Causal Models for Growing Networks

Gecia Bravo-Hermsdorff[1,a] **Lee M. Gunderson**[2,b] **Kayvan Sadeghi**[2,c]

[1]School of Informatics, University of Edinburgh, Edinburgh, Scotland, UK
[2]Department of Statistical Science, University College London, London, England, UK
[a,b,c]{gecia.bravo@gmail.com, l.gunderson@ucl.ac.uk, k.sadeghi@ucl.ac.uk}

Abstract

Real-world networks grow over time; statistical models based on node exchangeability are not appropriate. Instead of constraining the structure of the *distribution* of edges, we propose that the relevant symmetries refer to the *causal structure* between them. We first enumerate the 96 causal directed acyclic graph (DAG) models over pairs of nodes (dyad variables) in a growing network with finite ancestral sets that are invariant to node deletion. We then partition them into 21 classes with ancestral sets that are closed under node marginalization. Several of these classes are remarkably amenable to distributed and asynchronous evaluation. As an example, we highlight a simple model that exhibits flexible power-law degree distributions and emergent phase transitions in sparsity, which we characterize analytically. With few parameters and much conditional independence, our proposed framework provides natural baseline models for causal inference in relational data.

1 Causality and Networks

The importance of causal and relational reasoning cannot be understated. Causality is fundamentally about explaining how interventions influence observations and extrapolating from such explanations [Deutsch, 2011]. Individuals can act on their environment, learning from the result of their interventions. Multiple individuals in the same environment naturally lead to networks of interactions. Individuals themselves are complex systems, composed of many interacting parts. Collective changes of these interactions allow the individual to learn, which in turn shapes these webs of interactions [Chazelle and Wang, 2019, Chazelle, 2012, Bravo-Hermsdorff, 2023]. From the microcosm of an individual to the macrocosm of many,[1] learning is an intrinsically compositional, temporal, and collective process [Coecke, 2023, Rovelli, 2021].

A wide range of systems can be described as growing temporal networks: trades [Adamic et al., 2017], financial transactions [Arnold et al., 2024], citations [Radicchi et al., 2011], media interactions [Goglia and Vega, 2024], etc. A guiding principle for useful null models is to exploit the (often approximate) symmetries of the system [Villar et al., 2023]. For network data, the classical notion of exchangeability of graph distributions implies invariance to relabeling of the nodes [Orbanz, 2017].

However, node-exchangeable models of networks, such as graphons [Lovász, 2012, Gunderson et al., 2024] and Exponential Random Graph Models (ERGMs) [Harris, 2013, Lauritzen et al., 2018], frequently have difficulty describing real-world networks. For example, they tend to struggle to describe sparse networks, essentially treating them all as equivalent to the network without edges [Orbanz and Roy, 2014]. While various modifications have been suggested to cope with these issues, many of the hallmarks of real-world networks do not sit comfortably in this framework.

Perhaps the reason (at least partially) is that real-world networks do not typically pop into existence fully-developed. For growing networks, the order in which the nodes arrive is arguably their most basic "feature". With this perspective, instead of invariance of the *distribution* of edges (with respect to node permutation and subsampling), we ask for invariance of the causal mechanisms *generating* the edges (with respect to node deletion and marginalization).

By systematically enumerating causal models with these properties, we find statistically-streamlined models for growing networks that exhibit emergent features characteristic of real-world networks, and offer a baseline framework for causal inference in relational data.

