Non-Stratified Chain Event Graphs: For Modelling Asymmetric Processes

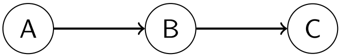
Aditi Shenvi

Savage Award (Theory and Methods), ISBA 2022

WARWICK

Probabilistic Graphical Models

PGM = Statistical Model of a Multivariate Process
+ Graph Representing Its Conditional
Independencies



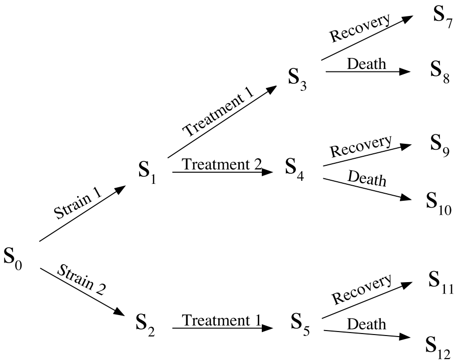
Benefits of PGMs:

- **Representational**: Useful communication tool.
- **Inferential**: Can leverage the factorised representation of the joint probability distribution for efficient inference.

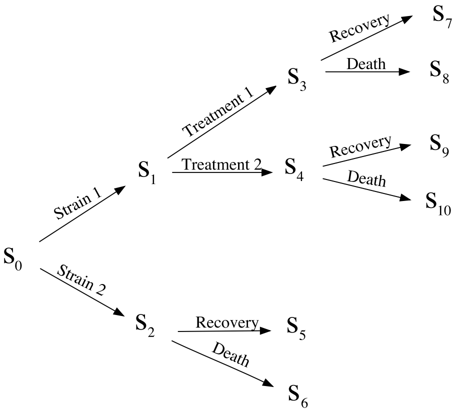
$$C \perp\!\!\!\perp A \mid B$$

$$p(a, b, c) = p(a)p(b|a)p(c|b)$$

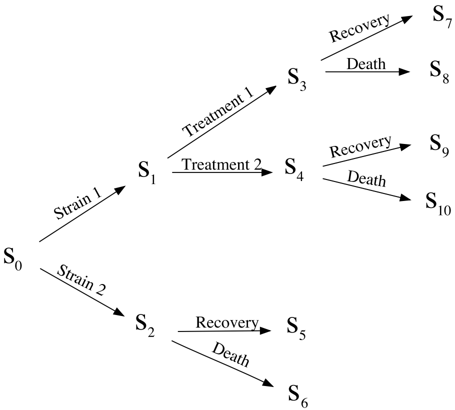
Asymmetric Processes



Asymmetric Processes



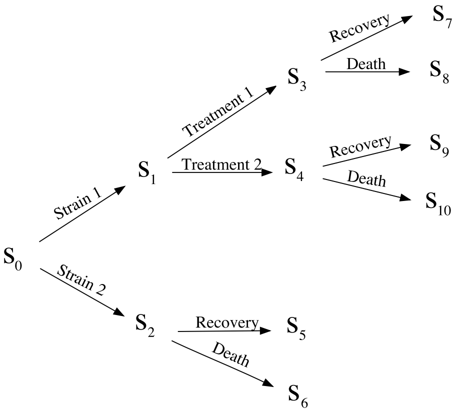
Asymmetric Processes



A process is **asymmetric** when:

- in simple words: when the tree of the process is not symmetric around its centre.

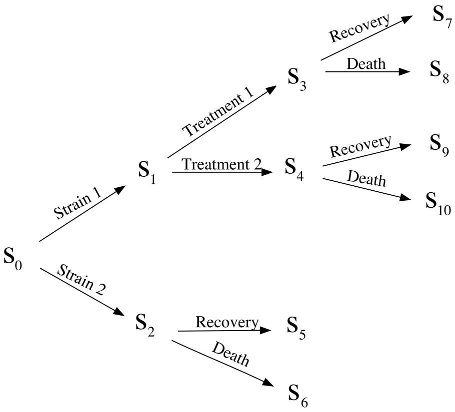
Asymmetric Processes



A process is **asymmetric** when:

- in simple words: when the tree of the process is not symmetric around its centre.
- more formally: when the event space of the process does not admit a natural product space structure.

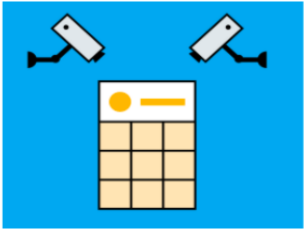
Asymmetric Processes

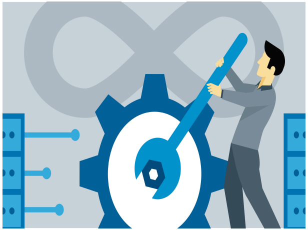


A process is **asymmetric** when:

- in simple words: when the tree of the process is not symmetric around its centre.
- more formally: when the event space of the process does not admit a natural product space structure.
- Here, $\mathbb{X}_S \times \mathbb{X}_T \times \mathbb{X}_O$

Asymmetric Processes





Outline and Goals

- Work with **probabilistic graphical models** for the representational and inferential benefits in modelling complex systems.
- However, these are typically **variable-based** and expect **structural symmetry**.

