ANR CODDDE meeting

Hadrien Hours

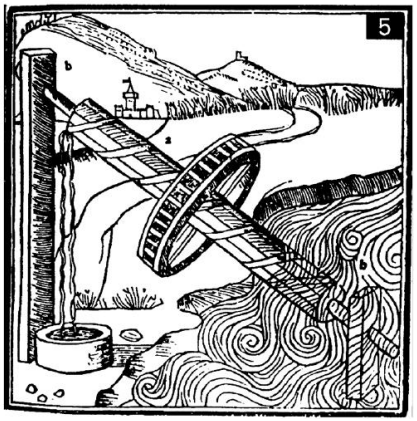
ENS Lyon,IXXI, DANTE team

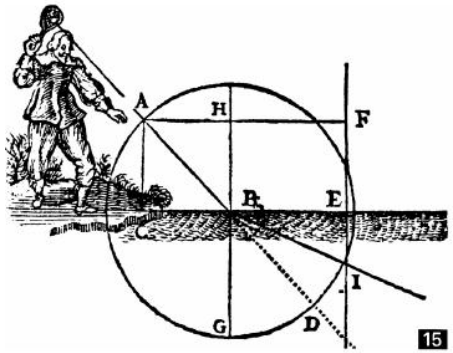
2015-01-27

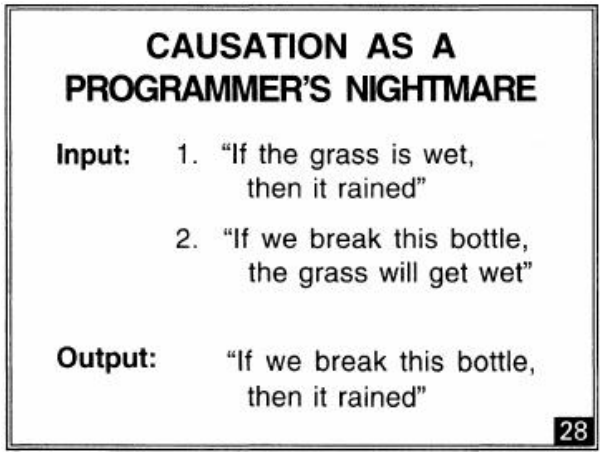
Introduction on Causality

Goals:

- Understand and model a system
- Predict behavior
- Formalize causal knowledge







Introduction on Causality

What is causality

- Set of causes and effects explaining a certain system behavior
- Historical
  - Aristotle, Galileo, Newton, David Hume, Pearson, Fisher, . . . ,
  - Judea Pearl (UCLA)
  - Peter Spirtes & Clark Glymour & Richard Scheines (CMU)

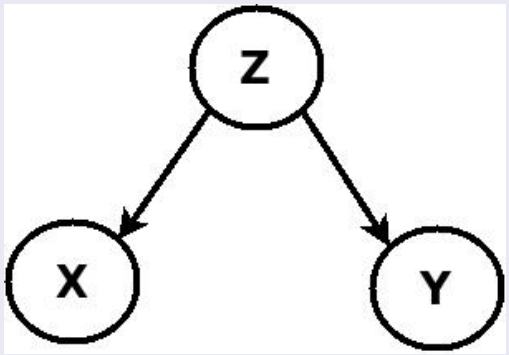
Why causality

- Complex systems: Interdependencies, Spurious associations
- Stable under interventions: Can predict the impact of change

Introduction on Causality

Correlation is not causation

Two parameters, X and Y, correlated



Spurious associations and latent variables

- People with yellow teeth higher probability to have lung cancer
- Windshield wiper and accidents

Causal model

Definitions

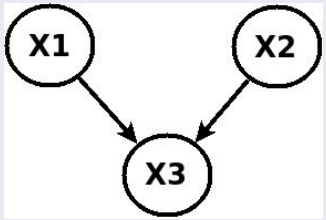
- *System:* $\{X_1, \ldots, X_p\}$
- *Causal study*: Understand the system causal dependencies:
  $\Rightarrow$ Causal model

Representations

- Structural equation models:

$$X_i := f_i(\prod_{j \neq i} X_j) + \epsilon_i$$

- Graphical representations



Causal model inference

Manual intervention

- Intervention
- Random experiments

Passive observation

- Suppose a model generating the observed system: hypothesis
- Infer such model through test relying on hypothesis
- *Constraints*: Determinism, time scale, model granularity

Ph.D. work

Telecommunication networks

- System:
  - Interconnected nodes providing connection and communications
- Observations:
  - Internet traffic (TCP) with probes within the network
- Complex system
  - No assumption regarding dependencies (linear), distributions (Normality)