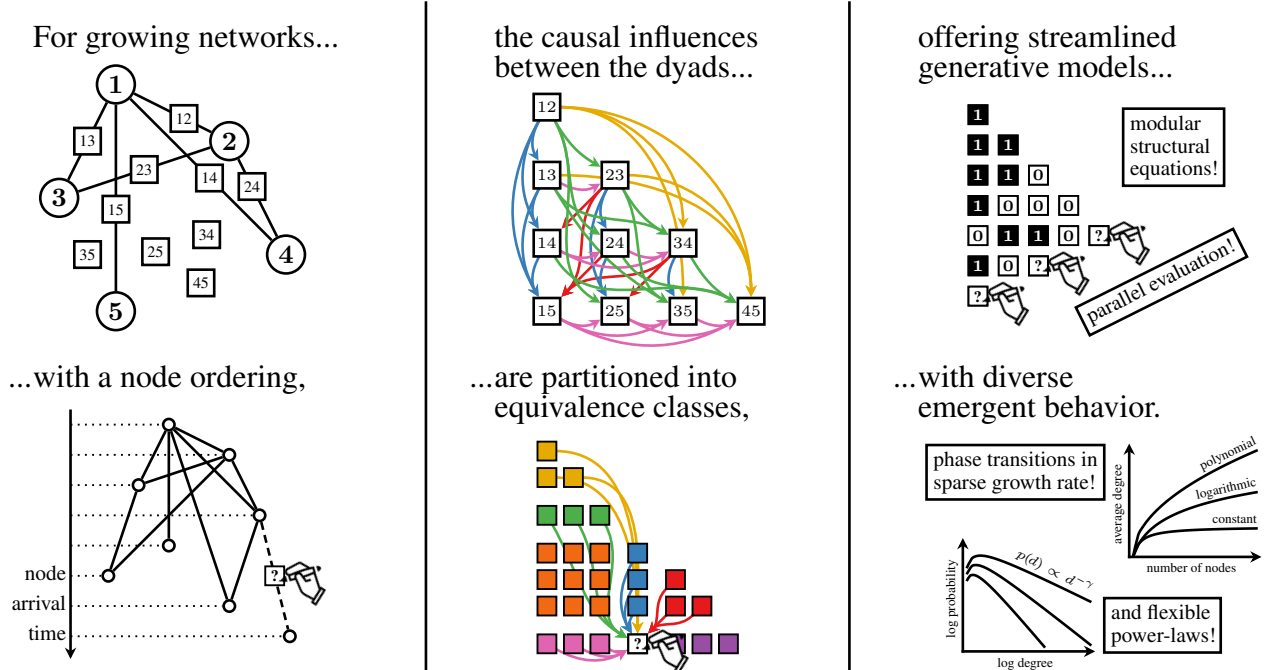[1]Such as the ants and their Aunt Hillary [Hofstadter, 1999].

Figure 1: **The narrative arc of the paper at a glance.**
Left column: The nodes of the growing network, represented as circles, have a total ordering. The variables in the model are indexed by the dyads (pairs of nodes), represented as squares (Sections 2.1 and 2.2). *Middle column:* Causal relationships between these dyad variables are represented as arrows in a causal DAG describing the generative process of the growing network. We classify the relevant equivalence classes of such causal arrows, represented by colors (Sections 2.3–2.6). *Right column:* Some combinations of these causal arrows are remarkably parallelizable, such as the model we call Distributed Affine Preferential Attachment (DAPA). In addition to a flexible power law, it naturally contains a phase transition between several well-studied sparse growth rates: from constant average degree, to logarithmic, to polynomial (Section 3.2).

We highlight one model in particular, which we call Distributed Affine Preferential Attachment (DAPA), and analytically characterize its asymptotic degree distribution, which exhibits a flexible power-law and a striking phase transition in the growth rate of the average degree. The choice of name is due to its surprisingly parallelizable causal structure. Indeed, most naturally-occurring growing networks are distributed and asynchronus systems, so it stands to reason that the causal models describing their generation might be more amenable to distributed and asynchronous computation.

Outline (see Figure 1). In Section 2, we present our framework in detail. In Section 3, we showcase a simple model with surprisingly rich behavior. In Section 4, we discuss applications of our framework to problems of generalization and causal inference. In Section 5, we discuss some promising sequels and conclude.

2 Our framework

We are concerned with describing causal models for growing networks. Both of these structures can be interpreted as graphs, so to avoid confusion, we refer to the data being generated as the ***growing network*** and the causal structure describing its generation as the ***causal DAG*** or ***meta-DAG***. For instance, in Figs. 2 and 3 the white nodes and black edges represent the growing network and the (colorful) directed arrows between these black edges represent the meta-DAG.

2.1 The Growing Network

The ***nodes*** in the growing network are indexed by the natural numbers $\mathbb{N}$ with the standard $\leq$ ordering. Intuitively, we can think of the nodes as "arriving" in that order, and then deciding with other nodes to connect to. We refer to specific node indices with lower-case letters (sometimes i and j, and sometimes a, b, c, d), with ordering implied lexicographically. Unspecified nodes are represented as open circles $\bigcirc$, with ordering implied by position (such as in Table 1).

