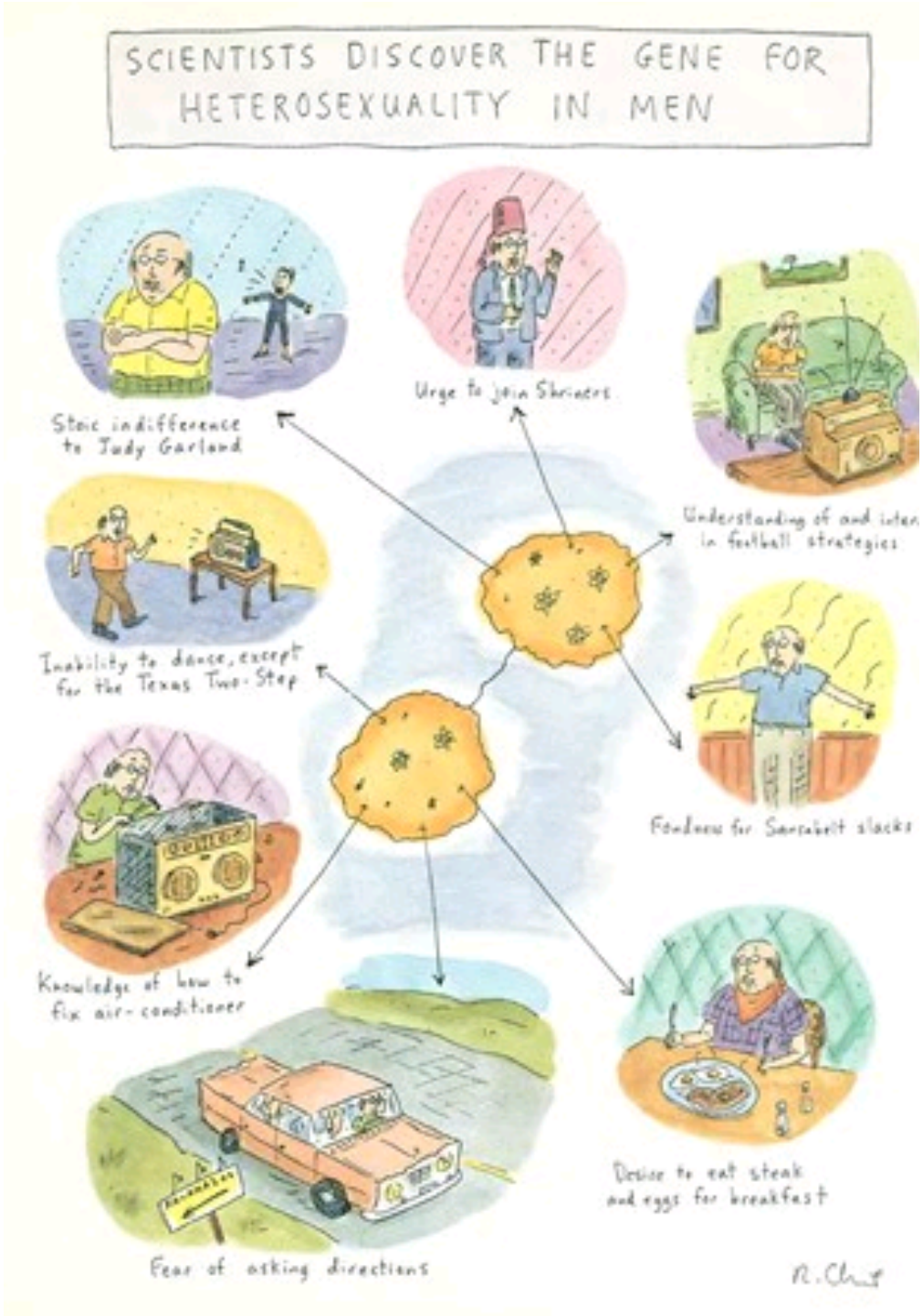
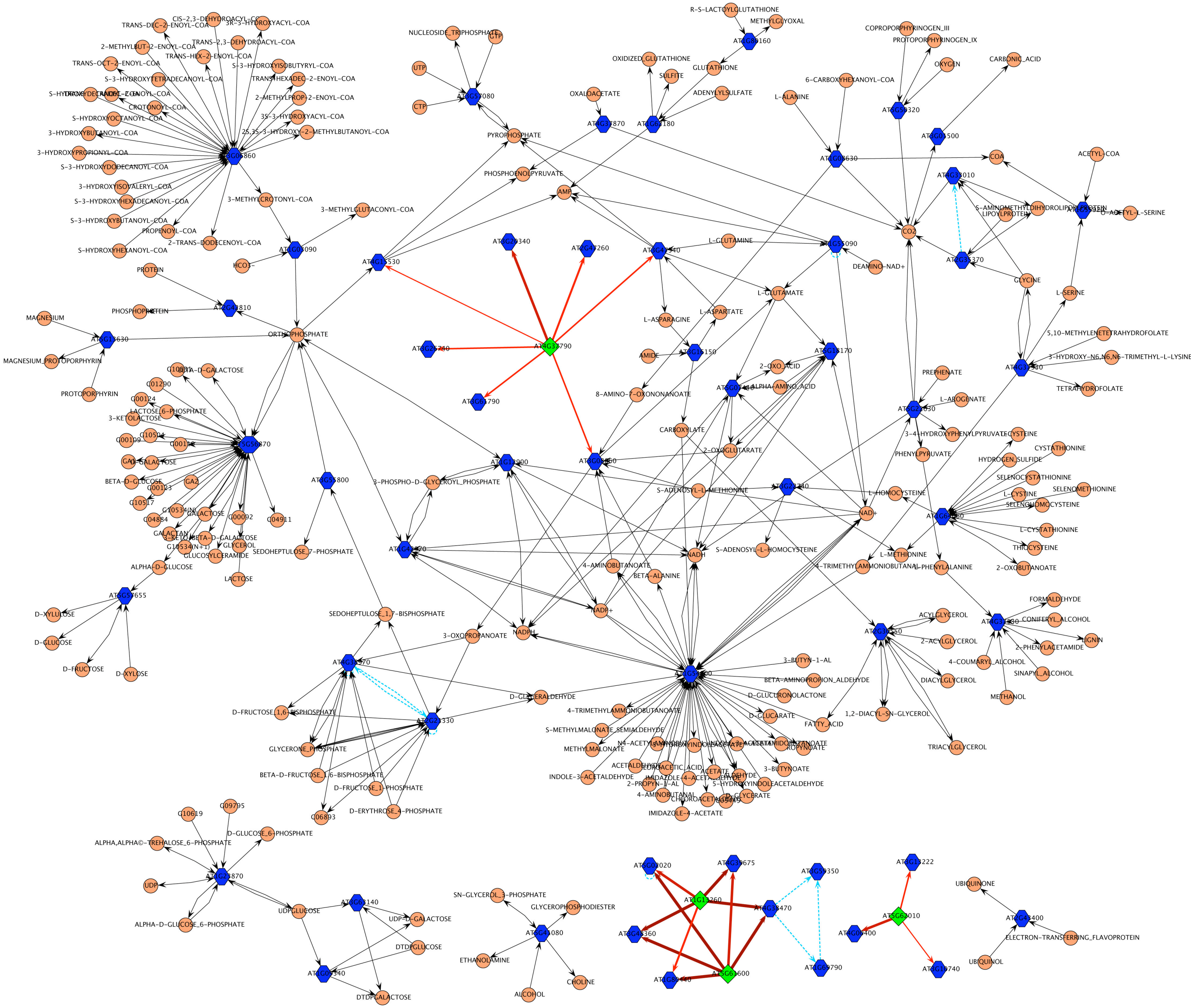
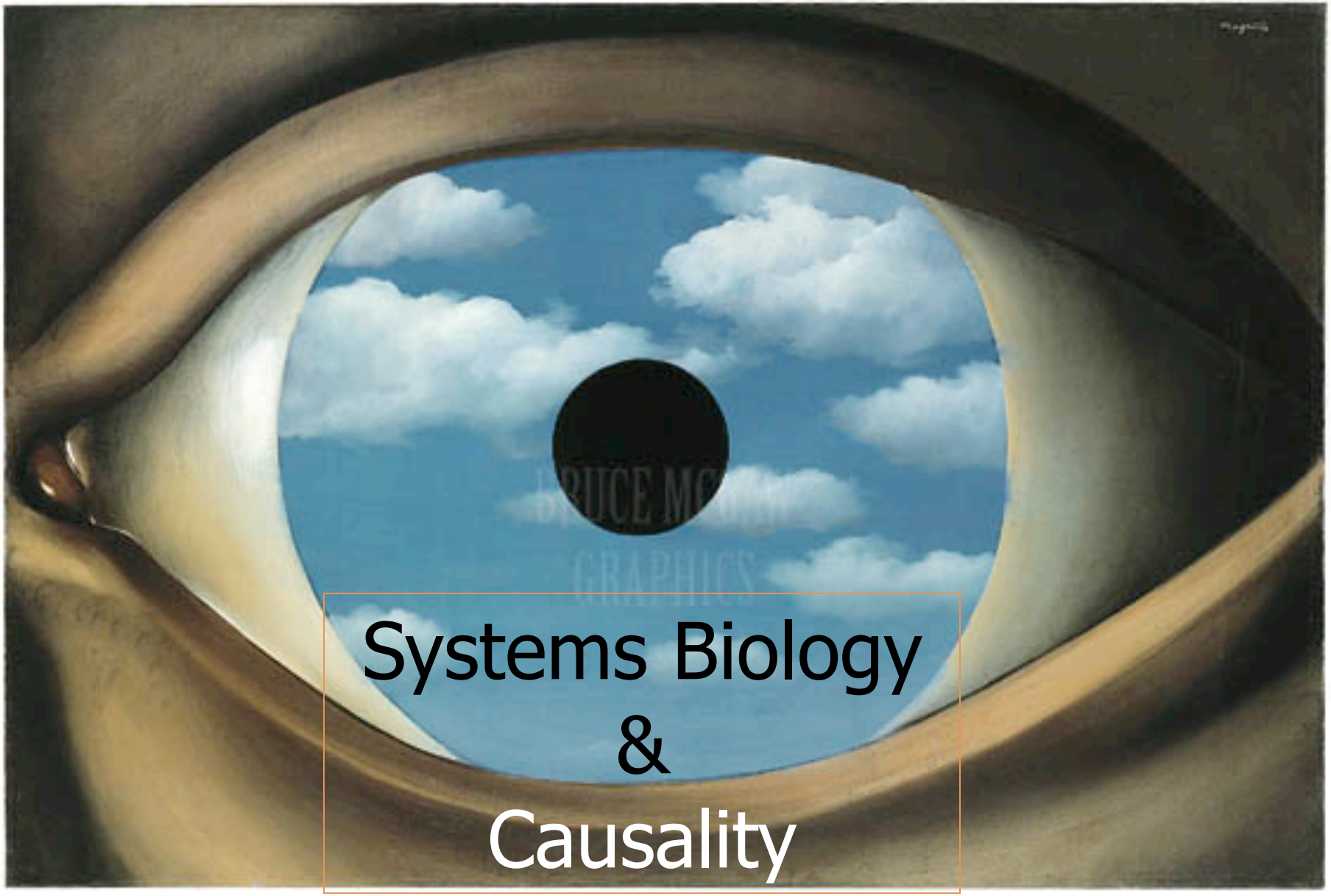