Aims:

1. What causes asymmetry and how do we model asymmetric processes?
2. How do we model longitudinal asymmetric processes with its different components evolving at different rates?
3. How can the above models be combined with other models, each describing a distinct part of a complex process?

PART I

Causes of Asymmetry

Asymmetric processes have different outcome spaces (or even an empty outcome space) given different outcomes of their ancestor variables. This happens because of:

- **Structural zeros**
  - Zero counts for a certain configuration of variables where a **non-zero value is logically restricted**.
  - e.g. Number of low-risk individuals who got treatment when treatment was only available to high-risk individuals.
- **Structural missing values**
  - Missing observations which have **no underlying meaningful value**.
  - e.g. Post-operative health status of individuals who had the illness but were not operated.

Bayesian Networks

"BNs are directed acyclic graphs in which nodes represent variables and edges represent informational or causal dependencies."
– Judea Pearl

- Structural Zeros: Violates positivity assumption. Hidden in the conditional probability tables.
- Structural Missing Values: Cannot be meaningfully accommodated.

$$X_S \rightarrow X_T \rightarrow X_O$$

STRAIN 1	STRAIN 2
0.2	0.8

STRAIN	TREATMENT 1	TREATMENT 2
1	0.4	0.6
2	0.15	0.85

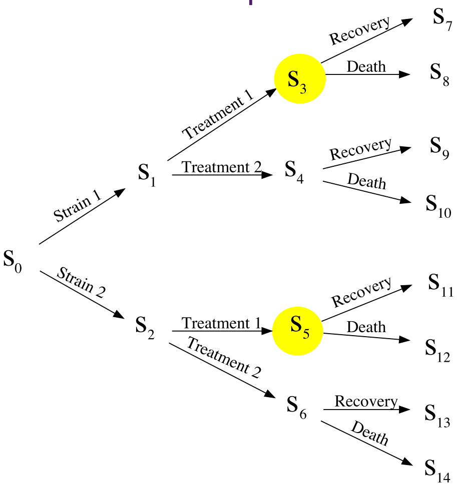
TREATMENT	OUTCOME Rec	OUTCOME Death
1	0.95	0.05
2	0.85	0.15

Chain Event Graphs

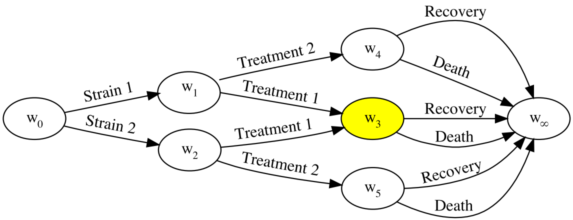
- Introduced in Smith and Anderson (2008)[1]
- Subject information often comes in structural form $\implies$ trees.
- Constructed from **event trees**.
- **Leverage local symmetries** to provide a compact representation.
- Generalise discrete BNs as a special class.

[1] Smith, J. Q., & Anderson, P. E. (2008). Conditional independence and chain event graphs. Artificial Intelligence, 172(1), 42–68.

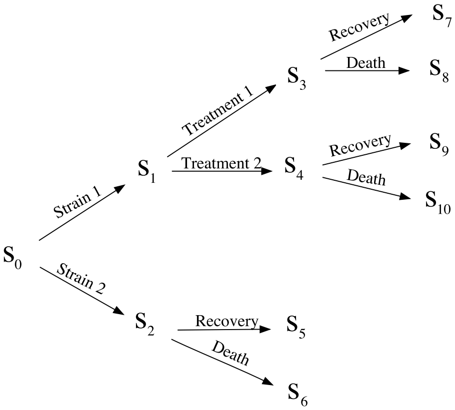
Chain Event Graphs

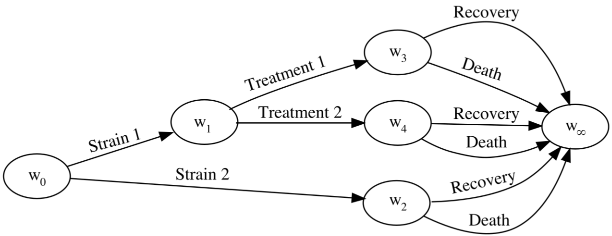


- Vertices s_i and s_j in same **'stage'** and assigned same colour iff $\boldsymbol{\theta}_{s_i} = \boldsymbol{\theta}_{s_j}$.
- Vertices are contracted when their rooted subtrees are isomorphic.
- **Stratified class:** CEGs for symmetric processes (hiding some technical details). Have analogous BN representations.