The random variables in our model are indexed pairs of distinct nodes (ij), which we refer to as ***dyad variables***, or simply ***dyads***. That is, we take the node ordering of the growing network as given and model distributions over

their dyadic connections. While the set of nodes is infinite, one can equivalently think of the growing network as a random process defining an infinite family of probability distributions over the $\binom{n}{2}$ dyad variables between the first n nodes, for each $n \in \mathbb{N}$. The dyads variables could in principle take values from any set of outcomes, but in our example model in Section 3 they are binary (for presence or absence of an edge).

2.2 The Causal Meta-DAG

The meta-DAG is a directed acyclic graph (DAG) that represents the causal relationships between the dyad variables of the growing network (see Appendix A Figs. 5, 6, and 7 for several examples). It refers to the generative model for the growing network. As such, the dyad variables of the growing network are represented as the vertices of the meta-DAG. For instance, in Figs. 2 and 3 the white nodes and black edges represent the growing network and the (colorful) directed arrows between these black edges represent the meta-DAG. To avoid confusion, we refer to the vertices of the meta-DAG as "dyads", using the word "nodes" only to refer to the nodes of the growing network.

We refer to these directed edges as ***causal arrows***. A causal arrow $(ab) \longrightarrow (cd)$ from a parent dyad (ab) to a child dyad (cd) indicates that the outcome of the parent dyad variable (ab) *can affect* the outcome of the child dyad variable (cd), whereas the outcome of the child dyad variable (cd) *cannot affect* the outcome of the parent dyad variable (ab). This is typically phrased in terms of performing interventions on variables, such as in the do-calculus of "hard" interventions [Pearl, 1994], but also applies to various notions of "soft" interventions [Lorenz and Tull, 2023, Bravo-Hermsdorff et al., 2024].

2.3 Invariances of the Causal Meta-DAGs

One of our reasons for defining the nodes to be countably infinite is that our notions of invariance and symmetries (Table 1, Theorems 1 and 2) are more natural to state.

Deleting a node $\bigcirc$ from the growing network leaves a set of nodes that is isomorphic to the original; there is a unique order-preserving map φ from the remaining nodes to the original nodes:

$$\varphi(i) = \begin{cases} i - 1 & \text{if} \quad \bigcirc < i \\ i & \text{if} \quad i < \bigcirc \end{cases}$$

Deleting a node also deletes the dyads containing that node. This relabeling induces a map from the remaining dyads to the original dyads:

$$\varphi\big((ij)\big) = (\varphi(i)\varphi(j))$$

Similarly, for each deleted dyad, any causal arrows referencing it (either as a parent or as a child) are also deleted.

And the relabeling maps the remaining causal arrows:

$$\varphi\Big((ij) \longrightarrow (kl)\Big) = \Big(\varphi((ij)) \longrightarrow \varphi((kl))\Big)$$

A meta-DAG is a set of causal arrows between dyad variables. We want to classify all meta-DAGs that are invariant to this action of node deletion and relabeling. That is, what sets of causal arrows are isomorphic to their image under the map φ:

$$\varphi\Big(\big\{\text{causal arrows}\big\}\Big) \cong \big\{\text{causal arrows}\big\}$$

To answer this question, note that the only property that is preserved by φ is the relative ordering of the nodes in the growing network. Indeed, if we define a set of causal arrows that makes reference to nodes that are any specific number of steps away (such as "immediate predecessor"), deleting nodes changes this property, and the set of causal arrows will not be invariant. Homogenize the use of "classes" and "types" I think for the 7 arrows we should use type and class for the 21 classes (lol) of meta DAGs.

With this in mind, consider a generic dyad (ij) between two nodes $i < j$. There are five ways that an arbitrary node, which we denote by an open circle $\bigcirc$, can relate to the nodes i and j:

$\bigcirc \quad i \qquad j$		"distant" nodes	$\bigcirc < i < j$
$\textcircled{i} \qquad j$		"past" node	$\bigcirc = i < j$
$i \;\bigcirc\; j$		"recent" nodes	$i < \bigcirc < j$
$i \qquad \textcircled{j}$		"current" node	$i < \bigcirc = j$
$i \qquad j \;\bigcirc$		"future" nodes	$i < j < \bigcirc$

Note that the set of "future" nodes is infinite, whereas the other four sets of node types are finite (with two of them having a single element).