SCIENTISTS DISCOVER THE GENE FOR HETEROSEXUALITY IN MEN
Stoic indifference to Judy Garland
Urge to join Shriners
Understanding of and inter in football strategies
Inability to dance, except for the Texas Two-Step
Fondness for Samsonite slacks
Knowledge of how to fix air-conditioner
Desire to eat steak and eggs for breakfast
Fear of asking directions



NUCLEOSIDE_TRIPHOSPHATE
UTP
CTP
PYROPHOSPHATE
PHOSPHOENOLPYRUVATE
AMP
OXALOACETATE
R-S-LACTOYLGLUTATHIONE
METHYLGLYOXAL
OXIDIZED_GLUTATHIONE
SULFITE
GLUTATHIONE
ADENYLYLSULFATE
COPROPORPHYRINOGEN_III
PROTOPORPHYRINOGEN_IX
OXYGEN
CARBONIC_ACID
6-CARBOXYHEXANOYL-COA
L-ALANINE
COA
ACETYL-COA
CO2
GLYCINE
L-SERINE
5,10-METHYLENETETRAHYDROFOLATE
TETRAHYDROFOLATE
DEAMINO-NAD+
L-GLUTAMINE
L-GLUTAMATE
L-ASPARAGINE
L-ASPARTATE
AMIDE
2-OXO_ACID
8-AMINO-7-OXONONANOATE
CARBOXYLATE
2-OXOGLUTARATE
PREPHENATE
L-AROGENATE
PHENYLPYRUVATE
L-CYSTEINE
CYSTATHIONINE
HYDROGEN_SULFIDE
SELENOCYSTATHIONINE
SELENOMETHIONINE
L-CYSTINE
SELENOHOMOCYSTEINE
L-CYSTATHIONINE
THIOCYSTEINE
2-OXOBUTANOATE
L-HOMOCYSTEINE
L-METHIONINE
L-PHENYLALANINE
S-ADENOSYL-L-METHIONINE
S-ADENOSYL-L-HOMOCYSTEINE
NAD+
NADH
NADP+
NADPH
4-AMINOBUTANOATE
BETA-ALANINE
4-TRIMETHYLAMMONIOBUTANAL
FORMALDEHYDE
CONIFERYL_ALCOHOL
LIGNIN
2-PHENYLACETAMIDE
4-COUMARYL_ALCOHOL
SINAPYL_ALCOHOL
METHANOL
ACYLGLYCEROL
2-ACYLGLYCEROL
DIACYLGLYCEROL
1,2-DIACYL-SN-GLYCEROL
TRIACYLGLYCEROL
FATTY_ACID
3-BUTYN-1-AL
BETA-AMINOPROPION_ALDEHYDE
D-GLUCURONOLACTONE
D-GLUCARATE
4-TRIMETHYLAMMONIOBUTANOATE
S-METHYLMALONATE_SEMIALDEHYDE
METHYLMALONATE
INDOLE-3-ACETALDEHYDE
3-BUTYNOATE
5-HYDROXYINDOLEACETALDEHYDE
IMIDAZOLE-4-ACETATE
2-PROPYN-1-AL
4-AMINOBUTANAL
ACETATE
TRANS-DEC-2-ENOYL-COA
2-METHYLBUT-2-ENOYL-COA
TRANS-2,3-DEHYDROACYL-COA
TRANS-OCT-2-ENOYL-COA
TRANS-HEX-2-ENOYL-COA
S-3-HYDROXYISOBUTYRYL-COA
S-3-HYDROXYTETRADECANOYL-COA
2-METHYLPROP-2-ENOYL-COA
CROTONOYL-COA
S-HYDROXYOCTANOYL-COA
3S-3-HYDROXYACYL-COA
3-HYDROXYBUTANOYL-COA
2S,3S-3-HYDROXY-2-METHYLBUTANOYL-COA
3-HYDROXYPROPIONYL-COA
S-3-HYDROXYDODECANOYL-COA
3-HYDROXYISOVALERYL-COA
S-3-HYDROXYHEXADECANOYL-COA
S-3-HYDROXYBUTANOYL-COA
PROPENOYL-COA
S-HYDROXYHEXANOYL-COA
2-TRANS-DODECENOYL-COA
3-METHYLCROTONYL-COA
3-METHYLGLUTACONYL-COA
HCO3-
PROTEIN
PHOSPHOPROTEIN
MAGNESIUM
MAGNESIUM_PROTOPORPHYRIN
PROTOPORPHYRIN
ORTHOPHOSPHATE
BETA-D-GALACTOSE
LACTOSE_6-PHOSPHATE
3-KETOLACTOSE
GALACTOSE
GALACTAN
GLYCEROL
GLUCOSYLCERAMIDE
LACTOSE
ALPHA-D-GLUCOSE
BETA-D-GLUCOSE
D-XYLULOSE
D-GLUCOSE
D-FRUCTOSE
D-XYLOSE
SEDOHEPTULOSE_7-PHOSPHATE
3-PHOSPHO-D-GLYCEROYL_PHOSPHATE
SEDOHEPTULOSE_1,7-BISPHOSPHATE
3-OXOPROPANOATE
D-GLYCERALDEHYDE
D-FRUCTOSE_1,6-BISPHOSPHATE
GLYCERONE_PHOSPHATE
BETA-D-FRUCTOSE_1,6-BISPHOSPHATE
D-FRUCTOSE_1-PHOSPHATE
D-ERYTHROSE_4-PHOSPHATE
C06893
C04911
C00092
C10619
C09795
ALPHA,ALPHA©-TREHALOSE_6-PHOSPHATE
D-GLUCOSE_6-PHOSPHATE
UDP
ALPHA-D-GLUCOSE_6-PHOSPHATE
UDP-GLUCOSE
UDP-D-GALACTOSE
DTDP-GLUCOSE
DTDP-GALACTOSE
SN-GLYCEROL_3-PHOSPHATE
GLYCEROPHOSPHODIESTER
ETHANOLAMINE
CHOLINE
ALCOHOL
UBIQUINONE
UBIQUINOL
ELECTRON-TRANSFERRING_FLAVOPROTEIN
AT4G37790
AT1G14260
AT5G61600
AT5G62010

Of Exactitude in Science…

- "In that Empire, the craft of Cartography attained such Perfection that the Map of a Single province covered the space of an entire City, and the Map of the Empire itself an entire Province.
- "In the course of Time, these Extensive maps were found somehow wanting, and so the College of Cartographers evolved a Map of the Empire that was of the same Scale as the Empire and that coincided with it point for point.
- "Less attentive to the Study of Cartography, succeeding Generations came to judge a map of such Magnitude cumbersome, and, not without Irreverence, they abandoned it to the Rigours of sun and Rain.
- "In the western Deserts, tattered Fragments of the Map are still to be found, Sheltering an occasional Beast or beggar; in the whole Nation, no other relic is left of the Discipline of Geography."
  - From *Travels of Praiseworthy Men (1658) by J. A. Suarez Miranda (The piece was written by Jorge Luis Borges and Adolfo Bioy Casares)*


Systems Biology
&
Causality

bud mishra

joint work with samantha kleinberg

professor of computer science, mathematics and cell biology

❖

courant, nyu school of medicine, cshl, tifr, & mssm

Hume's Problem

- ***Starting point for virtually all contemporary discussions of causation is David Hume's contribution to the topic***
- Hume sought a total reform of philosophy
- In particular, he aimed to abandon the a priori search for theoretical explanations that supposedly give us insight into the ultimate nature of reality, replacing such (to him) unintelligible propositions with empirical, descriptive inquiry

Hume's Challenge

- "[We] improve by experience, and learn the qualities of natural objects, by observing the effects which result from them. …
- "…[It] is not reasoning which engages us to suppose the past resembling the future, and to expect similar effects from causes which are, to appearance, similar."

