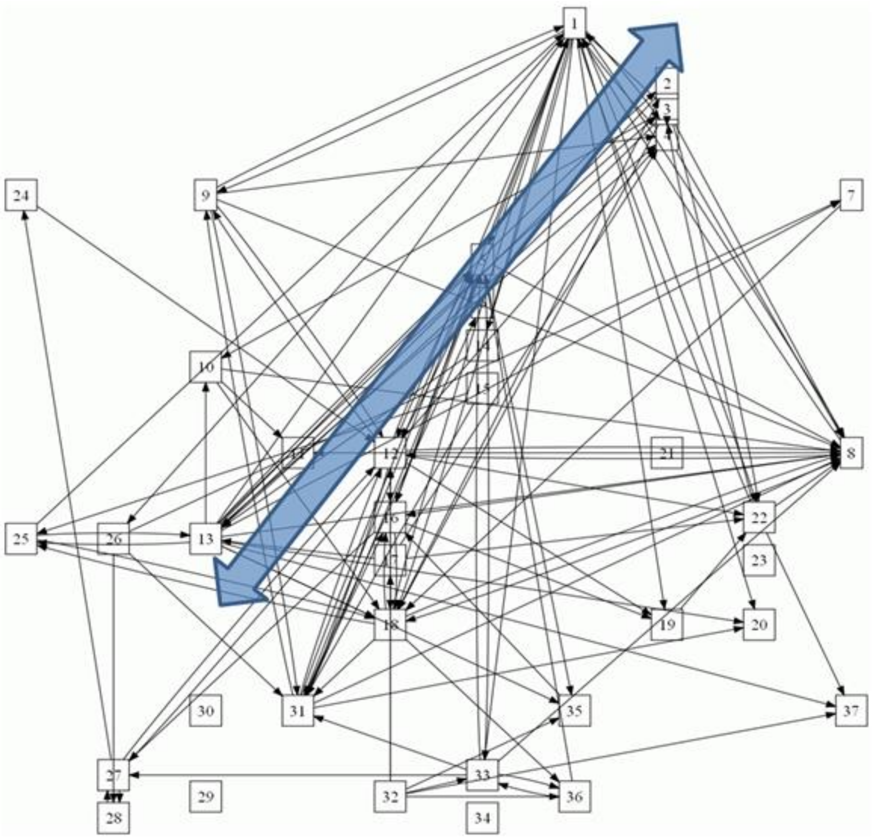
[CJQ, JE, NK, TPC. ISIT 2013] [CJQ, NK, TPC. NGH, JCNS 2011]

Neuroscience analysis





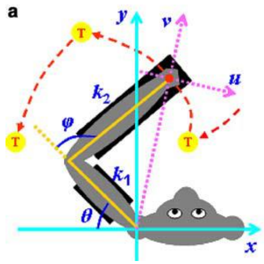
- Primate hand movement experiment
- Simultaneously recorded brain cells in motor control region
- We analyzed activity of individual neurons
- Many edges along upward-right diagonal correspond to direction of information propagation in this brain region

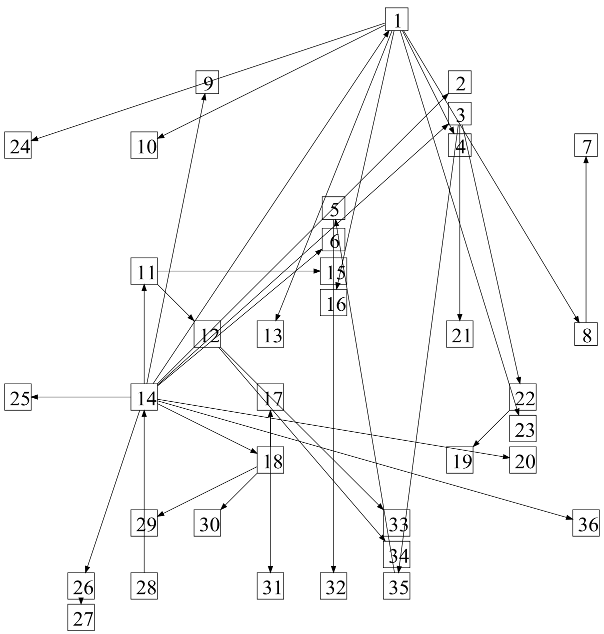


[Rubino et al Nat. Neuro 2006] [CJQ, NK, TPC, NGH JCNS 2011]