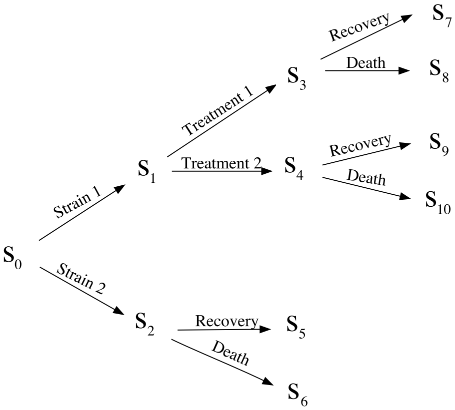


Chain Event Graphs



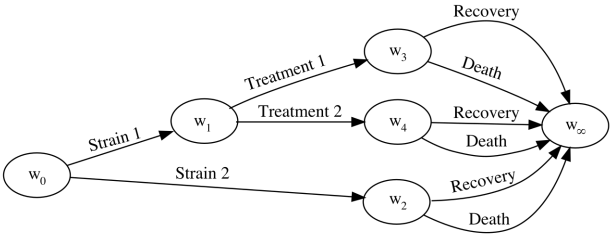


Chain Event Graphs



Non-Stratified CEG Class

- Strongly differentiates CEGs from BNs!
- My thesis was a first systematic theoretical exposition of this class.



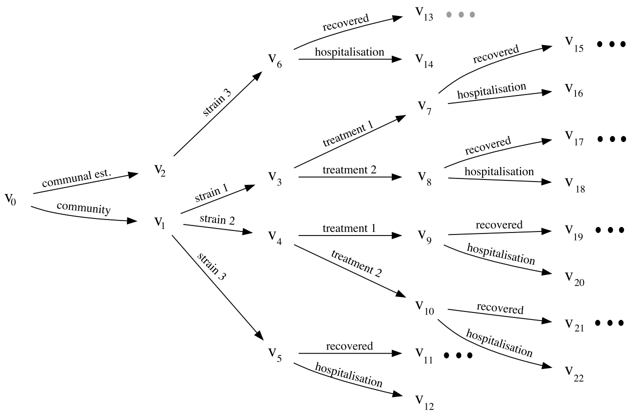
PART II

Longitudinal Asymmetric Processes

- To explicitly describe **temporal effects** in complex dynamic systems which are asymmetric and evolve in **continuous time**.
- Consider cold, flu or covid-19?
  - Overlap of symptoms
  - Different timeline of presentation
- Interested in not only **what happens next** but also **when it happens**.
- Alternative models:
  - Continuous time BNs
  - Graphical event history models

[1] Image from "Flu and COVID-19: how to tell the difference this winter and stay safe, UAB News"

Continuous Time Dynamic CEGs

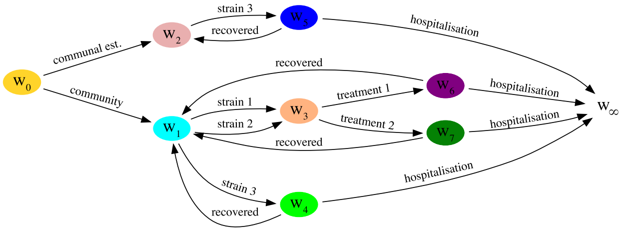


- Constructed from infinite event trees.
- Each edge e has conditional holding time variable $H(e)$.
- Vertices s_i and s_j in same **stage** iff $\boldsymbol{\theta}_{s_i} = \boldsymbol{\theta}_{s_j}$ and **for mapped edges,** $H(e)$ **&** $H(e')$ **follow same distribution**.
- Vertices contracted when their rooted infinite subtrees are isomorphic.
- **Time-invariant covariates**: Incorporated with iid holding times which have no discriminatory power.

CT-DCEG Special Class

Definition (Finite and Regular CT-DCEG)

Say that a CT-DCEG $\mathcal{D}$ is finite if its vertex set is finite, and regular if its conditional transition parameters in $\boldsymbol{\Phi}_{\mathcal{D}}$ and its conditional holding time random variables in $\boldsymbol{\mathcal{H}}_{\mathcal{D}}$ are all time homogeneous



Semi-Markovian Connection

Definition (Semi-Markov Process)

Consider a stochastic process $\mathbf{Z} = \{Z_t, t \geq 0\}$ on a discrete state space $\mathbf{S}$. The state occupied at the nth transition is given by $\mathbf{X} = (X_n)_{n \in \mathbb{N}}$, the jump times by $\mathbf{T} = (T_n)_{n \in \mathbb{N}}$ and the holding time in X_n before moving to X_{n+1} by $\boldsymbol{\tau} = (\tau_n)_{n \in \mathbb{N}}$ where $\tau_n = T_n - T_{n-1}$. The process $\mathbf{Z}$ is an SMP when

$$p(X_{n+1} = j, \tau_{n+1} \leq t \mid X_n, \ldots, X_0; \tau_n, \ldots, \tau_1) = p(X_{n+1} = j, \tau_{n+1} \leq t \mid X_n = i),$$

