Gaining confidence in inferred networks

Leo P.M. Diaz[1], Thalia E. Chan[2] and Michael P.H. Stumpf[1,2]

[1] Melbourne Integrative Genomics, School of Mathematics and Statistics, University of Melbourne, Parkville VIC 3010, Australia
[2] Centre for Integrative Systems Biology and Bioinformatics, Department of Life Sciences, Imperial College London, London SW7 2AZ, UK

1. Uncertainty in models

Discrepancies between models (**Figure 1**) demonstrate the extent of **structural uncertainty** and call for reliable validation procedures.

Different algorithms produce very different models – how to reliably choose one?

2. Mixing patterns

A reasonable **assumption** is to expect mixing patterns: genes involved in related biological processes preferentially interacting together.

What properties can we expect biological networks to display?

3. Assortativity

The assortativity coefficient is a **heuristic** measuring the presence of mixing patterns.

How to detect and measure mixing patterns?

Background

Reverse-engineering of networks is fraught with **uncertainty**: in the absence of ground truth, *in silico* modelling approaches are the most reliable model **validation** avenues; however, they are only informed by our current, most likely partial, and potentially misleading knowledge of the system.

This circularity can lead to **biases** when benchmarking inference algorithms by favouring the most suitable in the given context. **Here we investigate a more reliable validation criterion.**
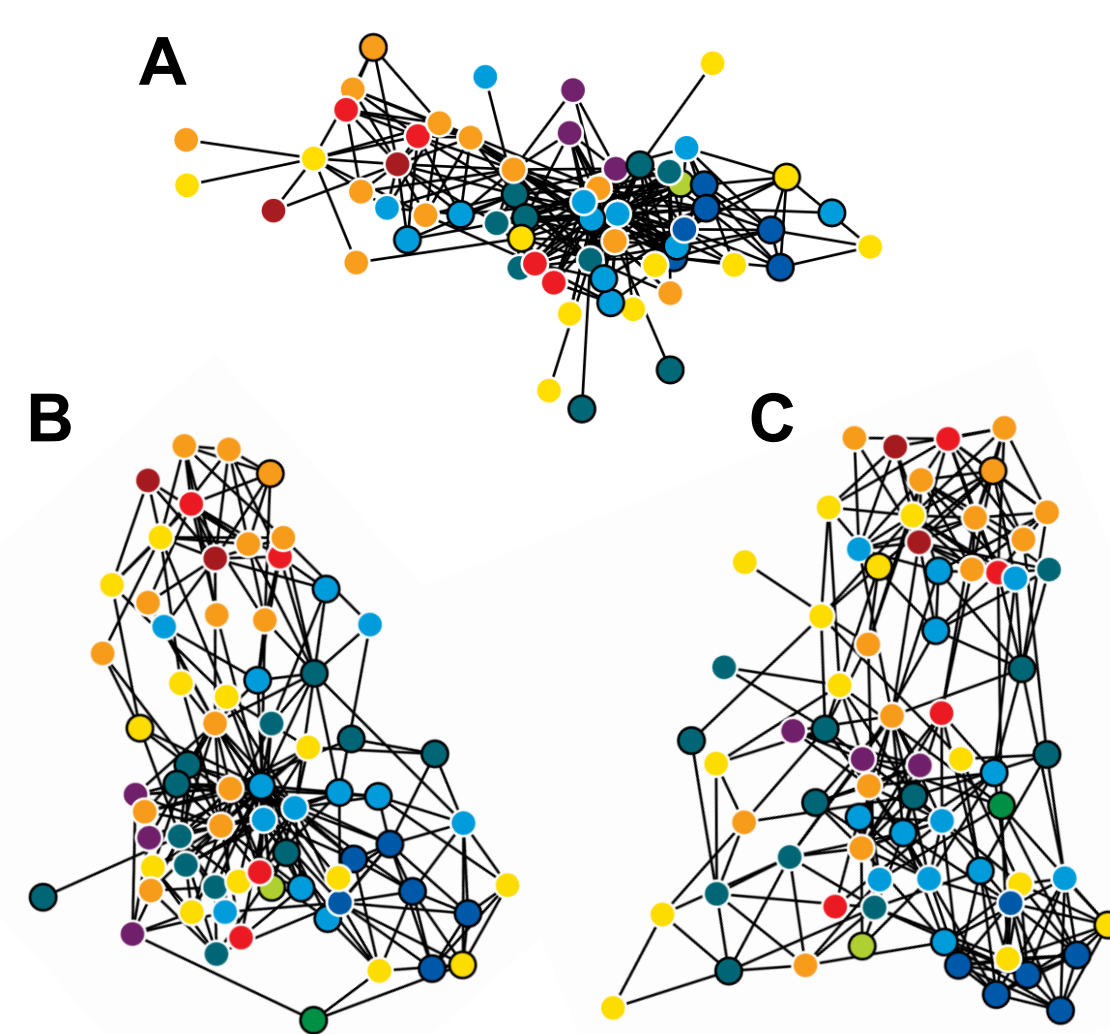