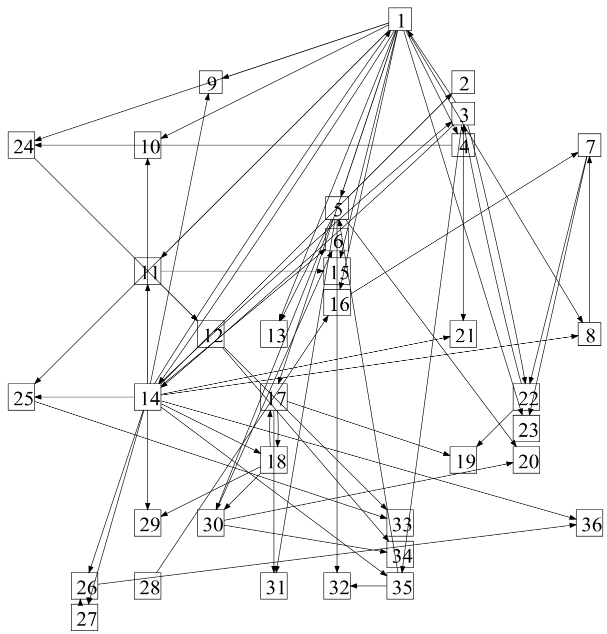
Neuroscience analysis

- Greedy approximation results

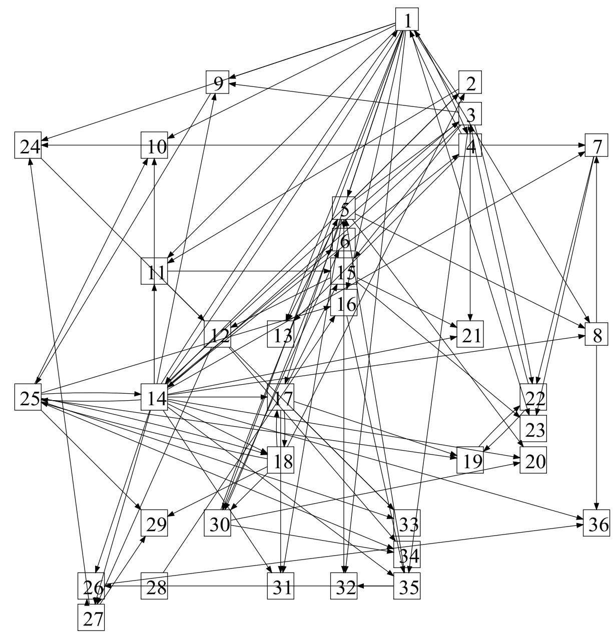




K = 1



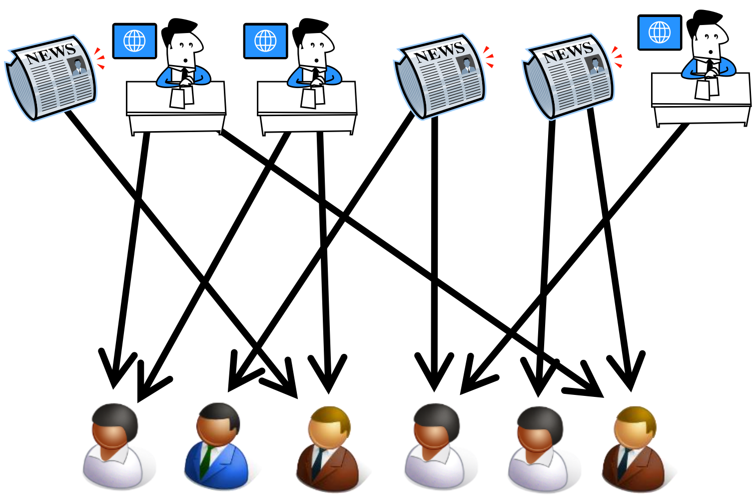
K = 2

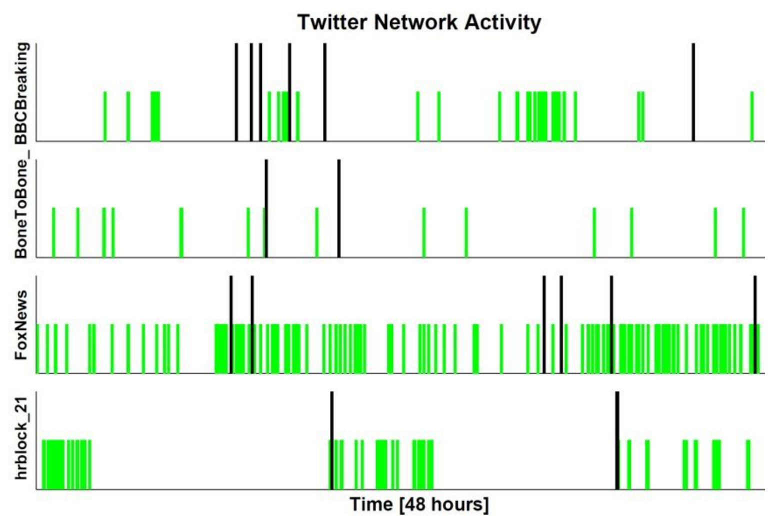


K = 3

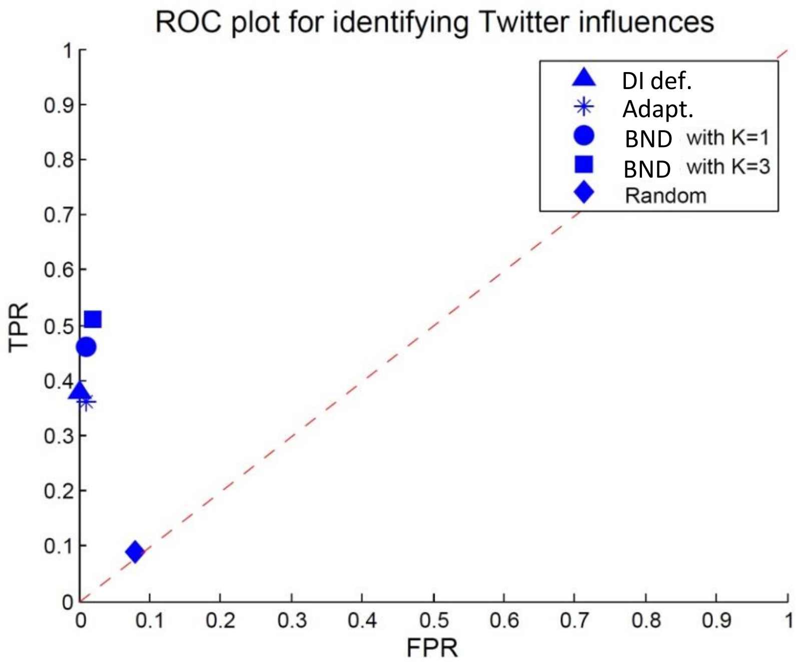
Twitter Analysis

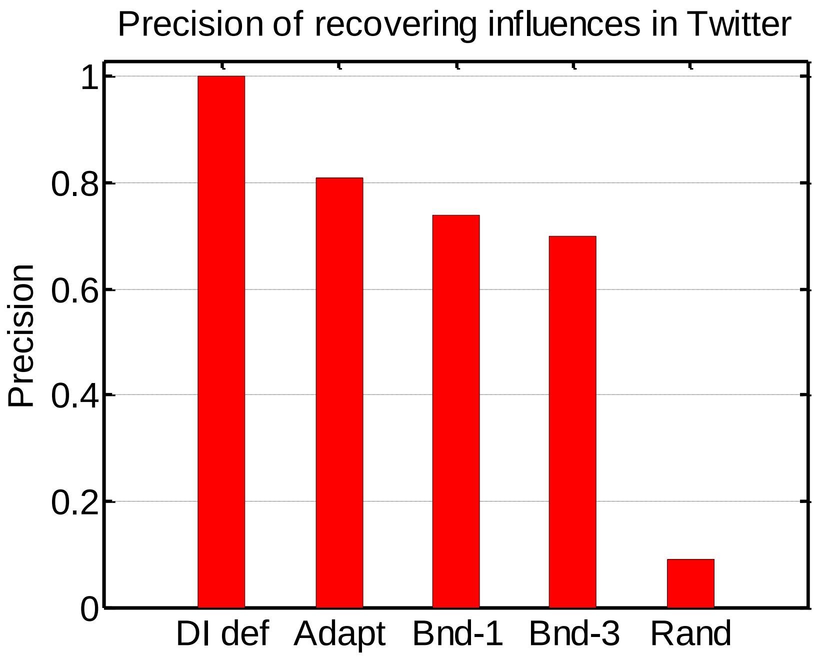