To enumerate the relationships between *dyads*, i.e., pairs of nodes, consider all the ways that *two* arbitrary nodes can relate to (ij). These two arbitrary nodes denote the parent dyad with a causal arrow to the child dyad: $(\bigcirc\bigcirc) \longrightarrow (ij)$. There are 12 types of deletion-invariant causal arrows[2] in total, which we enumerate in Appendix B. However, 5 of those options contain "future" node(s) in the parent dyad, leading to a causal meta-DAG in which child dyads have infinitely many parents.

To a avoid issues associated with infinitely many variables we require that our causal meta-DAGs have finite ancestral sets, where an ancestral set contains the parents of all variables in the set [Lauritzen, 1996]. Thereby, we exclude these 5 options in which the parent dyad has at least one "future" node. This ensures that the probability distribution resulting from a model following one such causal meta-DAG is unique [Peters and Halpern, 2021], and also makes the model straightforward to sample from.

---

[2]Plus the identity arrow from a dyad to itself makes 13.

The remaining 7 types of arrows (defined in Table 1) constitute the choices one has for constructing deletion-invariant meta-DAGs with finite ancestral sets; including any causal arrow of a given type requires one to include *all* arrows of that type.

○	ⓘ	j		HUB	$(ab) \longrightarrow (ac)$
	ⓘ ○	j		PATH	$(ab) \longrightarrow (bc)$
○	i	ⓙ		OLD	$(ac) \longrightarrow (bc)$
	i ○	ⓙ		NEW	$(bc) \longrightarrow (ac)$
○ ○	i	j		FAR	$(ab) \longrightarrow (cd)$
○	i ○	j		MID	$(ac) \longrightarrow (bd)$
	i ○ ○	j		NEAR	$(bc) \longrightarrow (ad)$
	ⓘ	ⓙ		SELF	$(ab) \longrightarrow (ab)$

Table 1: **The seven classes of deletion-invariant causal arrows with finite ancestral sets.** The *left* column represents the relative ordering of the parent dyad (○○) and the child dyad (ij). and the child dyad (ij). The *middle* column are the names we have given to each type of causal arrow. The *right* column shows the causal arrows from the parent dyad to the child dyad in terms of lexicographic node indices $a < b < c < d$. See Figs. 2 and 3 for the representation of the causal arrows in the growing network and Appendix A Figs. 5 and 6 for their representation in the causal meta-DAGs.

2.4 Seven Types of Causal Arrows

Recall that the variables in our growing network models are the dyads; the nodes have no intrinsic properties other than their (relative ordering and) dyadic relationships with other nodes. At face value, this might appear overly simplistic. For example, in a growing network of citations between publications, this would mean that decisions about which publications to cite are determined solely by their interconnected bibliographies, and not explicitly on their content or quality. However, abstracting the notions of publication and citation to include notions like "words/concepts/people" and "usage/reference/ideologies", it is not unreasonable to think of the content of such objects as encoded in the structure of what they reference.[3]

---

[3]This "relation-centric" perspective appears across many fields of mathematics. To quote Mazur [2008] on an essential tenet of category theory: *"Mathematical objects are determined by—and understood by—the network of relationships they enjoy with all the other objects of their species."*

2.4.1 Arrows Between Dyads Sharing a Node

The first four classes of causal arrows in Table 1 are from a parent dyad to a child dyad that have share a node in common. Fig. 2 displays these relations in the growing network, and Appendix A Fig. 5 shows their associated causal meta-DAGs.
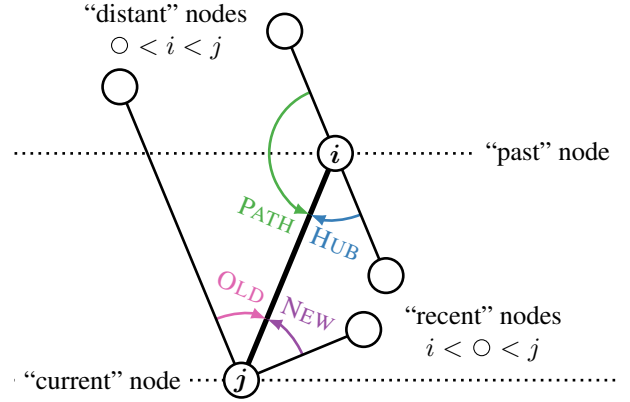

Figure 2: **Types of causal arrows between dyads that share a node.** The black edges represent the dyads of the growing network and colors represent different classes of causal arrows between them.