Hume's Balls

- "Here is a billiard-ball lying on the table, and another ball moving towards it with rapidity. They strike; and the ball, which was formerly at rest, now acquires a motion… There was no interval betwixt the shock and the motion.
- "Contiguity in time and place is therefore a requisite circumstance to the operation of all causes. 'Tis evident likewise, that the motion, which was the cause, is prior to the motion, which was the effect.
- "Priority in time, is therefore another requisite circumstance in every cause. But this is not all. Let us try any other balls of the same kind in a like situation, and we shall always find, that the impulse of one produces motion in the other.
- "Here, therefore is a third circumstance, viz. that of a constant conjunction betwixt the cause and effect. Every object like the cause, produces always some object like the effect.
- "**Beyond these three circumstances of contiguity, priority, and constant conjunction, I can discover nothing in this cause…**"

Causality

- Main approaches:
  - Regularity
  - Process
  - Counterfactual
  - Probabilistic
  - Statistical

Regularity: Mackie

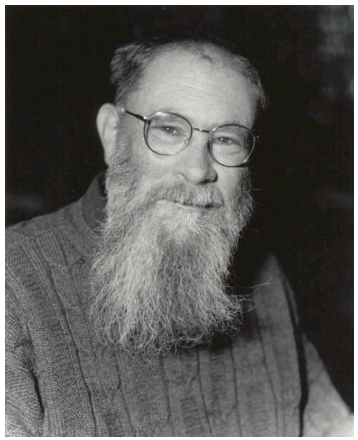
- C is necessary condition of event E if whenever E occurs, C also occurs
- C is a sufficient condition of E if whenever C occurs E also occurs
- "C causes E" is:
  - **an insufficient but non-redundant part of an unnecessary but sufficient condition(INUS)**

John Leslie Mackie. The cement of the universe. Clarendon Press, 1974.

Process Theory: Salmon & Dowe

- **Propagation**
  - Causal process transmits a signal, pseudo process cannot
  - Causal influence propagated through space and time
- CQ is anything science says is universally conserved (e.g. energy, momentum); Causal Process is defined by world lines of an object possessing a CQ
- **Interactions:**
  - Exchanges
  - Intersections
- Causal Interaction: intersection of world lines involving exchange of a CQ

Counterfactuals: Lewis



- Beyond Regularity: Hume also provides a different interpretation:
- *We may define a cause to be an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second. Or, in other words where, if the first object had not been, the second never had existed.*
- Counterfactual: $A \Box\!\!\rightarrow C$: if A was true, C would be true .. If A had not occurred, C would not have occurred.

David Lewis. Causation. The Journal of Philosophy, 70(17):556–567, oct 1973.

Probabilistic Causality: Suppes

- Causes raise probability of their effects
- Causes are temporally prior to their effects
- Relationships are between events
- C is a prima facie cause of E if it is earlier than E and $\mathbf{P(E|C) > P(E)}$
- C, a prima facie cause of E, is a spurious cause of E if there is an S, earlier than C s.t.:

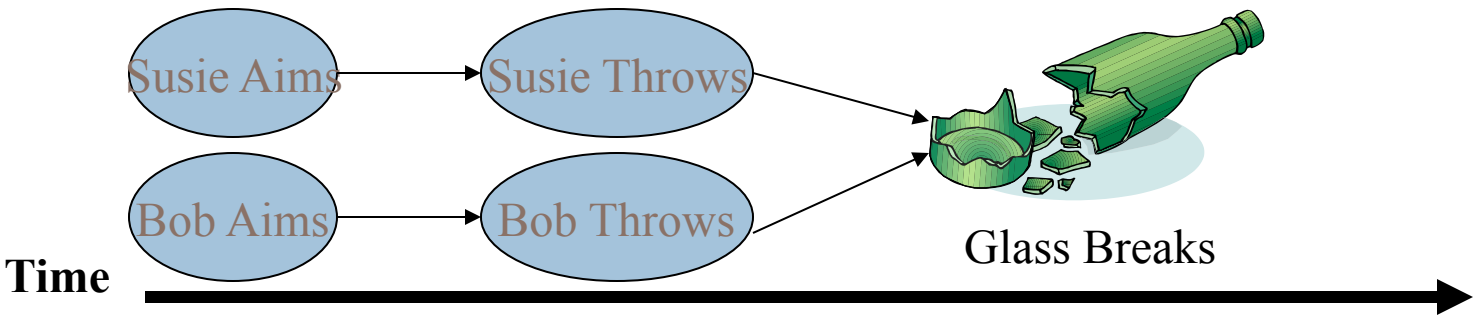
$$\mathbf{P(E|C \wedge S) = P(E|S), \text{ and } P(E|C \wedge S) \geq P(E|C)}$$

- A non-spurious prima facie cause is a genuine cause

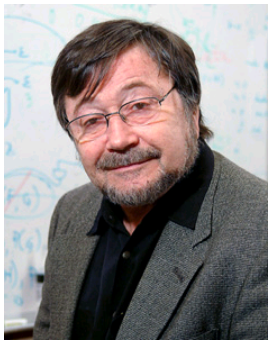
Patrick Suppes. A probabilistic theory of causality. North-Holland, 1970.

Problems with Probabilistic Causality

- Causal chains
- Simpson's Paradox
- Symmetric redundant causation & Preemption
- Many others: e.g., causation by omission, determinism, etc.

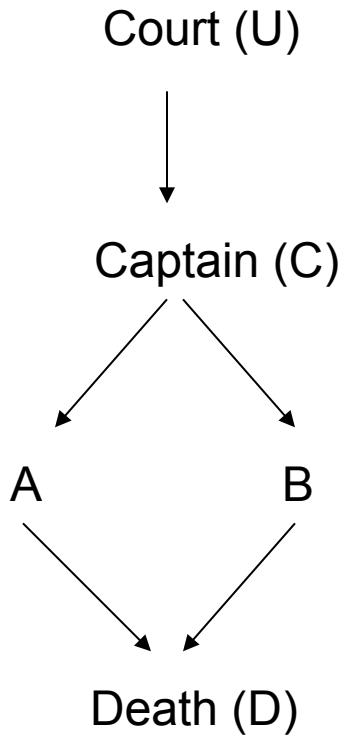


Causality: Pearl



- Structural Equation Model
  - Each variable is a function of its parents and background variables
    - $C = U$, $A = C$, $B = C$, $D = A \vee B$
  - Counterfactual queries:

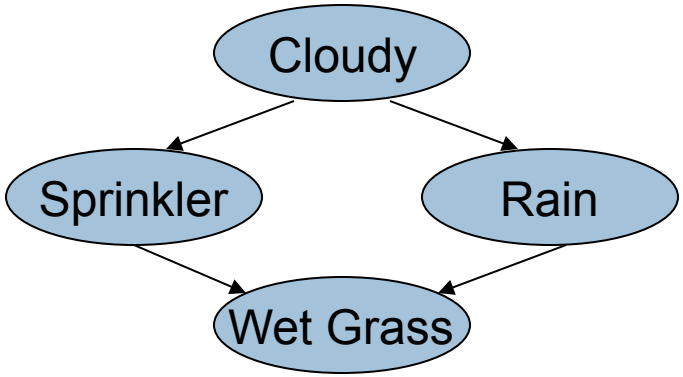
    $D \mapsto D^{\neg A}$



Judea Pearl. Causality: Models, Reasoning, and Inference. Cambridge University Press, 2000.

Graphical Models

- Graphical Model: directed or undirected graph where nodes are variables and (missing) edges represent conditional (in)dependences
  - Compact way to represent joint probability distributions



Problem

- Many types of time course data
  - Neuroscience: Neural spike trains
  - Finance: Stock price movements
  - Internet and Social Networks: Click streams on the internet
  - Biology: Gene expression levels
- How can we find underlying structure of system?
  - Why are the genes co-regulated?
  - What is causing their behavior?


FLIP-TOP BOX
Marlboro
Smoking causes ageing of the skin
FLIP-TOP BOX
Marlboro
Smoking kills
FLIP-TOP BOX
Marlboro
Smoking kills
FLIP-TOP BOX
Marlboro
Smoking causes fatal lung cancer
€ 23,-
€ 23,-
FLIP-TOP BOX
Marlboro
Smoking kills
FLIP-TOP BOX
Marlboro
Smoke contains benzene, nitrosamines, formaldehyde and hydrogen cyanide
FLIP-TOP BOX
Marlboro
Smoking causes fatal lung cancer
FLIP-TOP BOX
Marlboro
Smoking causes fatal lung cancer
€ 23,-
€ 23,-
FLIP-TOP BOX

Motivation

- It is frequently said "smoking causes lung cancer"
  - But, what about other ways of developing cancer, and other conditions required to develop cancer?
- **Goal**: Find details of this relationship
  - How probable is it that someone will get cancer if they smoke?
  - How long will this take to happen?