Semi-Markovian Connection

Definition (Semi-Markov Process)

Consider a stochastic process $\mathbf{Z} = \{Z_t, t \geq 0\}$ on a discrete state space $\mathbf{S}$. The state occupied at the nth transition is given by $\mathbf{X} = (X_n)_{n \in \mathbb{N}}$, the jump times by $\mathbf{T} = (T_n)_{n \in \mathbb{N}}$ and the holding time in X_n before moving to X_{n+1} by $\boldsymbol{\tau} = (\tau_n)_{n \in \mathbb{N}}$ where $\tau_n = T_n - T_{n-1}$. The process $\mathbf{Z}$ is an SMP when

$$p(X_{n+1} = j, \tau_{n+1} \leq t \mid X_n, \ldots, X_0; \tau_n, \ldots, \tau_1) = p(X_{n+1} = j, \tau_{n+1} \leq t \mid X_n = i),$$

- Every finite CT-DCEG has an associated (**approximate**) SMP representation.

Semi-Markovian Connection

Definition (Semi-Markov Process)

Consider a stochastic process $\mathbf{Z} = \{Z_t, t \geq 0\}$ on a discrete state space $\mathbf{S}$. The state occupied at the nth transition is given by $\mathbf{X} = (X_n)_{n\in\mathbb{N}}$, the jump times by $\mathbf{T} = (T_n)_{n\in\mathbb{N}}$ and the holding time in X_n before moving to X_{n+1} by $\boldsymbol{\tau} = (\tau_n)_{n\in\mathbb{N}}$ where $\tau_n = T_n - T_{n-1}$. The process $\mathbf{Z}$ is an SMP when

$$p(X_{n+1} = j, \tau_{n+1} \leq t \,|\, X_n, \ldots, X_0; \tau_n, \ldots, \tau_1) = p(X_{n+1} = j, \tau_{n+1} \leq t \,|\, X_n = i),$$

- Every finite CT-DCEG has an associated (**approximate**) SMP representation.
- An SMP is completely defined by its initial distribution and its renewal kernel:

$$p_{ij} = \begin{cases} \theta^*_{ij}, & \text{if } \mathbb{1}_{\mathcal{D}}(v_i \to v_j) = 1 \\ 0, & \text{otherwise} \end{cases} \qquad \text{(transition matrix entries)}$$

$$F_{ij}(t) = \mathbb{P}(\tau_{ij} \leq t \,|\, X_{n+1} = v_j, X_n = v_i) \qquad \text{(cumulative cond. holding time dist.)}$$

$$Q_{ij}(t) = p_{ij} F_{ij}(t) \qquad \text{(renewal kernel)}$$

Why not just use a semi-Markov process?

- States for an SMP are often defined *a priori* rather than being identified through a mixture of elicitation and learning.

Why not just use a semi-Markov process?

- States for an SMP are often defined *a priori* rather than being identified through a mixture of elicitation and learning.
- An SMP is a simple graph (loops permitted), unlike a CT-DCEG which can be a multigraph – so some information is condensed.

Why not just use a semi-Markov process?

- States for an SMP are often defined *a priori* rather than being identified through a mixture of elicitation and learning.
- An SMP is a simple graph (loops permitted), unlike a CT-DCEG which can be a multigraph – so some information is condensed.
- Colouring of CT-DCEG gives information about the conditional independencies and local symmetries.

Why not just use a semi-Markov process?

- States for an SMP are often defined *a priori* rather than being identified through a mixture of elicitation and learning.
- An SMP is a simple graph (loops permitted), unlike a CT-DCEG which can be a multigraph – so some information is condensed.
- Colouring of CT-DCEG gives information about the conditional independencies and local symmetries.
- HOWEVER, the SMP representation of a CT-DCEG is useful.

Probability Propagation in CT-DCEGs

- Revising probabilities in light of evidence.

[2] The specifications of acceptable evidence requires some care, which this presentation will overlook.

Probability Propagation in CT-DCEGs

- Revising probabilities in light of evidence.
- Evidence may refer to vertices or edges visited (evolutionary evidence $\mathscr{E}$) or related to transition/holding times (temporal evidence $\mathscr{T}$).

[2] The specifications of acceptable evidence requires some care, which this presentation will overlook.

Probability Propagation in CT-DCEGs

- Revising probabilities in light of evidence.
- Evidence may refer to vertices or edges visited (evolutionary evidence $\mathscr{E}$) or related to transition/holding times (temporal evidence $\mathscr{T}$).
- Evidence might be certain or uncertain.[2]

[2] The specifications of acceptable evidence requires some care, which this presentation will overlook.

Probability Propagation in CT-DCEGs

- Revising probabilities in light of evidence.
- Evidence may refer to vertices or edges visited (evolutionary evidence $\mathscr{E}$) or related to transition/holding times (temporal evidence $\mathscr{T}$).
- Evidence might be certain or uncertain.[2]
- Typically $\mathscr{E}$ and $\mathscr{T}$ are concerned with some finite window $\mathcal{C}$ of CT-DCEG $\mathcal{D}$.