The HUB causal arrows mediate the tendency to reference a past concept due to how others have referenced it since then (e.g., citing a seminal paper).

The PATH causal arrows mediate the tendency to reference a concept due to the earlier concepts that it referenced (e.g., citing a paper due to its remarkable bibliography).

The OLD and NEW causal arrows describe two ways in which one could sequentially decide which previous concepts to reference. A publication deciding what to cite using OLD would do so chronologically, making decisions about the oldest publications first, and allowing those choices to modulate its decisions about which recent publications to cite as well. Using NEW is exactly the reverse; a publication begins by deciding which recent references to cite, and allowing these choices to modulate its decisions about older publications.

Realistically, bibliographies are assembled more holistically, requiring both OLD and NEW causal influences. But since these causal arrows point in opposite directions, they cannot both be included in a causal meta-DAG, as they would introduce cycles. However, in Section 5, we describe a way to include both OLD and NEW causal arrows by using structural equations that can be unrolled into an asynchronous generative process.

2.4.2 Arrows between Dyads Not Sharing a Node

The next three classes of causal arrows in Table 1 are from a parent dyad to a child dyad with *no* nodes in common. Fig. 3 displays these relations in the growing network, and Appendix A Fig. 6 shows their associated meta-DAGs.
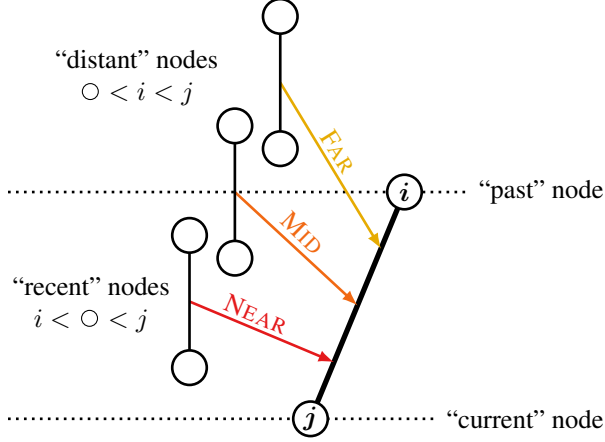

Figure 3: **Types of causal arrows between dyads that do not share a node.** Horizontal dashed lines indicate the relative ordering of the nodes in the parent dyad ($\circ\circ$) with respect to the nodes in the child dyad (ij).

These types of causal arrows can be interpreted as a sort of "context". The FAR causal arrows allow current citation decisions to be influenced by the way historical documents referenced each other. The MID causal arrows allow for influence from how more recent publications have been referencing historical documents. And the NEAR causal arrows allow for influence from how recent publications have been referencing each other.

2.4.3 The Arrow from a Dyad to Itself

Technically, the SELF arrow in Table 1 is a directed cycle, so we do not include it on our meta-DAGs. Nevertheless, it can be reasonably interpreted in a number of ways, such as the intrinsic stochasticity of the dyad variable or as a mechanism for interventions on it. For the enumeration of Theorem 2, it plays a useful role as the identity operator.

2.5 Invariance to Node Deletion

To enumerate all the possible deletion-invariant causal meta-DAGs with finite ancestral sets, we consider all combinations of these 7 causal arrow types that do not result in directed cycles. For all causal arrows except OLD and NEW, the largest node index necessarily increases from parent dyad to child dyad, so any combination of those 5 causal arrow types cannot contain a cycle. To any of these $2^5 = 32$ subsets, we may add either OLD or NEW (but not both) to create a valid deletion-invariant causal DAG.

Theorem 1 (*Deletion-invariant causal meta-DAGs*). There are $2^5 3 = 96$ deletion-invariant causal DAGs with finite ancestral sets, given by the subsets of $\{$FAR, PATH, MID, HUB, NEAR, OLD, NEW$\}$ that do not contain both OLD and NEW.