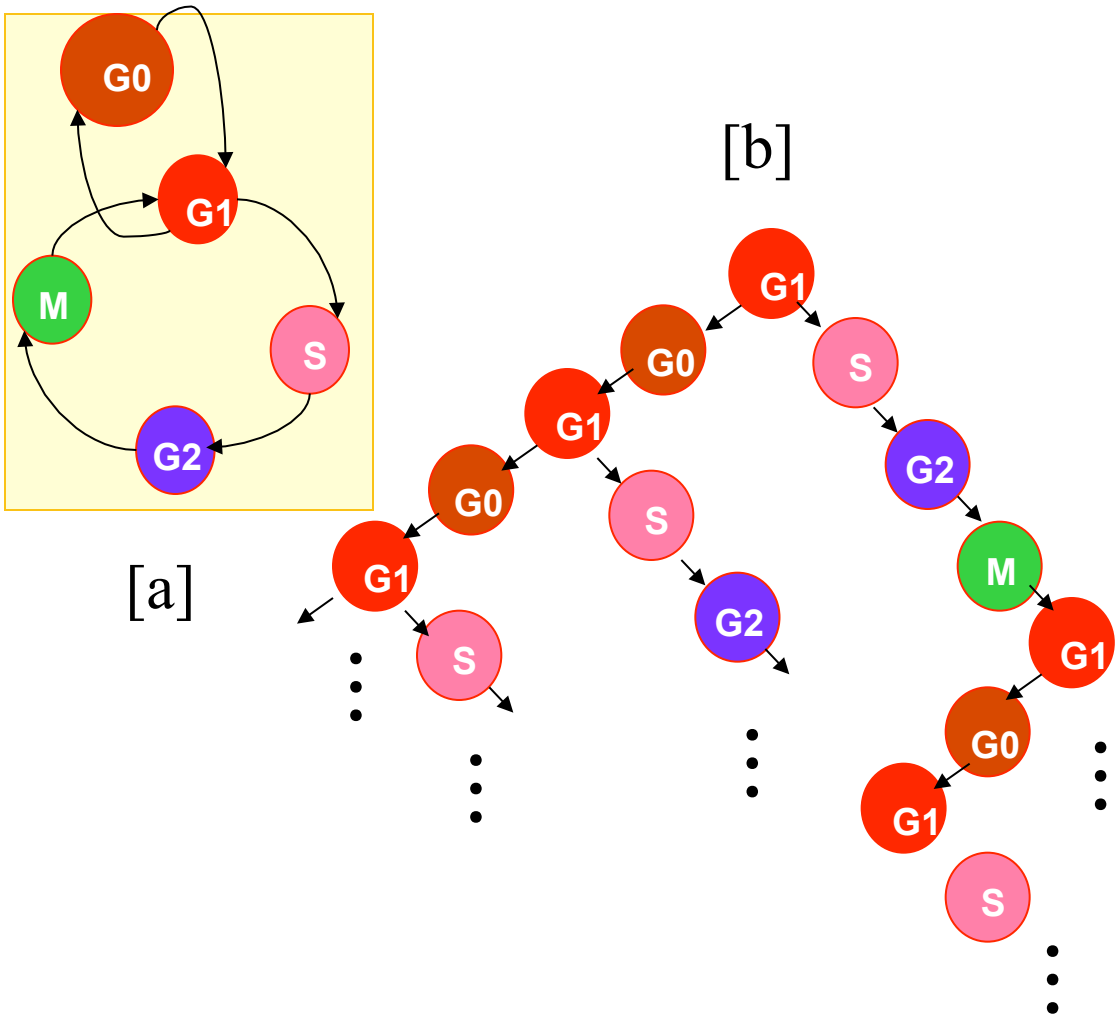
Chance & Time

- Compare:
  - A. Smoking causes lung cancer with probability ≈ 1 after 90 years
  - B. Smoking causes lung cancer with probability = ½ in less than 10 years.
  - Different implications!
- Also, consider other conditions that will make cancer more likely

Desiderata

- A (philosophically) sound notion of causality.
  - It should be able to work with the kinds of data that are available, in a variety of domains
- A (logically) rigorous method of expressing these notions of causality.
  - It should capture a notion of probabilistic nature of the data
  - It should be able reason about time; time must be metric, capturing a notion of locality
- An (algorithmic) automated method for finding all prima-facie causes
  - Model Checking
- A (statistically) sound method for finding all genuine causes.

Computation Tree



- Finite set of states; Some are initial states
- **Total** transition relation: every state has at least one next state i.e. infinite paths
- There is a set of basic environmental variables or features ("atomic propositions")
- In each state, some atomic propositions are true

Basics of PCTL

- Probabilistic extension of CTL
  - Transitions are probabilistic
- Formulas interpreted over structures $<S, s_i, T, L>$
  - S: finite set of states
  - $s_i \in S$: an initial state
  - T: transition probability function,

$$T: S \times S \rightarrow [0,1] \text{ such that for all } s \text{ in } S$$

$$\sum_{t \in S} T(s, t) = 1$$

  - L: a labeling function assigning atomic propositions to states
    L: $S \rightarrow 2^A$

PCTL Formulas

- Atomic propositions a in A
- Boolean connectives ($\neg$, $\wedge$, $\vee$, $\rightarrow$)
- State formulas:
  - Atomic propositions
  - $\neg f$, $f \wedge g$, $f \vee g$, $f \rightarrow g$
  - $[h]_{\geq p}$ and $[h]_{>p}$,$0 \leq p \leq 1$
- Path formulas:
  - $f\ U_{\leq t}\ g$, $f\ W_{\leq t}\ g$, where t is non-negative or infinity; f and g are state formulas, & h is a path formula

Overview of Semantics

- s satisfies a in AP if a in L(s)
- $\neg, \wedge, \vee, \rightarrow$ are defined as normal
- $[f]_{\geq p}$ (resp. $[f]_{>p}$) holds for a state s if the sum of probabilities of paths from s satisfying f is $\geq$ p ($>$p)
- U is strong until, W is weak until

Some expressible properties

$$1.\ Af \equiv [f]_{\geq 1}$$

$$2.\ Ef \equiv [f]_{>0}$$

$$3.\ G^{\leq t}_{\geq p} f \equiv fW^{\leq t}_{\geq p} false$$

$$4.\ F^{\leq t}_{\geq p} f \equiv trueU^{\leq t}_{\geq p} f$$

$$5.\ AGf \equiv fW^{\leq \infty}_{\geq 1} false$$

$$6.\ AFf \equiv trueU^{\leq \infty}_{\geq 1} f$$

$$7.\ EGf \equiv fW^{\leq \infty}_{>0} false$$

$$8.\ EFf \equiv trueU^{\leq \infty}_{>0} f$$

Model Checking

- **Basic steps**
  - Modeling
    - Convert system into standardized format
  - Specification
    - State properties we want the system to satisfy
  - Verification
    - Test whether model satisfies these properties
- To see if structure K satisfies formula f:
  - Take subformulas of f
  - Label each state in K with subformulas that are true within that state (beginning with atomic propositions)
  - If initial state is in set of states labeled with f then

    K satisfies f

Checking a probabilistic formula:

- For state s, P(t,s) is sum of probabilities for set of paths starting in s satisfying formula
- If t< 0, define P(t,s)=0
- For t ≥ 0:

$$f\ U^{\leq t}_{\geq p} g, t \neq \infty$$

$P(t,s)=$

if g in labels(*s*)
    1
else if f not in labels(*s*)
    0
else

$$\sum_{s' \in S} T(s,s') \cdot P(t-1,s')$$

Leads to

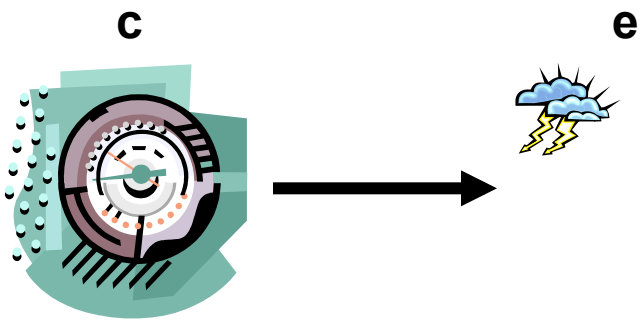
$$f_1 \rightsquigarrow_{\geq p}^{\leq t} f_2 \equiv AG[(f_1 \rightarrow F_{\geq p}^{\leq t} f_2)]$$

- Derived operator
- "for all paths, at every state, if f_1 then eventually f_2 within t time units with probability at least p"
- Means that there can be any number of transitions between f_1 and f_2
- Transitions must happen within t time units

Types of causes

- **Prima facie**: Positively associated with effect; *Potential* causes
- **Spurious**: No (or little) influence on effect; Other causes account better for the effect
- **Genuine**: Non-spurious prima facie causes
- **Supplementary**: Two prima facie causes may aid each other in producing effect
- Next, define these in terms of PCTL