[2] The specifications of acceptable evidence requires some care, which this presentation will overlook.

Probability Propagation in CT-DCEGs

- Revising probabilities in light of evidence.
- Evidence may refer to vertices or edges visited (evolutionary evidence $\mathscr{E}$) or related to transition/holding times (temporal evidence $\mathscr{T}$).
- Evidence might be certain or uncertain.[3]
- Typically $\mathscr{E}$ and $\mathscr{T}$ are concerned with some finite window $\mathcal{C}$ of CT-DCEG $\mathcal{D}$.

Compute $P(Q|\mathscr{E}, \mathscr{T}, \mathcal{C})$ for query Q using **local computations**.

[3] The specifications of acceptable evidence requires some care, which this presentation will overlook.

Propagation Scheme for the Finite Window of Interest

- Finite window of interest $\mathcal{C}$ is essentially a large CEG evolving in continuous time.

[1] Thwaites, P. A., Smith, J. Q., & Cowell, R. G. (2008). Propagation using chain event graphs, UAI.

Propagation Scheme for the Finite Window of Interest

- Finite window of interest $\mathcal{C}$ is essentially a large CEG evolving in continuous time.
- Observing evidence typically **reduces** the complexity of a CEG.

[1] Thwaites, P. A., Smith, J. Q., & Cowell, R. G. (2008). Propagation using chain event graphs, UAI.

Propagation Scheme for the Finite Window of Interest

- Finite window of interest $\mathcal{C}$ is essentially a large CEG evolving in continuous time.
- Observing evidence typically **reduces** the complexity of a CEG.
- Thwaites et al (2008)[1] proposed an evidence propagation algorithm for evolutionary evidence in CEGs.

[1] Thwaites, P. A., Smith, J. Q., & Cowell, R. G. (2008). Propagation using chain event graphs, UAI.

Propagation Scheme for the Finite Window of Interest

- Finite window of interest $\mathcal{C}$ is essentially a large CEG evolving in continuous time.
- Observing evidence typically **reduces** the complexity of a CEG.
- Thwaites et al (2008)[1] proposed an evidence propagation algorithm for evolutionary evidence in CEGs.
- We extended this to incorporate point temporal evidence of unconditional holding times at a given vertex.

[1] Thwaites, P. A., Smith, J. Q., & Cowell, R. G. (2008). Propagation using chain event graphs, UAI.

Propagation Scheme for the Finite Window of Interest

- Finite window of interest is essentially a large CEG evolving in continuous time.
- Observing evidence typically **reduces** the complexity of a CEG.
- Thwaites et al (2008)[1] proposed an evidence propagation algorithm for evolutionary evidence in CEGs.
- We extended this to incorporate point temporal evidence of unconditional holding times at a given vertex.

Path A – Avg 20 mins
Path B – Avg 2 hours
Path C – Avg 10 hours

[1] Thwaites, P. A., Smith, J. Q., & Cowell, R. G. (2008). Propagation using chain event graphs, UAI.

Vanilla CEG Propagation

Calculate messages:
Set $\Phi(w_\infty) = 1$.
For each edge $e = (w, w', l)$, calculate:

$$\tau_e(w' \mid w) = p(w, w', l) \tag{1}$$

For each vertex w, calculate:

$$\Phi(w) = \textstyle\sum_{e \in E(w)} \tau_e(w' \mid w). \tag{2}$$

Update:

$$\hat{p}(e) = \begin{cases} \dfrac{\tau_e(w' \mid w)}{\Phi(w)}, & \text{if } e \in \text{simplified graph} \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

Temporal CEG Propagation

Calculate messages:

Set $\Phi(w_\infty) = 1$.

For each edge $e = (w, w', l)$, calculate:

$$\tau_e(w' \mid w) = p(w, w', l) \tag{1}$$

$$\tau_e^{t_w}(w' \mid w) = p(H_{(w,w',l)} = t_w). \tag{2}$$

For each vertex w, calculate:

$$\Phi(w) = \sum_{e \in E(w)} \tau_e(w' \mid w), \tag{3}$$

$$\Phi^{t_w}(w) = \sum_{e \in E(w)} \tau_e(w' \mid w)\, \tau_e^{t_w}(w' \mid w). \tag{4}$$

Update:

$$\hat{p}(e) = \begin{cases} \dfrac{\tau_e(w' \mid w)\, \tau_e^{t_w}(w' \mid w)}{\Phi^{t_w}(w)}, & \text{if } e \in \text{simplified graph} \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

Dynamic CT-DCEG Propagation Scheme

Based on Kjærulff's strategy[1] for dynamic Bayesian networks:

- Unroll the graph over its time-slices;
- Separate into three submodels: Past, Current and Future;
- Propagation for each as appropriate.