By requiring invariance to node deletion, we have described the set of causal DAGs over dyad variables that are "the same" for any subset of nodes that are actually included in the growing network. To illustrate this idea, consider indexing publication nodes by the precise time that the authors made their final edit, in terms of milliseconds since 1 Jan 1970. One can think of the resulting citation network as being initialized with over a trillion potential nodes, then deleting the vast majority of those that did not contain a publication. In a sense, invariance to node deletion is invariance to "that which never existed".

2.6 Invariance to Node Marginalization

Similarly, one might be interested in invariance to "that which was not observed". Variables that exist but are not observed cannot simply be deleted from a causal model; in order to preserve the causal and probabilistic relationships between the remaining variables, they must be *marginalized*. For a causal DAG, marginalizing a variable is a two-step process: first add arrows from all of its parents to all of its children, then delete that variable [Richardson and Spirtes, 2002]. In general, one must also account for the stochastic component of this now-unobserved latent variable, often represented by introducing bidirected arrows between the children. We will ignore such bidirected arrows for now, and address them momentarily.

In our setting, the random variables are the dyads between the nodes of the growing network. So marginalization of a node in the growing network means marginalizing all dyad variables containing that node. Of the 96 deletion-invariant causal meta-DAGs, 21 are also invariant to node marginalization, shown in Appendix C Fig. 8. They are precisely the meta-DAGs that are *transitively closed*, that is, all dyad variables have the property that all of its parent dyads have arrows pointing to all of its child dyads.

To obtain the transitive closure of a causal meta-DAG, one can repeatedly perform the first step of marginalizing a variable, adding arrows from parents to children *without* deleting the variable, until no more arrows can be added. For example, if the causal model contains the arrows $(ab) \rightarrow (ij)$ and $(ij) \rightarrow (rs)$, then it must also include the arrow $(ab) \rightarrow (rs)$. As this first step is the only difference between deletion and marginalization, taking the transitive closure renders them equivalent, ensuring invariance to marginalization as well as deletion. Of the 96 deletion-invariant causal meta-DAGs in Theorem 1, there are 21 that are also invariant to marginalization (Theorem 2, see also Appendix C Fig. 8).

With the directed causal arrows accounted for, we return to the bidirected arrows mentioned earlier. Indeed, these must be included if one wishes to preserve conditional independence statements between the remaining variables. But for our setting, these statements do not distinguish between many of the resulting causal structures. For example, any causal meta-DAG with the FAR causal arrow would require bidirected arrows between every dyad (marginalizing the first two nodes will suffice), resulting in zero conditional independence statements for 13 of the 21 meta-DAGs in Appendix C Fig. 8.

However, causal models encode more than just conditional independence; they also encode how the distribution might change as a result of performing interventions. Consider the question of which dyad variables $(\bigcirc\bigcirc)$ *cannot* be affected by an intervention on the outcome of (ij). For causal DAGs, the answer is unaffected by the presence of bidirected arrows; $(\bigcirc\bigcirc) \perp \mathrm{do}\big((ij)\big)$ if there is *no* arrow $(ij)\!\longrightarrow\!(\bigcirc\bigcirc)$ in its transitive closure. We therefore consider invariance (to node deletion and contraction) of this interventional structure to define and distinguish between the 21 meta-DAGs in Theorem 2.

Theorem 2 (*Deletion and marginalization-invariant causal meta-DAGs*). There are 21 causal meta-DAGs with finite ancestral sets whose interventional structure is invariant to both node deletion and marginalization.

2.6.1 Computing the Transitive Closure

To take the transitive closure of a set of causal arrows, consider all the ways those arrows can compose. That is, for all cases in which the child dyad of the first causal arrow is the parent dyad of the next causal arrow, which types of causal arrows could point from the parent dyad of the first arrow to the child dyad of the second?

For example: $\text{PATH} \blacktriangleright \text{PATH} \blacktriangleright = \{\text{FAR} \blacktriangleright\}$, since the effect of traversing two PATH causal arrows is always equivalent to traversing a single FAR causal arrow. Some compositions have the potential to be equivalent to multiple arrow types: $\text{HUB} \blacktriangleright \text{OLD} \blacktriangleright = \{\text{MID} \blacktriangleright, \text{PATH} \blacktriangleright, \text{FAR} \blacktriangleright\}$. Technically, this composition depends on the order: $\text{OLD} \blacktriangleright \text{HUB} \blacktriangleright = \{\text{MID} \blacktriangleright\}$, but this noncommutativity will not be important for our purposes.