Prima facie causes



- c has non-zero probability
- Probability of e given c is greater than general probability of e

$$F_{>0}^{\leq\infty} c$$

$$c \overset{\leq t}{\underset{\geq p}{\rightsquigarrow}} e$$

$$F_{<p}^{\leq\infty} e$$

ε -spurious causes

- **A la Patrick Suppes:** if there is an earlier x s.t.

$$P(e|c \wedge x) = P(e|x), \&$$
$$P(e|\ c \wedge x) \geq P(e|\ c)$$

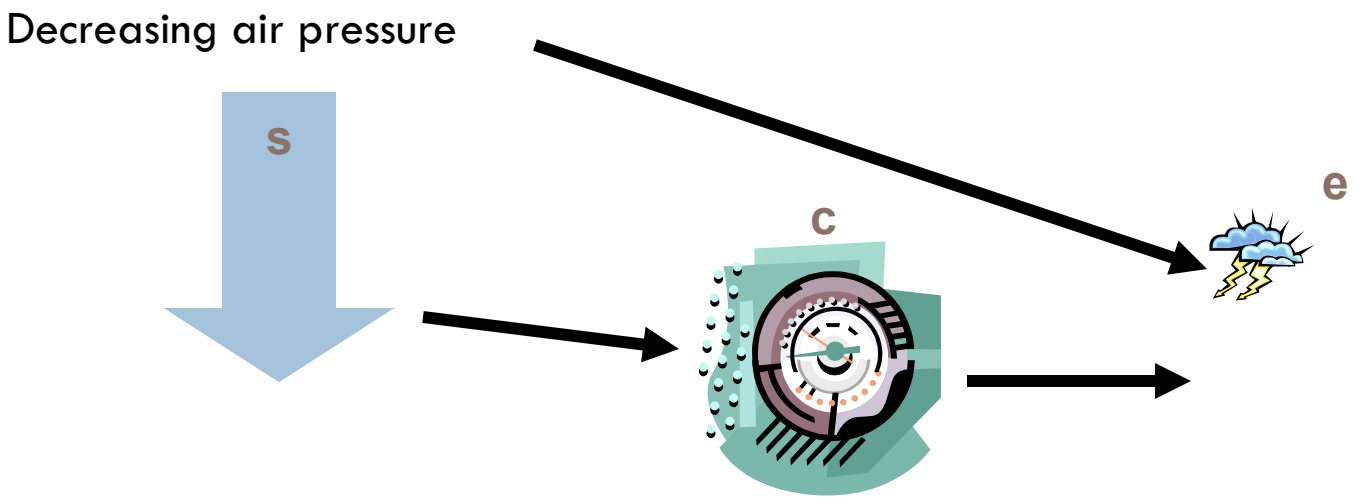
  Also, ε-spuriousness:

$$|P(e|c \wedge x) - P(e|x)| < \varepsilon$$

- **A la Ellery Eells:** look at any factors earlier than effect, for set of n, 2^n ways of holding these fixed. Compute average difference in probability, with respect to these background contexts

Finding spurious causes

- X = set of prima facie causes of e \ c
- c = one prima facie cause
- For each, estimate the probability of transitioning to e state from c $\wedge$ x state vs ($\neg$c) $\wedge$ x state
  - E.g., Probability of rain given decreasing air pressure AND falling barometer, versus decreasing air pressure and NOT falling barometer



Calculating spuriousness

- Need not consider *all* other events; just other prima facie causes of e
- Why?
  - Provides a way to narrow down the factors that must be considered

$$\varepsilon_x = P(e \mid c \wedge x) - P(e \mid \neg c \wedge x)$$

$$\varepsilon_{avg} = \sum_{x \in X} \varepsilon_x / |X|$$

Definitions

- Spurious Cause
  - c is an ε-spurious cause of e if:
    - c is a prima facie cause of e
    - and $\varepsilon_{avg} < \varepsilon$
- Genuine Cause
  - c is a genuine cause of e if it is a non-spurious prima facie cause

What ε?

- Could use background knowledge
- Perform simulations
- BUT, we are testing systems with a lot of data
  - Can use this to our advantage
  - Multiple hypothesis testing

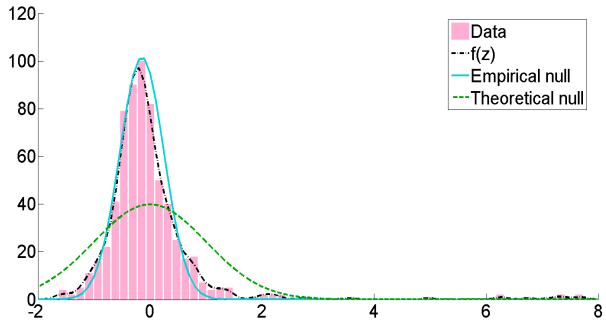
FDR

	# not rejected	# rejected	totals
# true null H	U	V (F+)	m_0
# non- true null H	T (F-)	S	m_1
totals	m - R	R	m

- FDR = V/R
- Local FDR (fdr)
  - For each hypothesis, compute probability of it being null

Two groups of data

- Two classes of prior probabilities
  - p_0 = Pr(uninteresting), $f_0(z)$ density
  - p_1 = Pr(interesting), $f_1(z)$ density
- Assume p_0 large.
- Mixture density:
  - $f(z) = p_0 f_0(z) + p_1 f_1(z)$
- Prob of being uninteresting given z-value z
  - $\text{fdr}(z) \approx \Pr(\text{null} \mid z) = p_0 f_0(z) / f(z)$



Steps

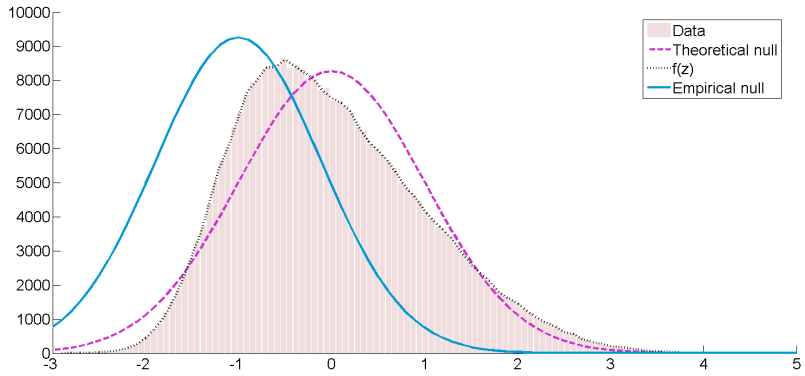
- 1. Estimate distribution of data, f(z)
  - E.g. splines or Poisson regression
- 2. Define null density $f_0(z)$ from data
  - One method is to fit to central peak of data.
- 3. Calculate fdr(z)
- 4. Call H_i where $fdr(z_i) <$ threshold interesting
  - Common threshold is 0.005

Causal Inference

- Enumerate logical formulas describing possible causes
- From experimental data determine prima facie causes
- Calculate ε for each, translate to z-values
- Take set of z values, calculate empirical null, label prima facie causes with z-value where fdr(z) < threshold as genuine

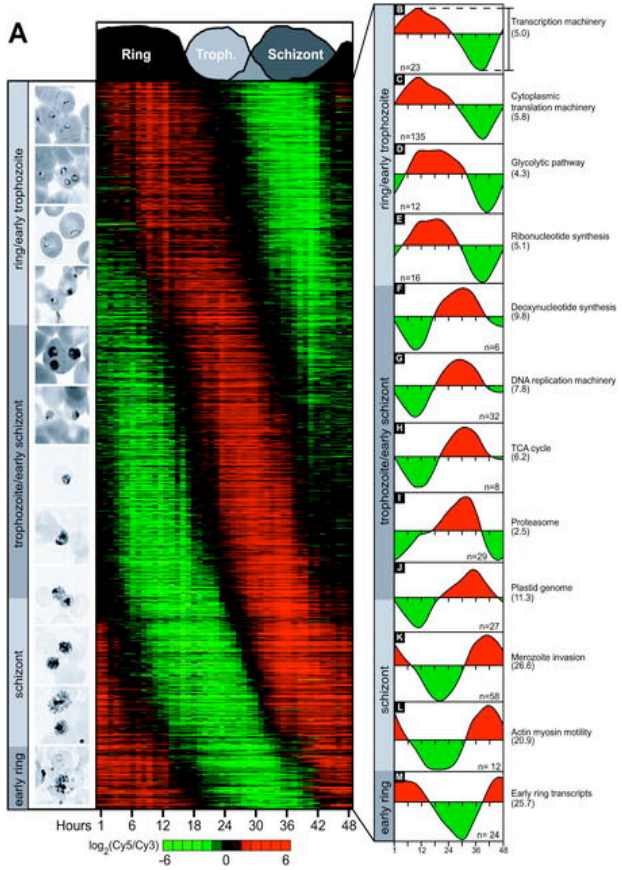
Cellular data

- Looked at relationships between pairs of genes where relationship takes place at next unit of time
- Empirical null: N(-1.00,0.89)
- Thousands of prima facie causes where f(z)< 0.1