Figure 1. Networks including the top 250 edges. **A.** MI. **B.** CLR. **C.** PIDC.

Methods

We compare the output of 5 **inference algorithms** (MI [1], CLR [2], PUC [3], PIDC [3], GENIE3 [4]), on one **single-cell dataset** of mouse ESCs differentiating into neurons [5].

We look at networks in the context of the **edge threshold**, that is the number of edges to be included from the top of the ranked list of edges.

Confidence in networks is measured via the **assortativity coefficient** $r = \frac{\mathrm{Tr}\,e - \|e^2\|}{1 - \|e^2\|}$ [6] where e is a square matrix counting the number of edges between nodes with any given property and $\|x\|$ is the sum of all elements in x.

Assortativity

The assortativity coefficient $r \in [0, 1]$ measures the preferential interaction between nodes with similar properties. Properties are defined here via biologically meaningful labels assigned to each gene. **Assortativity**:

- tends to 0 as the network becomes complete;
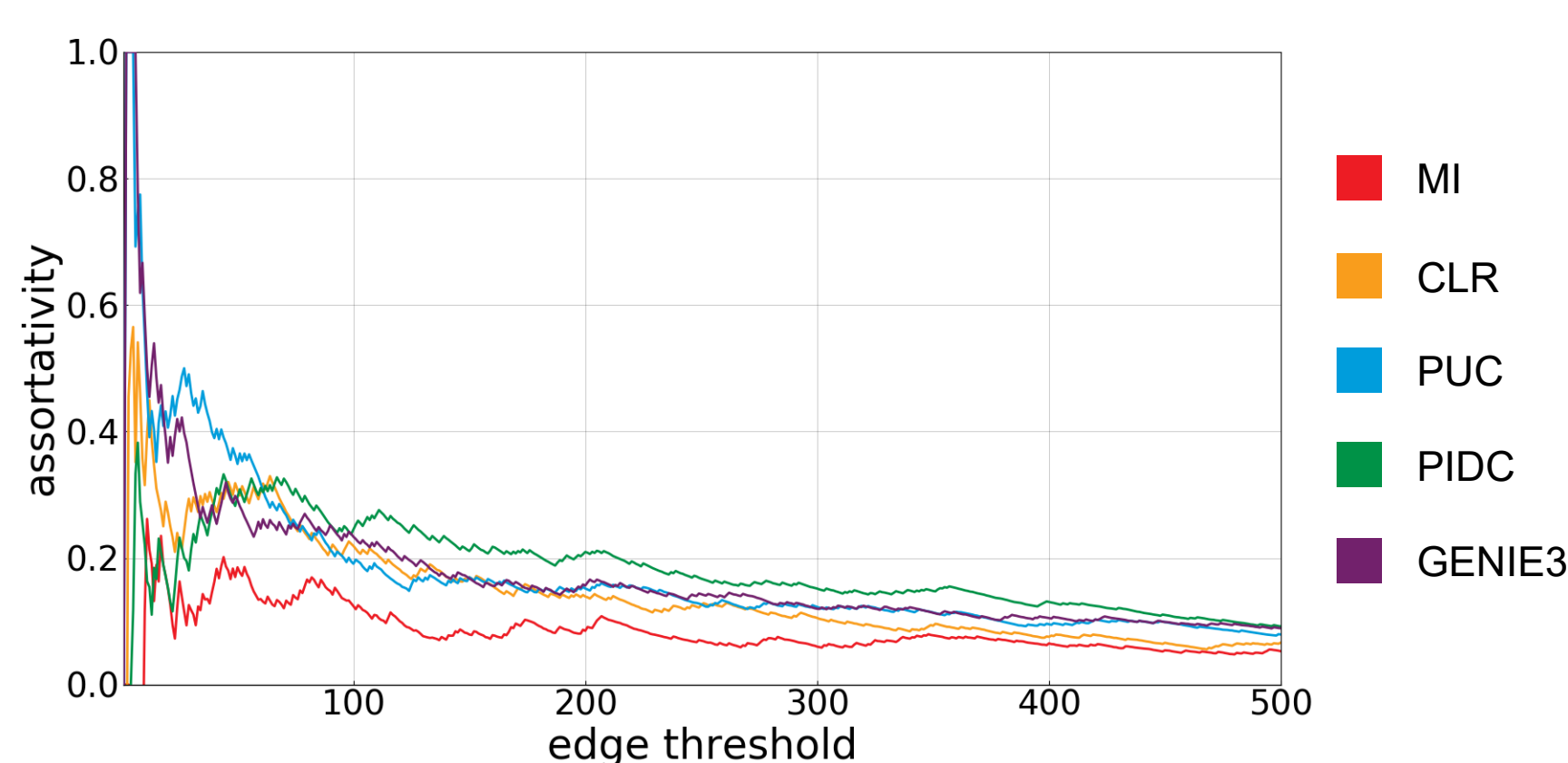- is higher for the first few hundred edges (**Figure 2**).

Figure 2. Assortativity coefficient for the 5 inference algorithms and for increasing number of edges (edge threshold).

Simulating **noise** and **randomness** allows us to study the behaviour of the assortativity coefficient (**Figure 3**). It is:

