

Découverte causale sur de larges séries temporelles

Application à des séquences d'images d'écosystème tumoral

Franck Simon¹, Louise Dupuis¹, Vincent Cabeli¹, Tiziana Tocci¹, Nikita Lagrange¹, Hervé Isambert^{1*}

¹ CNRS UMR168, Institut Curie, Université PSL, Sorbonne Université, Paris, France

* corresponding authors: herve.isambert@curie.fr

Nous présentons ici la version temporelle de la méthode de découverte causale, MIIC^{1,2}, qui apprend des réseaux causaux pour un large éventail de données mixtes et incluant des variables latentes. MIIC a été appliqué à des données biologiques ou biomédicales, comme des données transcriptomiques et génomiques¹ ou des données de dossiers médicaux de patients². Les séries temporelles, comme des images vidéo de cellules vivantes, contiennent cependant des informations sur la dynamique des variables d'intérêt, qui peut en principe faciliter la découverte de processus fonctionnels de cause à effet, en partant de l'hypothèse que les événements futurs ne peuvent pas causer des événements passés.

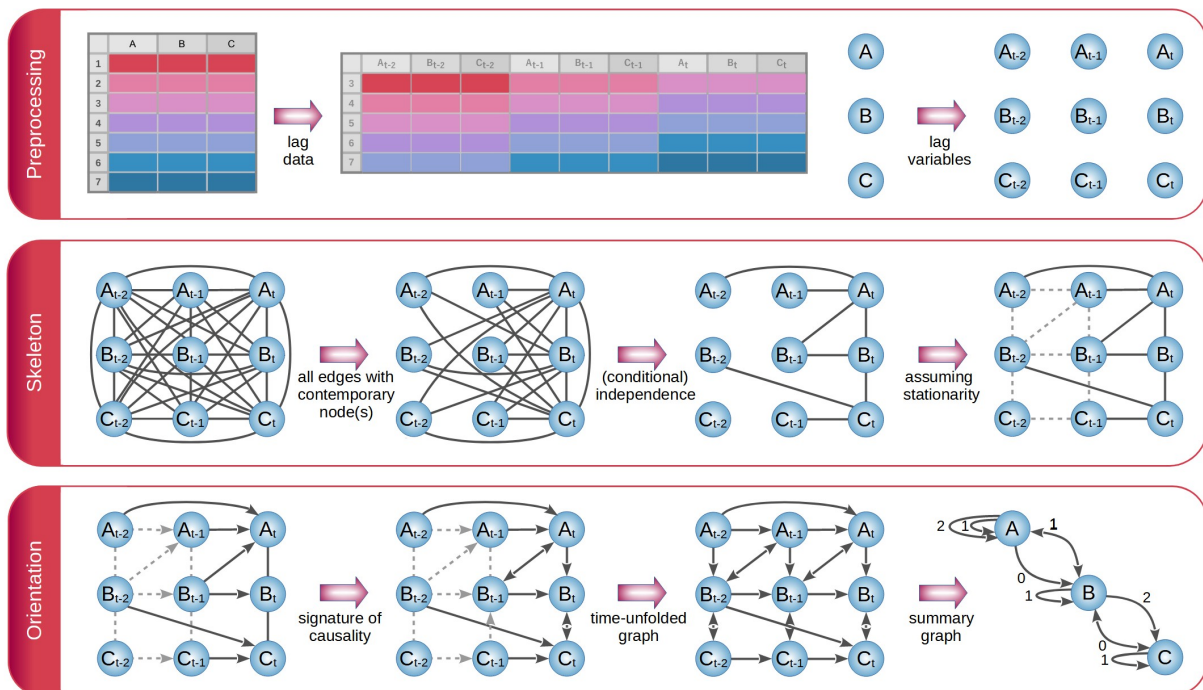


Figure 1. Schéma algorithmique de la version temporelle de MIIC^{1,2} (Simon et al, en préparation).