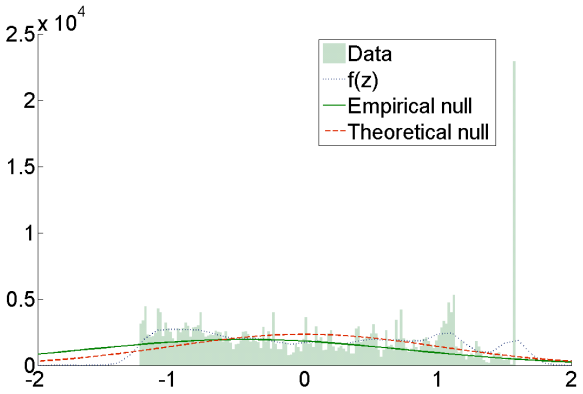


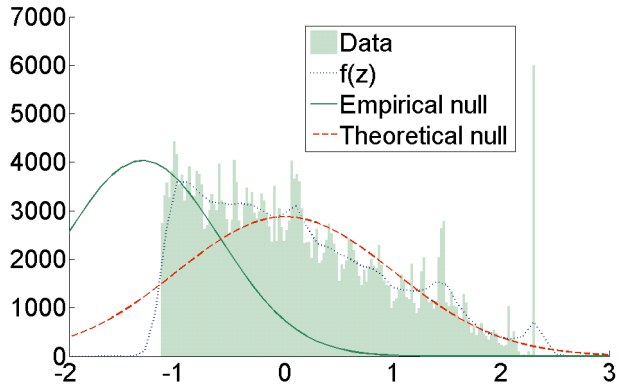
Microarray data

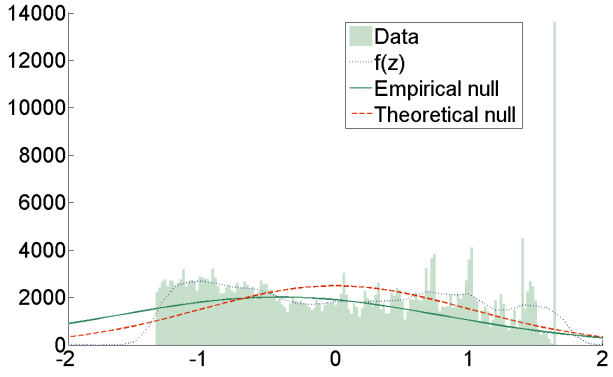
- Microarray gene expression data from the 48-hour Intraerythrocytic developmental cycle(IDC) of *P. falciparum*
- Most deadly form of malaria
- IDC (blood stage) is stage that produces all malaria symptoms
- All genes active at some point during IDC