News Agency	Twitter Account Handle
ABC News	**ABC**
ABC World News	**ABCWorldNews**
Agence France-Presse	**AFP**
Al Jazeera English	**AJEnglish**
The Associated Press	**AP**
Al Arabiya English	**AlArabiya_Eng**
BBC Breaking News	**BBCBreaking**
BBC News (World)	**BBCWorld**
Drudge Report	**DRUDGE_REPORT**
Fox News	**FoxNews**
The Jerusalem Post	**Jerusalem_Post**
NBC News	**NBCNews**
Reuters Top News	**Reuters**
Reuters World	**ReutersWorld**
CNN Breaking News	**cnnbrk**
The New York Times	**nytimes**





Twitter Analysis

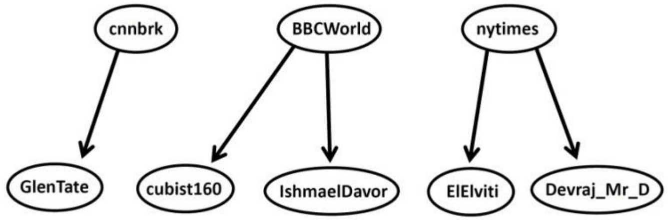




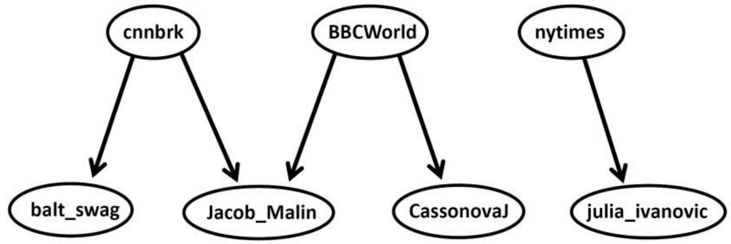
$$TPR = \frac{TP}{TP + FN} \qquad FPR = \frac{FP}{FP + TN}$$