[1] Kjærulff, U. (1992). A computational scheme for reasoning in dynamic probabilistic networks, UAI.

Dynamic CT-DCEG Propagation Scheme

Based on Kjærulff's strategy[1] for dynamic Bayesian networks:

- Unroll the graph over its time-slices;
- Separate into three submodels: Past, Current and Future;
- Propagation for each as appropriate.

Challenges:

- No natural concept analogous to time-slices in continuous time.

[1] Kjærulff, U. (1992). A computational scheme for reasoning in dynamic probabilistic networks, UAI.

Dynamic CT-DCEG Propagation Scheme

Based on Kjærulff's strategy[1] for dynamic Bayesian networks:

- Unroll the graph over its time-slices;
- Separate into three submodels: Past, Current and Future;
- Propagation for each as appropriate.

Challenges:

- No natural concept analogous to time-slices in continuous time.
- Not appropriate to base on stopping times as **path may evolve at a different rates!**

[1] Kjærulff, U. (1992). A computational scheme for reasoning in dynamic probabilistic networks, UAI.

Dynamic CT-DCEG Propagation Scheme

Based on Kjærulff's strategy[1] for dynamic Bayesian networks:

- Unroll the graph over its time-slices;
- Separate into three submodels: Past, Current and Future;
- Propagation for each as appropriate.

Challenges:

- No natural concept analogous to time-slices in continuous time.
- Not appropriate to base on stopping times as **path may evolve at a different rates!**
- New concept: **Passage-slices**

[1] Kjærulff, U. (1992). A computational scheme for reasoning in dynamic probabilistic networks, UAI.

Dynamic CT-DCEG Propagation Scheme

Based on Kjærulff's strategy[1] for dynamic Bayesian networks:

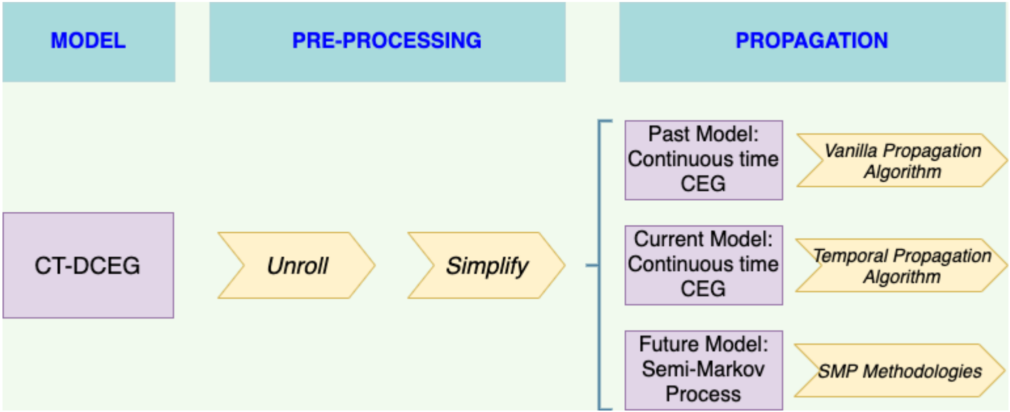
- Unroll the graph over its time-slices;
- Separate into three submodels: Past, Current and Future;
- Propagation for each as appropriate.

Challenges:

- No natural concept analogous to time-slices in continuous time.
- Not appropriate to base on stopping times as **path may evolve at a different rates!**
- New concept: **Passage-slices**
- Defined directly from the graph of the CT-DCEG without reference to the underlying holding time distributions.

[1] Kjærulff, U. (1992). A computational scheme for reasoning in dynamic probabilistic networks, UAI.

Dynamic CT-DCEG Propagation Scheme



Contributions for Part I

- Short-term modelling of Asymmetric Processes:
  - Developed the **non-stratified** class of Chain Event Graphs for asymmetric processes.
  - They **explicitly** accommodate these asymmetries **within the statistical model and the graph**.
  - Presented a **general construction algorithm with an optimal stopping criterion** for transforming a coloured tree to CEG and showed this transformation is bijective.
  - Demonstrated through the falls intervention example that they are **more suitable to using a Bayesian network** for asymmetric processes.
  - Developed a Python package: `cegpy`[1] for modelling with this class (version 1 to be released next month!)

[1] Joint work with Gareth Walley, Peter Strong and Kasia Kobalczyk

Contributions for Part II

- Modelling of Longitudinal Asymmetric Processes evolving in Continuous Time:
  - Developed a **dynamic, continuous-time extension** of the non-stratified CEG class with bespoke semantics.
  - Developed a **dynamic propagation scheme** and extended the propagation algorithm to **incorporate temporal evidence**.
  - Explored **model selection** for finite and regular CT-DCEGs using an object-oriented approach.

