Network archeology: phase transition in the recoverability of network history

Jean-Gabriel Young[1,2], Laurent Hébert-Dufresne[1,3,4], Edward Laurence[1,2], Charles Murphy[1,2], Guillaume St-Onge[1,2], Patrick Desrosiers[1,2,5].

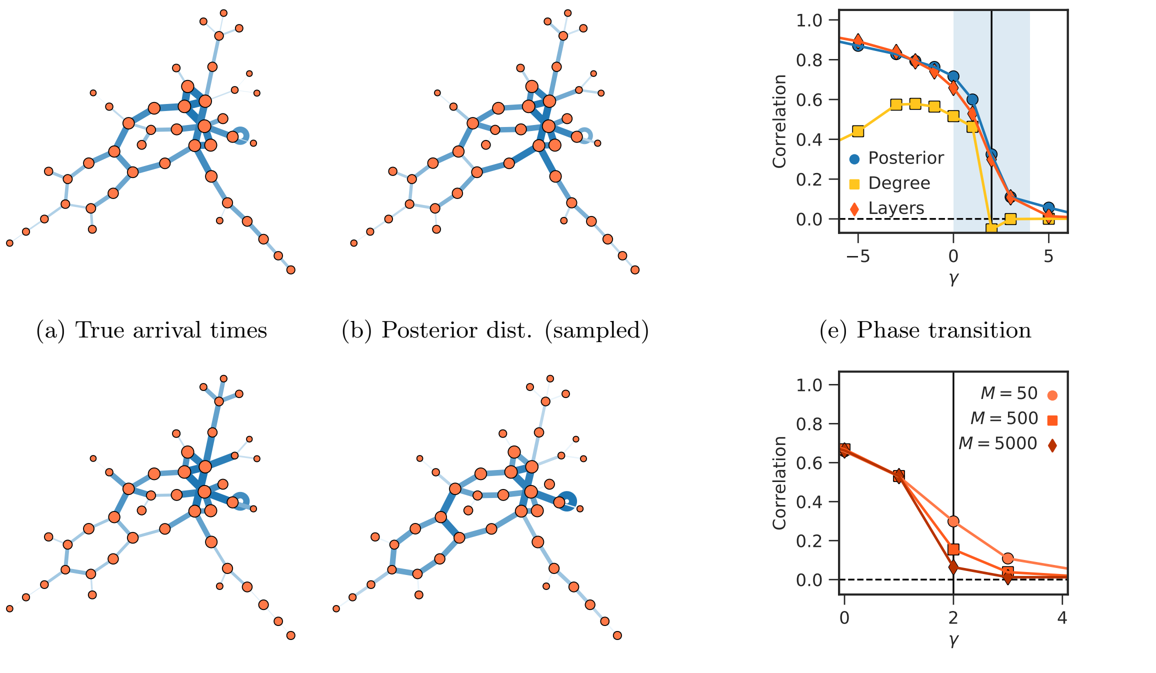
1. Département de physique, de génie physique et d'optique, Université Laval, Québec (QC), G1V 0A6, Canada
2. Centre interdisciplinaire de modélisation mathématique de l'Université Laval, Québec (QC), G1V 0A6, Canada
3. Department of Computer Science, University of Vermont, Burlington (VT), 05405, USA
4. Vermont Complex Systems Center, University of Vermont, Burlington (VT), 05405, USA
5. Centre de recherche CERVO, Québec (QC), G1J 2G3, Canada

Network growth processes can be seen as generative models for the structure and history of complex networks. We show that this point of view leads to a natural inference task, that of reconstructing the past states of a growing network, given its current structure—a difficult permutation inference problem.

Using a generalization of the preferential attachment model as the generative model [1], we introduce an importance sampling algorithm that allows us to estimate the distribution of past states, conditioned on the observed structure. We report a phase transition of the accuracy of this algorithm when it is applied to artificial networks generated by the model itself. This transition appears to be information-theoretic in nature, since simpler methods based on network properties also undergo a sharp drop in quality at the same point.

Despite the existence of a phase transition, we find that non-trivial inference is possible in a large portion of the parameter space, including a region where many real networks are found. This implies that our method can be used to infer the past states of networks with no known temporal information. Our work therefore opens the way to many important applications, such as ordering past events in networks, or generating plausible past states for real networks that are well modeled by preferential attachment.

[1] P. L. Krapivsky, S. Redner, and F. Leyvraz, *Phys. Rev. Lett.*, **85**(21), p. 4629 (2000).
[2] L. Hébert-Dufresne, J.A. Grochow, and A. Allard, *Sci. Rep.*, **6**, p. 31708 (2016).



(a) True arrival times (b) Posterior dist. (sampled) (e) Phase transition

(c) Degree-based estimation (d) Layer-based estimation [2] (f) Layer-based method on large networks

(a-d) Time of arrivals of edges for a network generated by a generalization of the preferential attachment model. The salient features of the model are a non-linear attachment kernel γ and the possibility for new links to connect pairs of existing nodes, with probability b ($b = 0.9$ and $\gamma = -1.1$ in the figure). Ticker and darker edges arrived earlier. **(e)** Average correlation of the ground-truth history (ordering of past states) with the estimated histories, for networks generated by the model with $b = 1, M = 50$ edges, and a varying kernel γ. The condensation threshold $\gamma_c = 2$ [1] is indicated with a vertical line. **(f)** The transition becomes sharper in larger networks (zoom on the shaded region of (e)).