Afin d'analyser des ensembles de données de séries chronologiques, la version temporelle de MIIC vise à apprendre un graphe déplié dans le temps, où chaque variable est représentée par une série de nœuds associés à sa valeur à différents moments relatifs, **Figure 1**. Un tel réseau déplié dans le temps est nécessaire pour tenir compte de la corrélation temporelle entre des échantillons successifs dans les données de séries chronologiques. En admettant que la dynamique peut être considérée comme stationnaire, le graphe recherché doit être invariant par translation dans le temps et une structure périodique peut lui être assignée *a priori*. De plus, la découverte causale peut être limitée à quelques pas de temps à partir du temps d'exécution, t , jusqu'à un maximum de $t - \tau_{\max}$, pas de temps passés.

La version temporelle de MIIC a été utilisée pour effectuer l'analyse de séquences d'images d'un écosystème tumoral reconstitué *ex vivo* à l'aide de la technologie tumor-on-chip. Ces séquences proviennent d'une étude de preuve de concept³ qui a déterminé les effets d'un médicament anticancéreux (les anticorps monoclonaux trastuzumab, nom de marque Herceptin, utilisés pour traiter des cancers du sein HER2+) sur un micro-environnement tumoral reconstitué comprenant des cellules cancéreuses, des cellules immunitaires, des fibroblastes associés au cancer (CAF) et des cellules endothéliales.

Les caractéristiques cellulaires telles que la géométrie cellulaire, la vitesse, la division, l'apoptose (*i.e.* mort cellulaire), les interactions transitoires et les contacts persistants entre cellules, ont été extraites des images brutes avec l'information de temps avec l'algorithme de segmentation et de tracking CellHunter³. Ces données ont été soumises en entrée de la version temporelle de MIIC pour inférer un réseau causal.