Summary and Contributions for Part III

- CEGs as part of a composite model for modelling criminal collaborations[1]
  - Developed a **dynamic weighted network** to model the intensity of communications between suspects.
  - Modelled a **collection of CT-DCEGs** – one for each suspect's personal criminal trajectory.
  - Combined these into an **integrated decision support system** using a decoupling methodology borrowed from multiregression dynamic models.
  - Demonstrated how these can be used by experts to customise **indicators** of the imminence of an attack.

More on this at Jim Smith's talk at the next session "Contributions to Bayesian Modelling 1" in room Bonaventure!

[1] Joint work with Oliver Bunnin and Jim Smith. Supported by the Alan Turing Institute's Defence and Security Programme.

Summary

- Asymmetric processes exist in many domains, naturally or by design.
- The non-stratified class of Chain Event Graphs has developed for modelling (longitudinal) asymmetric processes.
- These can be used in combination with other models as part of a composite model.
- Collaborators: Jim Q. Smith (Warwick), Oliver Bunnin (Formerly at Turing, now at NatWest), Robert Walton (Barts Medical School), Sandra Eldridge (QMUL)
- Key References:
  - Shenvi et al (2018). Modelling with non-stratified chain event graphs, In International Conference on Bayesian Statistics in Action. Springer.
  - Shenvi, Bunnin & Smith (2022+). A Bayesian Decision Support System for Counteracting Activities of Terrorist Groups. arXiv:2007.04410 (Accepted subject to revisions RSSA)
  - Shenvi & Smith (2020). Propagation for dynamic continuous time chain event graphs. arXiv preprint arXiv:2006.15865.
- Software: cegpy (https://github.com/g-walley/cegpy) – Collaborators: Gareth Walley, Peter Strong and Kasia Kobalczyk

Thank You!

Contact: Aditi.Shenvi@warwick.ac.uk
Website: https://ashenvi10.github.io/

References

- Bunnin, F. O., Shenvi, A., & Smith, J. Q. (2022+). A Bayesian Decision Support System for Counteracting Activities of Terrorist Groups. arXiv:2007.04410 (Accepted subject to revisions RSS: Series A)
- Shenvi, A., & Smith, J. Q. (2020). Propagation for dynamic continuous time chain event graphs. arXiv preprint arXiv:2006.15865.
- Shenvi, A., & Smith, J. Q. (2020). Constructing a chain event graph from a staged tree, In Proceedings of the Tenth International Conference on Probabilistic Graphical Models. PMLR.
- Shenvi, A., & Smith, J. Q. (2019). A Bayesian dynamic graphical model for recurrent events in public health. arXiv preprint arXiv:1811.08872.
- Shenvi, A., Smith, J. Q., Walton, R., & Eldridge, S. (2018). Modelling with non-stratified chain event graphs, In International Conference on Bayesian Statistics in Action. Springer.
- Smith, J. Q., & Shenvi, A. (2018). Assault crime dynamic chain event graphs. Working Paper. http://wrap.warwick.ac.uk/104824/

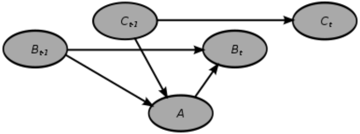
EXTRA SLIDES

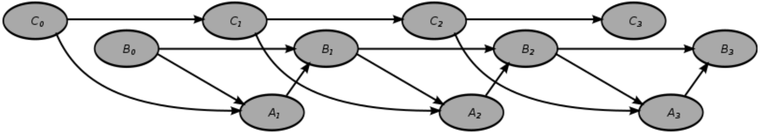
Main Challenges

- The CEG can get too large for moderate to large number of variables posing a representational challenge.
  - **Possible Solution:** Object oriented approach to model construction.
- The edge counts on an event tree can get too sparse for very large trees.
  - **Possible Solution:** Several ways to handle this for model selection. Useful to have more prior information to restrict the model space or use an object oriented approach.

Unrolling

- Need concept analogous to time-slices.

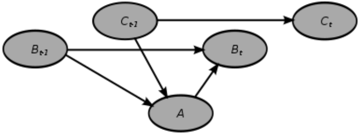


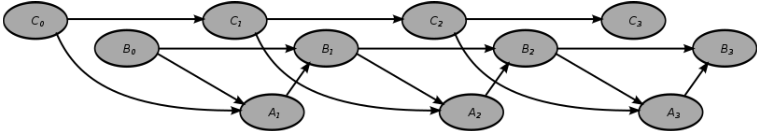


[1] Images from https://en.wikipedia.org/wiki/Dynamic_Bayesian_network

Unrolling

- Need concept analogous to time-slices.
- Not appropriate to base on stopping times as **path may evolve at a different rates!**