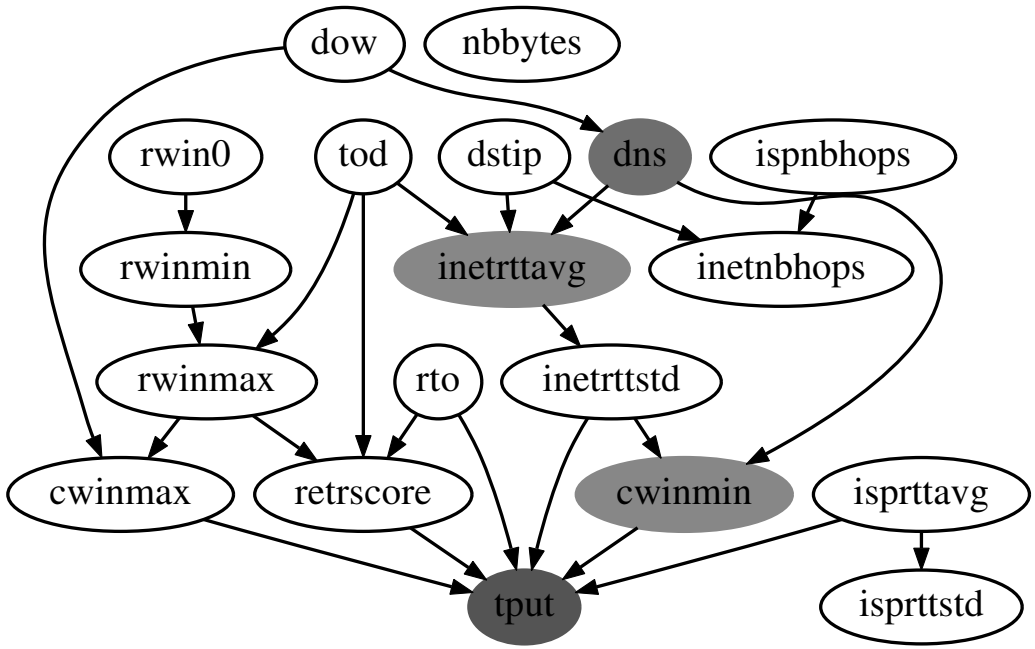
Graphical causal models

- Bayesian networks
- Directed Acyclic Graphs (DAGs)
- Graphical criteria to predict **intervention**

Ph.D. work: CDN

Goal: Understand the impact on CDN performance if choosing one DNS service instead of another

- Distribution throughput for users of DNS *a* if delay follows distribution of users of DNS *b*
- Distribution throughput for users of DNS *a* if server configurations of users of DNS *b*



Results Measure impact of DNS on CDN performance: redirection and server configuration

Study of the evolution of the mention network of Tweeter

Objective: Study the co-evolution of information diffusion and mention network

- If ***user A*** and ***user B*** share the same information: more likely to mention each other
- If ***user A*** and ***user B*** are exposed to the same information: more likely to mention each other

Approach

- Use follower network to follow information diffusion
- Use hashtags for information sharing

Twitter mention network

Twitter mention

- Dynamic interactions
- Classical approach: followers
- Different social interactions

Target model

- Structural properties
- Comparison
  - with follower network properties
  - information exposure / shared
- Use a causal approach to capture social impact (second step)

Causality as a tool to assess social bound dynamics

Goal: Impact of information diffusion on the dynamics of the mention network evolution

Why causality

- Many factors influencing social network evolution
- Many possible latent variables
- Time series (#JeSuisCharlie,...): Granger, Non deterministic dynamic systems

How

- Linguistic to capture social concepts (common / opposite interests)
- Capture of hashtags and communities as proxies

First results and open questions

Results

- Evolution of mention graph per day during on year.
- Mention network structure $\neq$ Follower network
  - Edge creation (reciprocal mention, time scale)
  - Triadic closure

Open questions

- Capture passive information exposition and active information exposition
  - Exposed information: Follower network
  - Shared information: hashtag clustering
- Capture ponctual events (Sport event, Political event)
- Introduce temporal decay in edge weights

References

- Causation, prediction and search, P. Spirtes, C. Glymour, R. Scheines, MIT Press, 2000
- Causality, J. Pearl, Cambridge University Press, 2009
- A causal approach to the study of TCP performance, H. Hours, E. Biersack, P. Loiseau, ACM TIST, 7-2, 2016
- A study of the impact of DNS resolvers on performance using a causal approach, H. Hours, E. Biersack, P. Loiseau, A. Finamore, M. Mellia, ITC 2015
- The directed closure processhybrid social-information networks, with an analysis of link formation on twitter, D.M. Romero and J. M. Kleinberg, ICWSM 2010
- COEVOLVE: A joint point process model for information diffusion and network co-evolution, M. Farajtabar, Y. Wang, M. Gomez-Rodriguez, S. Li, H. Zha, L. Song, CoRR, 2015
- The role of information diffusion in the evolution of social networks, L. Weng, J. Ratkiewicz, N. Perra, B. Goncalves, C. Castillo, F. Bonchi, R. Schifanella, F. Menczer, A. Flammini. KDD '13