$$\frac{TP}{TP + FP}$$

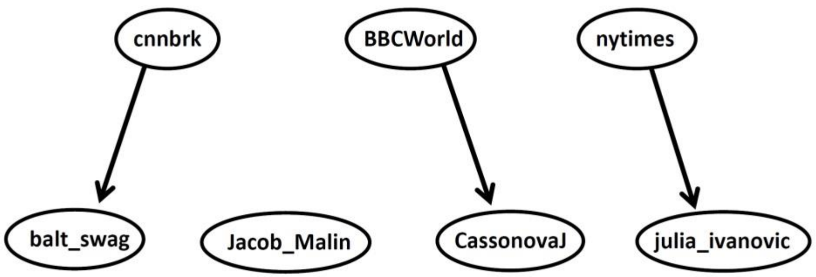
All correct



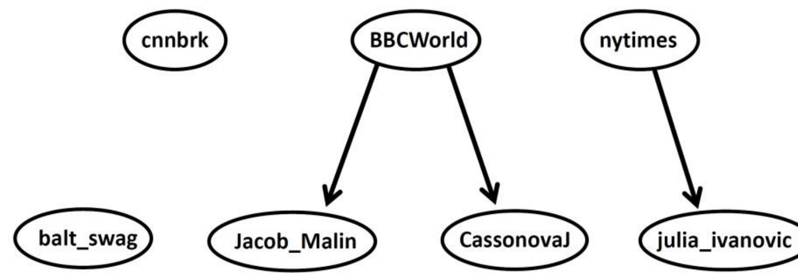
Twitter Analysis



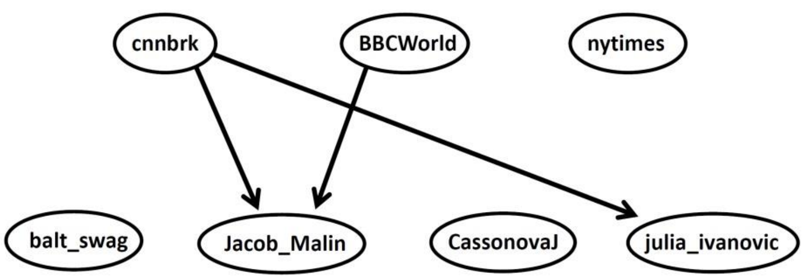
Ground truth



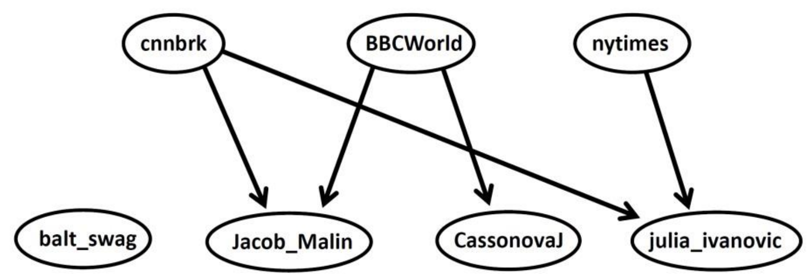
DI definition



Bounded In-degree K = 1



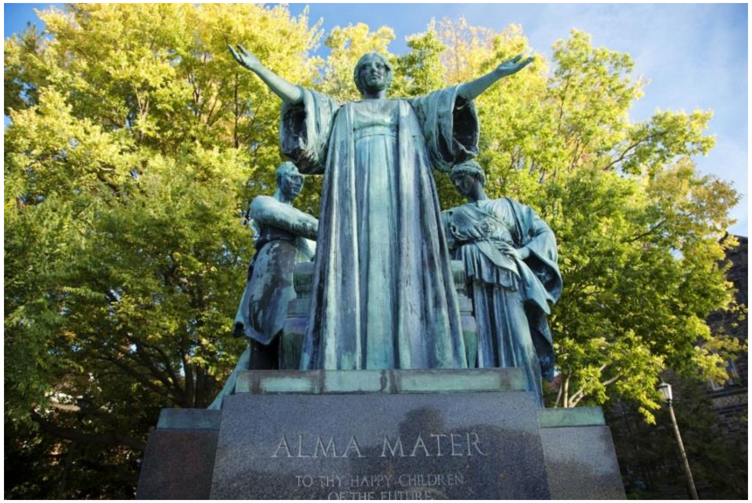
Adaptive



Bounded In-degree K = 3

Thanks

- Krell and CSGF
- Advisors:
  - Negar Kiyavash (UIUC)
  - Todd Coleman (UCSD)
- Collaborators
  - Ali Pinar (Sandia)
  - Nicholas Hatsopoulos (U Chicago)
  - Jalal Etesami (UIUC)



(UIUC Alma Mater statue)