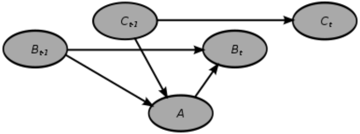


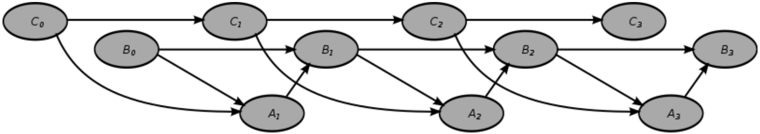


[1] Images from `https://en.wikipedia.org/wiki/Dynamic_Bayesian_network`

Unrolling

- Need concept analogous to time-slices.
- Not appropriate to base on stopping times as **path may evolve at a different rates!**
- New concept: **Passage-slices**

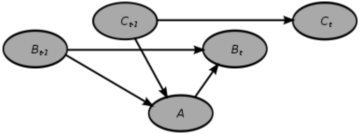


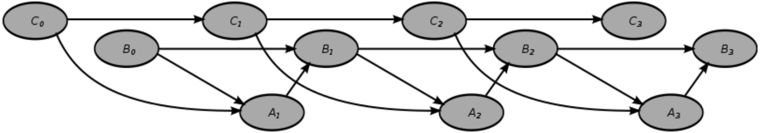


[1] Images from `https://en.wikipedia.org/wiki/Dynamic_Bayesian_network`

Unrolling

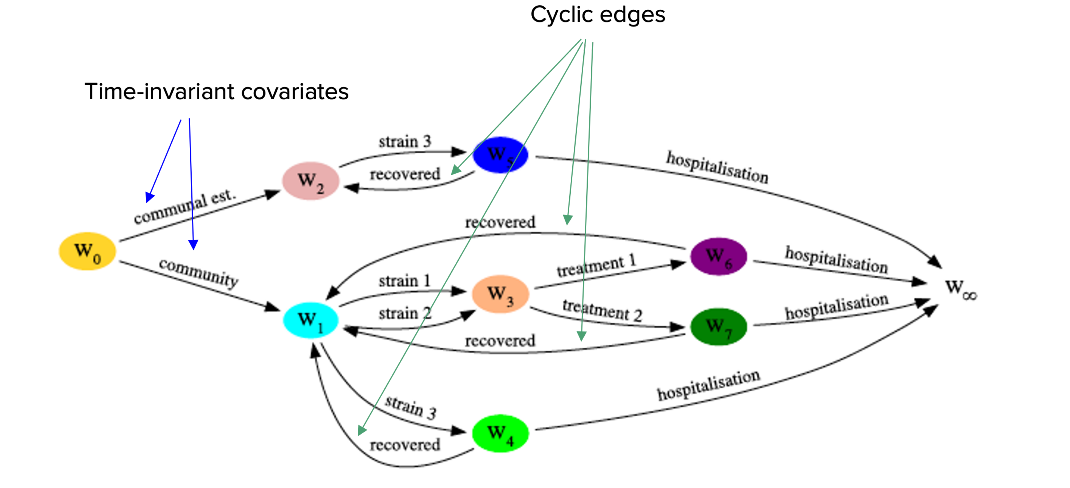
- Need concept analogous to time-slices.
- Not appropriate to base on stopping times as **path may evolve at a different rates!**
- New concept: **Passage-slices**
- Defined directly from the graph of the CT-DCEG without reference to the underlying holding time distributions.

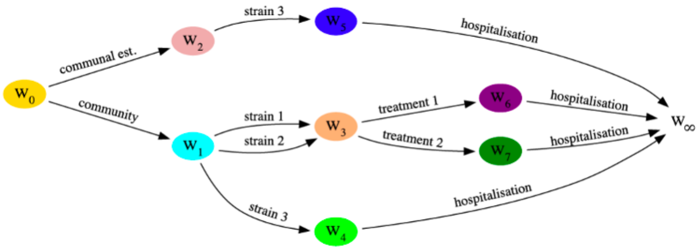




[1] Images from https://en.wikipedia.org/wiki/Dynamic_Bayesian_network

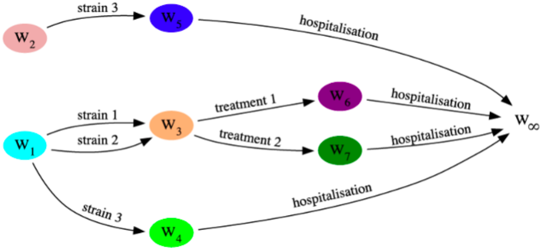
Infection Example





For first passage-slice:

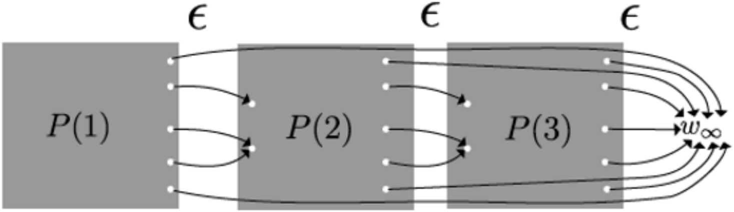
- Remove cyclic edges



For subsequent passage-slices:

- Begin where cyclic edge terminated
- Remove cyclic edges

Unrolling



Infection Process Unrolled