- **informative** – the signal dissipates with noise; assortativity is a function of the non-random network topology and label assignment, and not an artefact of e.g. high modularity;
- **reliable** – robust to low levels of noise i.e. tends to 0 gradually.
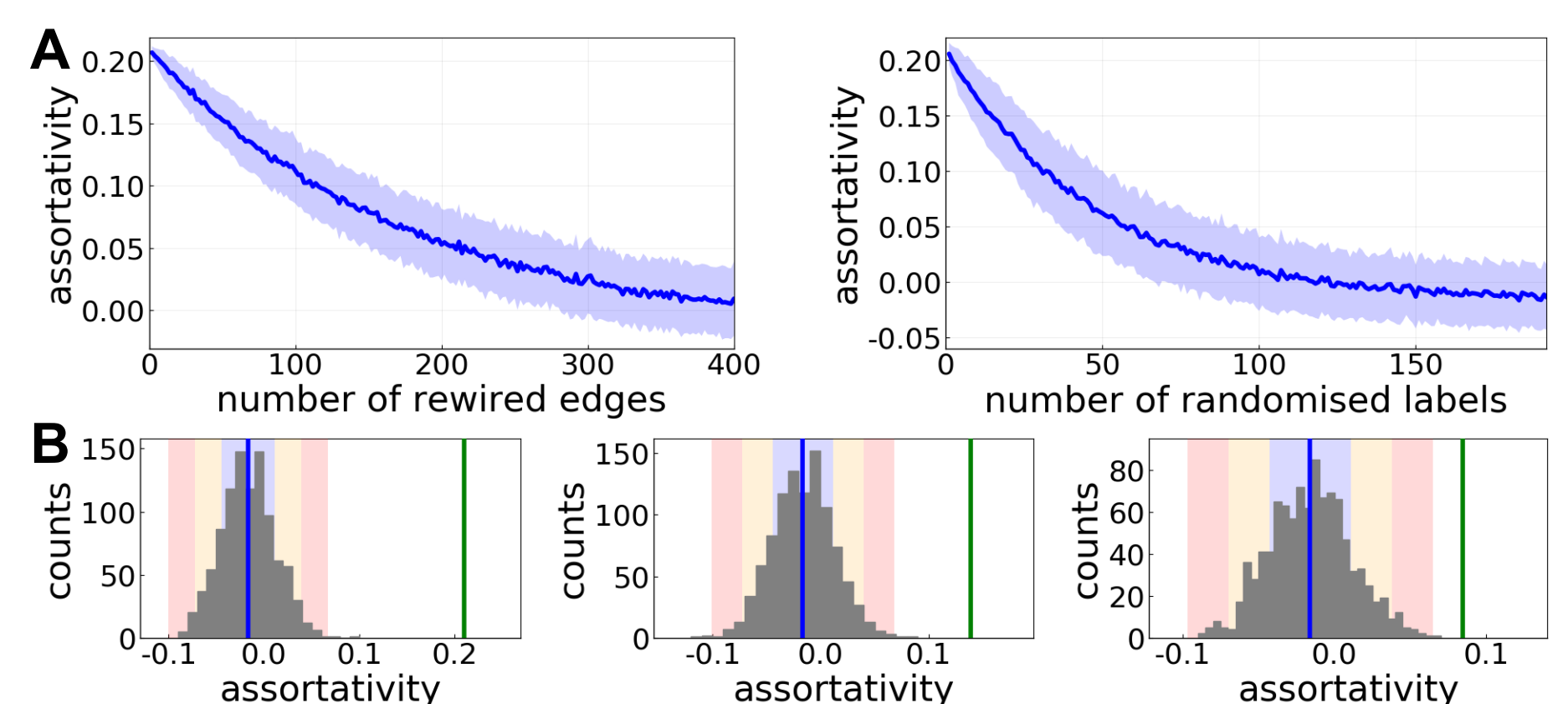

Figure 3. Impact of noise on assortativity for top 200 edges. **A.** PIDC. **B.** Expectation for random PIDC, CLR, and MI networks.

Conclusion

The assumption of mixing patterns measured via the assortativity coefficient allows us to **gain confidence** in inferred networks: fixing a relatively high assortativity threshold produces biologically **meaningful** networks (**Figure 4**).