Define the composition of two *sets* of arrow types as the union over all combinations of arrows from the two sets. Then, to take the transitive closure, begin with an initial set of causal arrows (along with the SELF arrow, which plays the role of the identity). Compose this set of arrows with itself to obtain a (possibly) larger set of causal arrow types. Continue this process until the set is no longer increasing to obtain the transitive closure.

For example, $\{\text{SELF} \blacktriangleright, \text{HUB} \blacktriangleright, \text{PATH} \blacktriangleright\}$ composed with itself results in $\{\text{SELF} \blacktriangleright, \text{HUB} \blacktriangleright, \text{PATH} \blacktriangleright, \text{FAR} \blacktriangleright\}$. Composing $\{\text{SELF} \blacktriangleright, \text{HUB} \blacktriangleright, \text{PATH} \blacktriangleright, \text{FAR} \blacktriangleright\}$ with itself results in the same set,

so it is transitively closed. Starting with the 96 deletion-invariant subsets of arrows from Theorem 1, one obtains the 21 fixed points of this process, corresponding to the meta-DAGs in Theorem 2.

Computationally, we implemented this by representing the arrows and their composition as matrices and matrix multiplication. While these 21 equivalence classes can easliy be found by hand, this quickly become unwieldily for generalizations of the procedure.[4]

3 Streamlined Network Models

Our classification of invariant causal meta-DAGs imposes requirements for where causal arrows can appear between the dyad variables, but so far we have said nothing of what the structural equations of a growing network might be. Indeed, one *could* use a different function for each dyad with a complex dependency on all of its parents, and such a model would still generate network distributions that are faithful to the causal structure of our meta-DAGs.

However, in the spirit of our framework of "invariance of causal mechanisms", we propose using the *same* structural equation for all dyad variables in the model. Since the number of parents of A natural choice is to use basic summary statistics of the different types of causal parents.

In this section, we present a simple model for binary (edge or no-edge) dyad variables that implements preferential attachment using HUB and PATH causal arrows, generating (simple undirected) growing networks with controllable power-law degree distributions (Section 3.2). More important than the model itself is the rather surprising insight it provides: *reducing* dependencies between the dyad variables leads to *increased* diversity in the asymptotic behavior of the growing network.

3.1 Basic Preferential Attachment

While the relationship between preferential growth and scale-free distributions had long been known [Eggenberger and Pólya, 1923, Simon, 1955, Price, 1965], the model of Barabási and Albert [1999] made it widely famous in the network science community. So much so that their names have become synonymous with power-law degree distributions.

The simplest statement of their model [Pósfai and Barabási, 2016] has a single parameter, m. Start with a clique of m nodes. Then, at each iteration, m nodes are selected proportional to their current degree, and a new node is added to the growing network by connecting it to

---

[4]For example, for simple hypergraphs with cardinality 3, there are already 37 type of causal arrows between the triads (subsets of 3 nodes) with finite ancestral sets.

these m selected nodes.

Many extensions to this model have been described [Ray, 2024]. Such as introducing parameters to control correlations between neighboring degrees [Avin et al., 2020] or to promote clustering [Eikmeier and Gleich, 2019]; adding node covariates [Bianconi and Barabási, 2001, Lee et al., 2015]; and considering alternative attachment functions [Krapivsky et al., 2000].

3.1.1 Poissonified Preferential Attachment

An interesting line of modifications involves removing statistical correlations between pairs of edges [Bollobás et al., 2007, Wang and Resnick, 2020]. Most models of preferential attachment have a parameter m, specifying the precise number of edges that each new node j makes with the previous nodes. This results in a small anticorrelation between the dyad variables within each iteration.

Alternatively, one could compute probabilities for each edge that are proportional to their degrees, such that m edges will be added in expectation [Van Der Hofstad, 2024].[5] By allowing the number of edges added to fluctuate appropriately, the anticorrelation in the previous model can be removed without qualitatively changing the asymptotic behavior; the degree distribution has the same power-law tail $p(d) \propto d^{-3}$ and average degree $\langle d \rangle = 2m$.