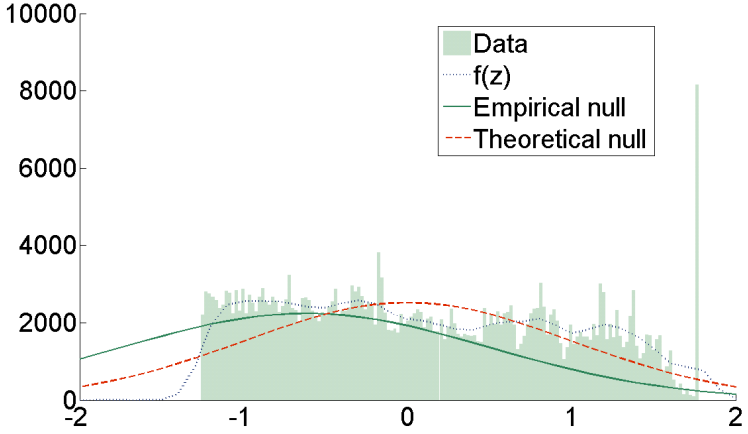


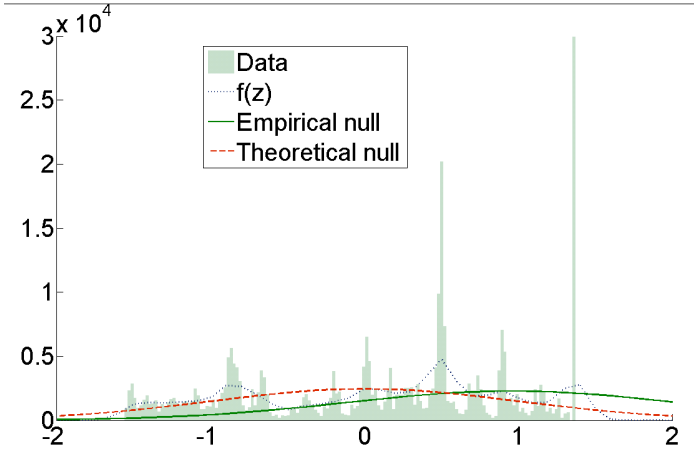
Analyzing Stage-by-stage

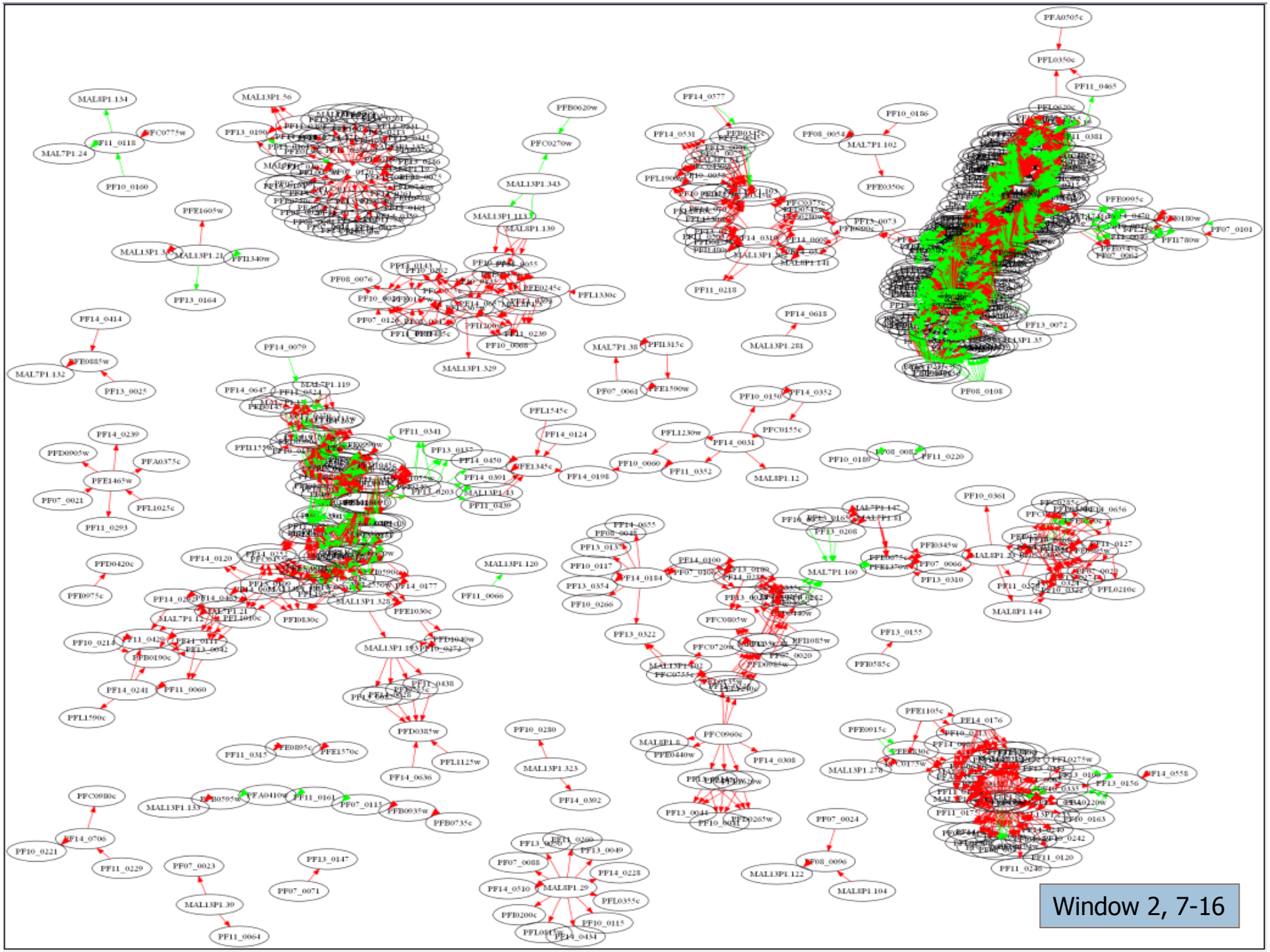


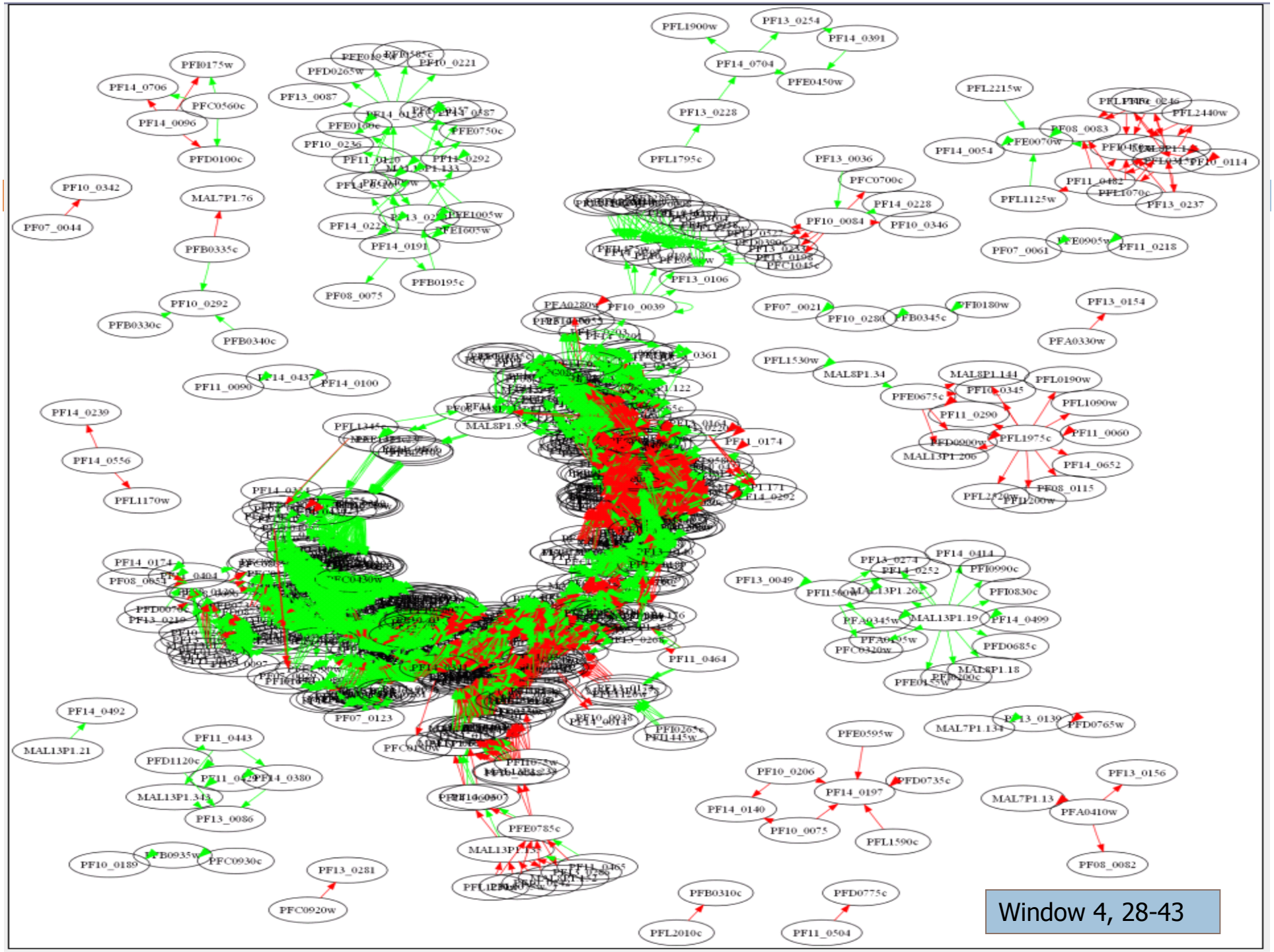






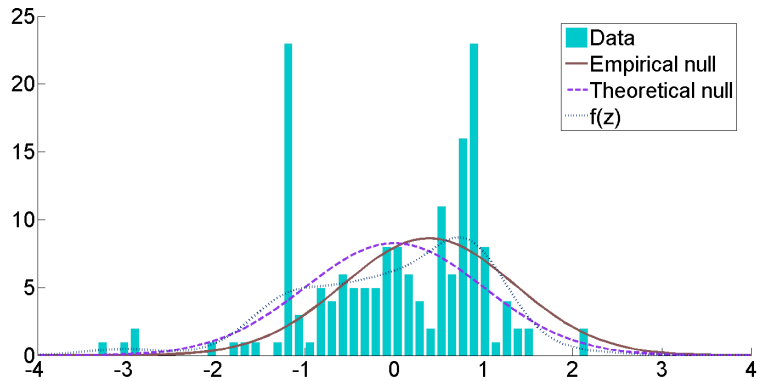



Window 2, 7-16

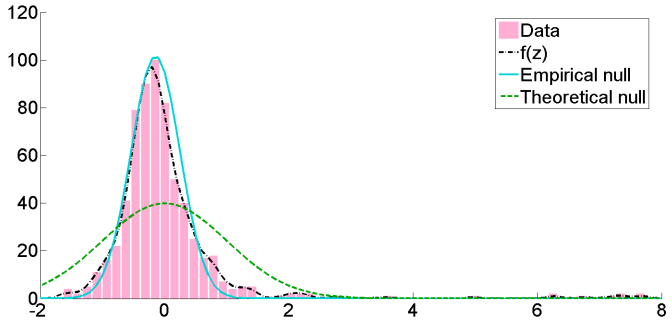

PFL1900w
PF13_0254
PF14_0391
PF14_0704
PFE0450w
PF13_0228
PFL1795c
PFL2215w
PFL2440w
PF08_0083
PF14_0054
PFE0070w
PF10_0114
PF11_0482
PFL1070c
PFL1125w
PF13_0237
PF13_0036
PFC0700c
PF14_0228
PF10_0084
PF10_0346
PF07_0061
PFE0905w
PF11_0218
PF10_0342
PF07_0044
MAL7P1.76
PFB0335c
PF10_0292
PFB0330c
PFB0340c
PFD0265w
PF10_0221
PF13_0087
PFE0160c
PFE0750c
PF10_0236
PF13_0292
PFE1005w
PFE1005w
PF14_0191
PFB0195c
PF08_0075
PF13_0106
PF10_0039
PF07_0021
PF10_0280
PFB0345c
PFI0180w
PF13_0154
PFA0330w
PFL1530w
MAL8P1.34
MAL8P1.144
PFL0190w
PFE0675c
PF10_0345
PFL1090w
PF11_0290
PFD0900w
PFL1975c
PF11_0060
MAL13P1.206
PF14_0652
PF08_0115
PFL2520w
PFL1200w
PF14_0239
PF14_0556
PFL1170w
PF11_0090
PF14_0100
PFL1345c
MAL8P1.93
PF11_0174
PF14_0174
PF08_0053
PF13_0049
PFI1580w
PF14_0414
PF14_0252
PFI0990c
PFI0830c
MAL13P1.262
PFA0345w
MAL13P1.19
PF14_0499
PFC0320w
PFD0685c
MAL8P1.18
PFE0155w
PF11_0464
PF07_0123
PF14_0492
MAL13P1.21
PF11_0443
PFD1120c
PF14_0380
MAL13P1.343
PF13_0086
PF10_0189
PFB0935w
PFC0930c
PF13_0281
PFC0920w
PFE0265c
PFI1445w
PFE0595w
PF10_0206
PFD0735c
PF14_0197
PF14_0140
PF10_0075
PFL1590c
MAL7P1.134
PF13_0139
PFD0765w
PF13_0156
MAL7P1.13
PFA0410w
PF08_0082
PFB0310c
PFL2010c
PFD0775c
PF11_0504
Window 4, 28-43

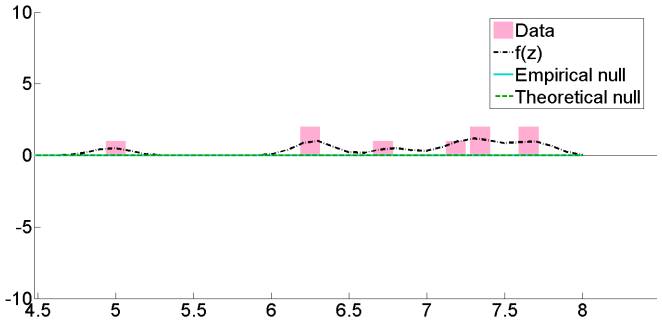
Political data

- Empirical null: N(0.39,0.96)
- No genuine causes with z>0, but look at z<0
  - 3 phrases with false discovery rate, fdr<0.1, all have z around -3
  - Homes, progress, lebanon
- What does this mean?
  - For example “had President Bush NOT said homes, his rating would have gone down”

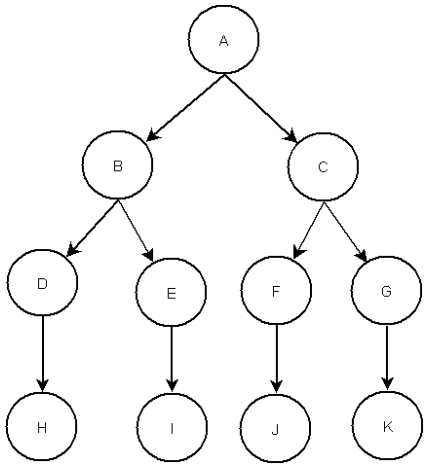


Neural data





- We used the multiple hypothesis testing framework
- Empirical null: N(-0.15,-0.39)
- Genuine causes have $z>3$