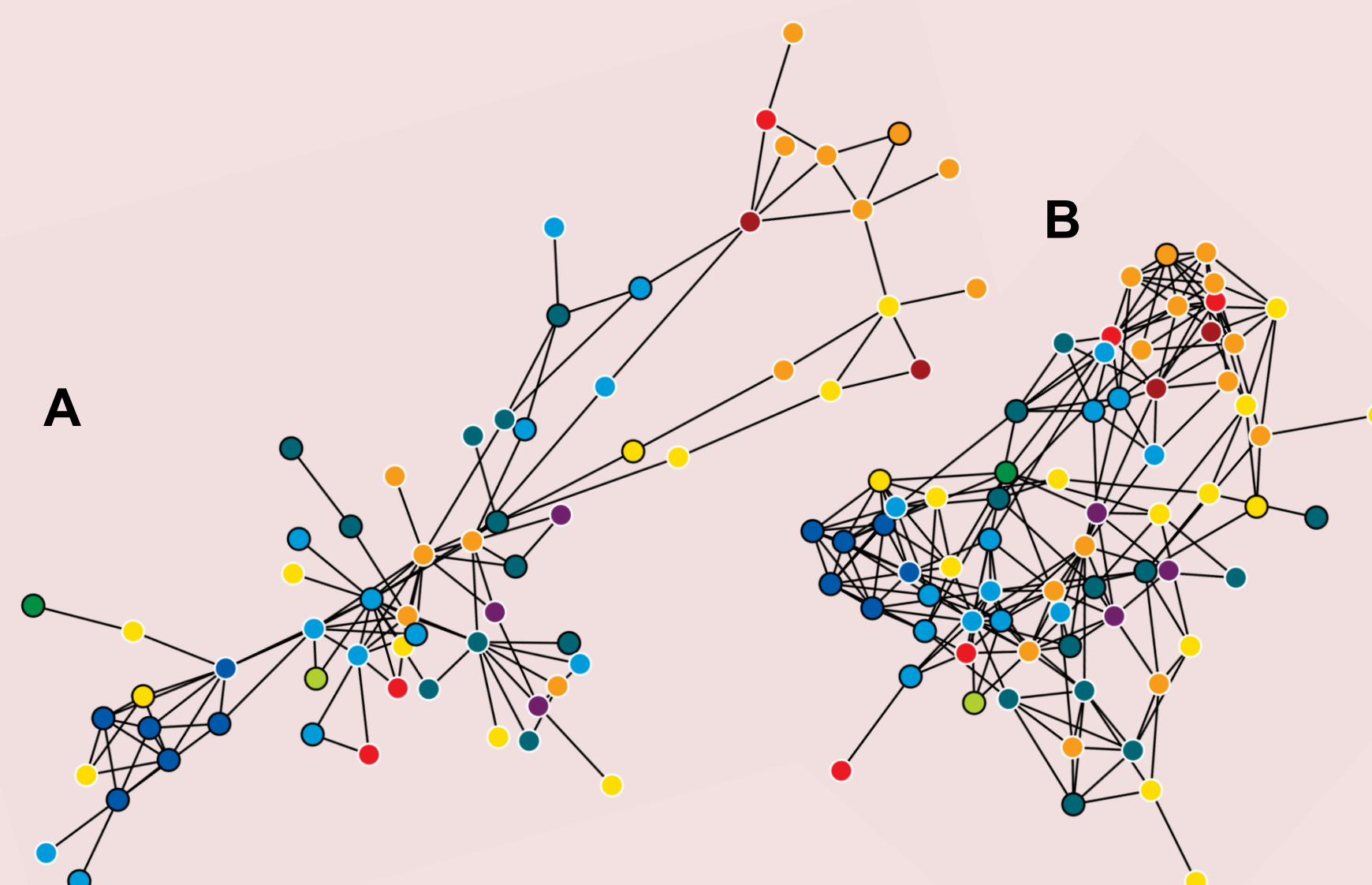This allows to **validate** models in the absence of ground truth.

Figure 4. Example networks with assortativity coefficient of $r = 0.2$. **A.** CLR at 111 edges. **B.** PIDC at 214 edges.

Future work:

- quantify the level of similarity between node properties for finer measure;
- implement a local measure of assortativity, for any given nodes.

References

1. S.S. Mc Mahon et al., *Semin. Cell Dev. Biol.* **35**, 98 (2014).
2. J.J. Faith et al, *PLoS Biol.* **5**, e8 (2007).
3. T.E. Chan, M.P.H. Stumpf & A.C. Babtie, *Cell Syst.* **5**, 251 (2017).
4. V.A. Huynh-Thu et al., *PLoS One* **5**, e12776 (2010).
5. P.S. Stumpf et al., *Cell Syst.* **5**, 268 (2017).
6. M.E.J. Newman, *Phys. Rev. Lett.* **89**, 208701 (2002).