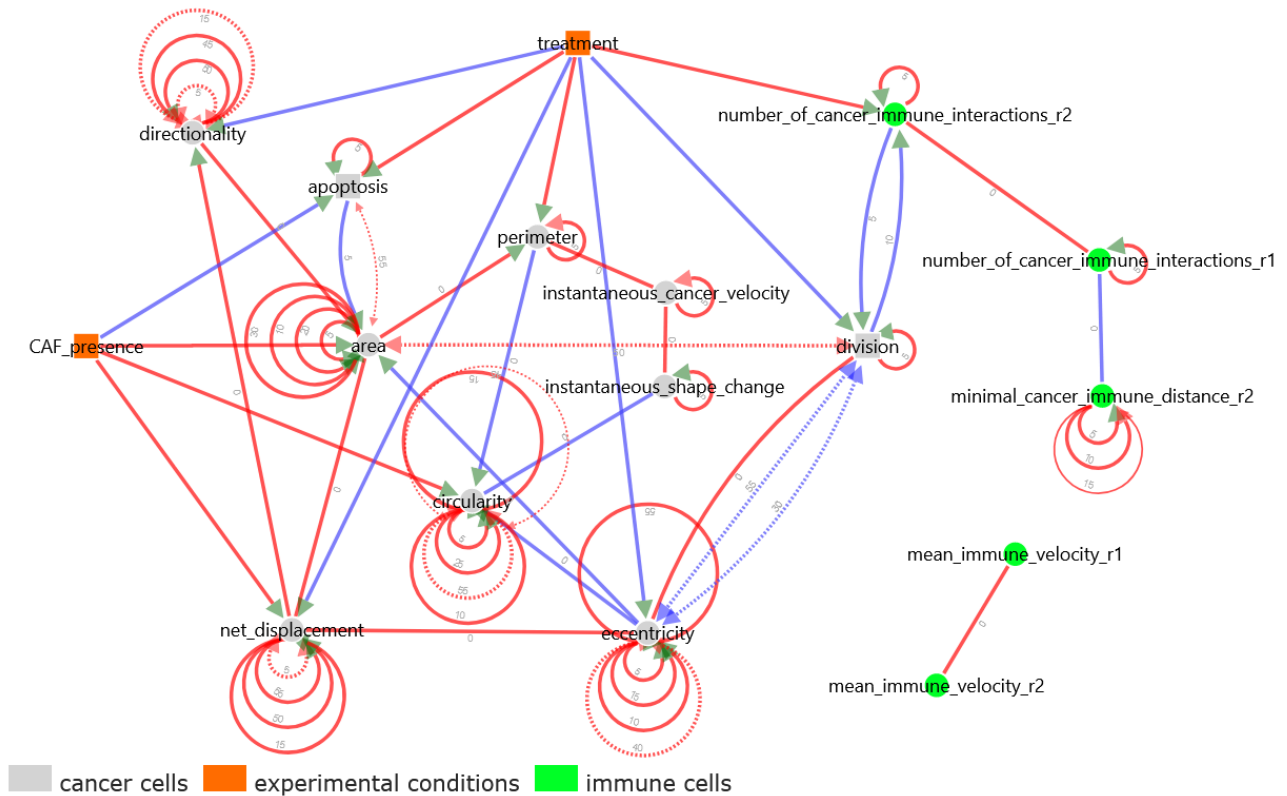


Figure 2. Réseau causal résumé obtenu à partir de données temporelles (50,000 images, dt=2 min) d'écosystème tumoral incluant des cellules cancéreuses et immunitaires dans 4 conditions: avec ou sans traitement et avec ou sans CAFs (Simon *et al*, en prép).

Le réseau causal temporel résumé obtenu, **Figure 2**, révèle de nouvelles découvertes biologiquement pertinentes en plus de confirmer les résultats connus d'études antérieures. En particulier, MIIC temporel découvre que les CAFs inhibent directement l'apoptose des cellules cancéreuses, indépendamment du traitement anticancéreux. MIIC temporel découvre également que le traitement réduit le périmètre des cellules cancéreuses et inhibe leur migration, ce qui n'a pas été rapporté jusqu'à présent non plus. De plus, MIIC temporel confirme les résultats connus d'études antérieures, il retrouve notamment que le traitement augmente l'apoptose des cellules cancéreuses et le nombre d'interactions entre cellules cancéreuses et immunitaires, et diminue le taux de division des cellules cancéreuses. De même, MIIC temporel retrouve l'effet que les CAFs stimulent la migration des cellules cancéreuses et augmentent leur aire et circularité.

Fait intéressant, MIIC temporel identifie également des effets multiples et éventuellement antagonistes avec des délais différents. Par exemple, MIIC temporel retrouve plusieurs relations antagonistes connues entre des caractéristiques morphodynamiques telles que la division cellulaire et l'excentricité. En effet, les phases tardives de la division cellulaire sont associées à une augmentation marquée de l'excentricité (arête rouge) mais la division cellulaire est précédée d'une nette diminution de l'excentricité, une à deux heures avant la cytokinèse (arêtes bleues), une fois que la décision de division a été prise (correspondant vraisemblablement à la cause latente) et que la cellule duplique en fait son matériel biologique (prophase).

Pour résumer, MIIC temporel permet de découvrir de nouvelles relations causales entre des caractéristiques cellulaires observées ou non et possiblement décalées dans le temps. Avec la disponibilité de données temporelles virtuellement illimitées dans de nombreux domaines, les méthodes de découverte causale sans hypothèse deviennent indispensables et nous pensons que MIIC temporel peut améliorer l'interprétation de ces données avec l'apport de la causalité.

Références

1. Verny L, Sella N, Affeldt S, Singh PP, Isambert H: Learning causal networks with latent variables from multivariate information in genomic data. *PLoS Comput Biol* **13**(10):e1005662 (2017).
2. Cabeli V, Verny L, Sella N, Uguzzoni U, Verny M, Isambert H: Learning clinical networks from medical records based on information estimates in mixed-type data. *PLoS Comput Biol* **16**(5):e1007866 (2020).
3. Nguyen, M. *et al*. Dissecting Effects of Anti-cancer Drugs and Cancer-Associated Fibroblasts by On-Chip Reconstitution of Immunocompetent Tumor Microenvironments. *Cell Reports* **25**, 3884–3893.e3 (2018).