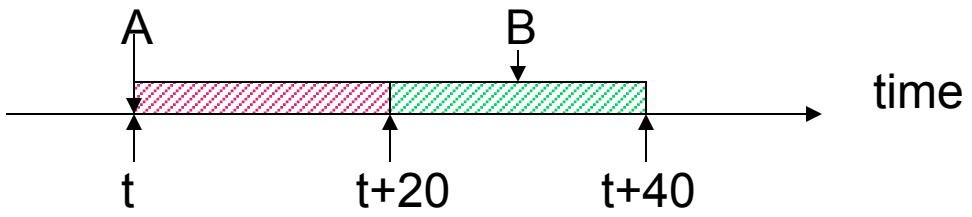
Neural spike trains

- Simulation of neural spike trains
- 26 neurons
- 5 causal structures
  - For each, 2 data sets generated for high and low noise
  - Relationships can be many to many
- 100,000 firings
- At each time point:
  - Neuron can fire randomly (dependent on noise level)
  - Neuron can be triggered by one of the neurons that causes it to fire

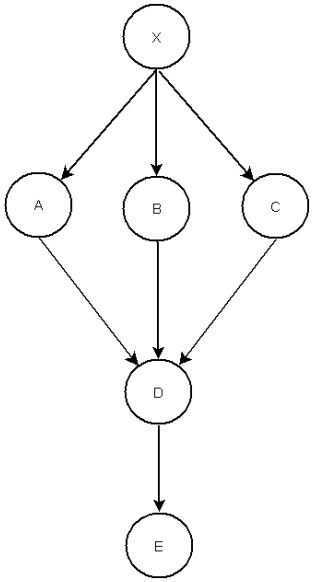
Data from 2006 KDD workshop on temporal data mining.
K.P. Unnikrishnan, Naren Ramakrishnan, P.S. Sastry.

Data set, continued

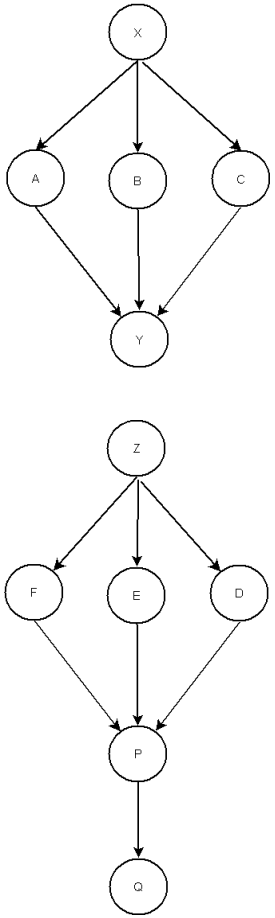
- Structures
  - All directed acyclic graphs (DAGs)
  - Range from chains of neurons to binary trees
- Known information
  - Neuron has 20 time unit refractory period
  - Window of 20 time units after refractory period when it can activate another neuron
  - Only simulated positive causal influence



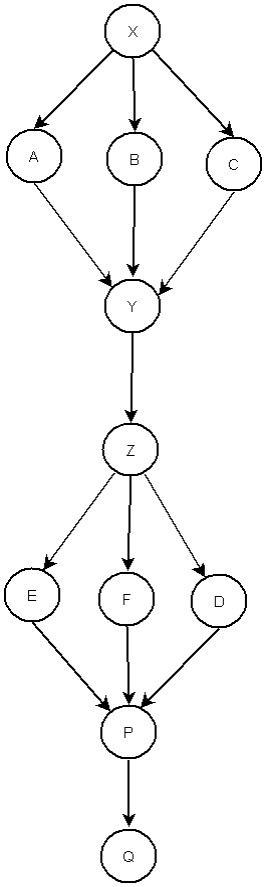
Patterns 1-3



Pattern 1



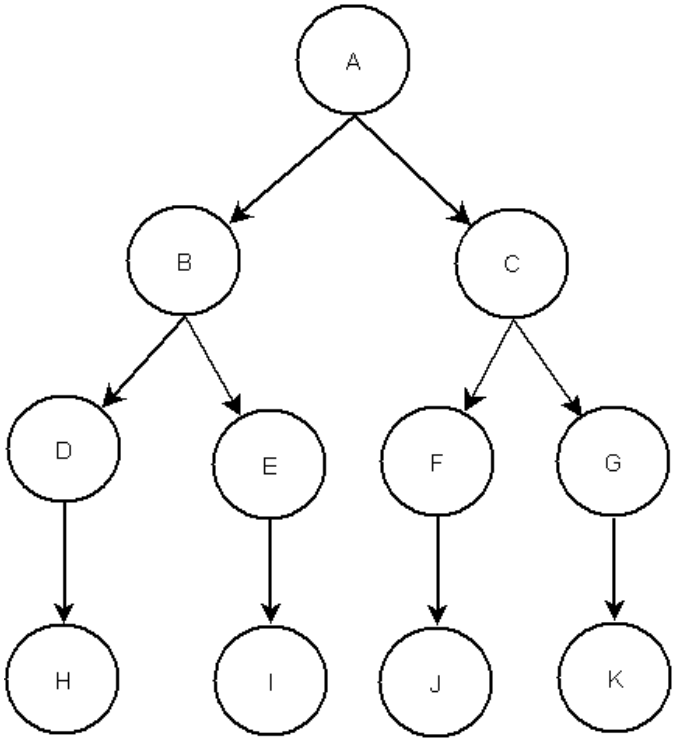
Pattern 2

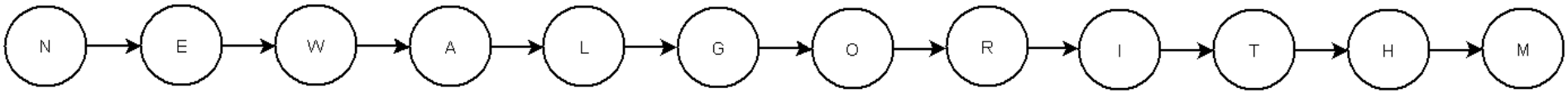


Pattern 3

Patterns 4 and 5

Pattern 4





Pattern 5

Results

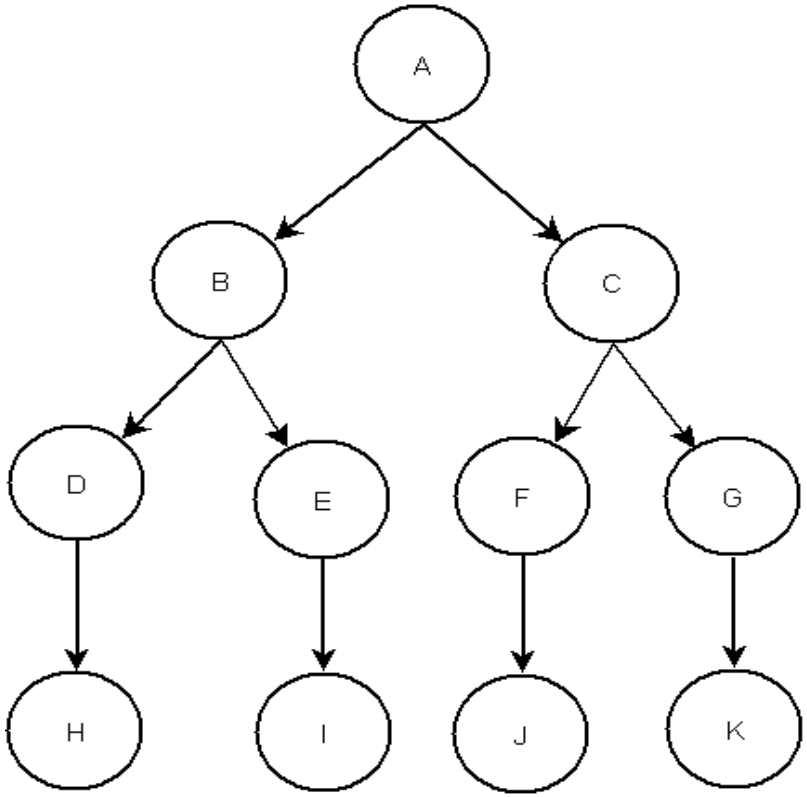
- Used known time window
  - Second condition for prima facie causality is then:

$$c \overset{\geq 20, \leq 40}{\underset{\geq p}{\rightsquigarrow}} e$$

- Found all structures
  - 100% of genuine causes found in low-noise datasets (i.e. prima facie causes, and not deemed spurious)
  - 92% in high-noise datasets

Pattern 4: Binary Tree

- Prior Work
  - Found D→I, E→H, F→K, G→J
  - Difficult to determine which was genuine cause
  - Had to disambiguate manually using prior knowledge about binary tree structure
- Using causality
  - Looking at average causal influence, actual parent was found as only genuine cause
  - Even though D and E (and F and G) have common cause, were able to distinguish their children



Future Applications

- Personalized Medicine
  - Patient data over long period of time
  - PatientsLikeMe
- Financial Data and Trading Rules
- Biological Data
  - Neuroscience
  - Cancer

Hume's Advice

- "Indulge your passion for science, says she, but let your science be human, and such as may have a direct reference to action and society.
- "Abstruse thought and profound researches I prohibit, and will severely punish, by the pensive melancholy which they introduce, by the endless uncertainty in which they involve you, and by the cold reception which your pretended discoveries shall meet with, when communicated.
- "**Be a philosopher; but, amidst all your philosophy, be still a man.**"