While this modification renders the dyad variables $(\bigcirc j)$ conditionally independent given the current degrees of the previous nodes, their outcomes still depend on the entire network up to the previous iteration. This is due to the fact that the sum of their edge probabilities has been scaled to be equal to m. Is this dependence on all previous variables necessary? How much dependence can we remove while still retaining the hallmark features of preferential attachment models?

3.2 Distributed Affine Preferential Attachment

What if the probability of an edge were to depend *only* on the degree statistics of the older node to which it might connect, with zero dependence on the degrees of the other nodes in the network? This would require only the HUB and PATH causal arrows (see causal meta-DAG in Fig. 7 *top-left*). Motivated by this, we consider the edge probabilities to be an affine function of the degree statistics corresponding to these two types of causal arrows, leading to the model we call *Distributed Affine Preferential Attachment*, or *DAPA*, as sampling from it is highly parallelizable (see

---

[5]Setting aside cases, where this cannot be done: $\frac{d_i}{\sum_i d_i} > \frac{1}{m}$, which corresponds to a probability greater than 1 for connecting to node i. Such cases are unlikely to happen, but are technically possible when $m > 2$.

Section 5).

Explicitly, we have the following probabilistic model for the edges of a growing network:

$$x_{ij} \sim \text{Bernoulli}(p_{ij}) \tag{1}$$

$$p_{ij} = \frac{\alpha + \theta_{\text{in}} d_i^{\text{in}} + \theta_{\text{out}} d_i^{\text{out}}}{j - 2 + \alpha + \beta} \tag{2}$$

$$d_i^{\text{in}} = \sum_{\bigcirc = i+1}^{j-1} x_{i\bigcirc} \qquad d_i^{\text{out}} = \sum_{\bigcirc = 1}^{i-1} x_{\bigcirc i} \tag{3}$$

where $x_{ij} = 1$ indicates an edge between nodes i and j (and $x_{ij} = 0$ indicates no edge).

Surprisingly, while drastically decreasing the dependence between the dyads, this model exhibits an increased diversity in its asymptotic behavior (see Fig. 4, Theorems 3 and 4, proofs in Appendix D).
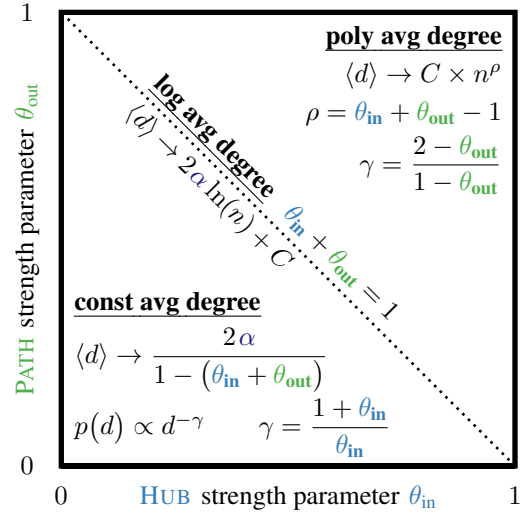

Figure 4: **Sparsity and power-laws in the DAPA model.**

3.2.1 Three Sparsity Regimes

Sparsity can be characterized in terms of the average degree as a function of the number of nodes $\langle d(n) \rangle = \frac{2E(n)}{n}$. For dense networks, the average degree grows linearly in the number of nodes: $\langle d(n) \rangle = \mathcal{O}(n)$, while for sparse networks, the average degree grows sublinearly, $\langle d(n) \rangle = o(n)$ [Van Der Hofstad, 2024]. Networks with power-law degree distributions are naturally sparse; in order for the degrees to span arbitrarily many orders of magnitude, "most" nodes have a vanishingly small degree compared to the maximum. Despite the apparent similarity of our DAPA model to preferential attachment models that "hard-code" the average degree, our model exhibits an emergent phase transition to sparse scalings with growing average degree.

Theorem 3 (*Phase transitions in the sparsity of the DAPA model*)**.** Our DAPA model exhibits three qualitatively different asymptotic behaviors for the average degree $\langle d(n) \rangle$

constant: $\dfrac{2\alpha}{1 - (\theta_{\text{in}} + \theta_{\text{out}})}$ $0 < \theta_{\text{in}} + \theta_{\text{out}} < 1$

logarithmic: $2\alpha \log(n) + C$ $\theta_{\text{in}} + \theta_{\text{out}} = 1$

polynomial: $C \times n^{\theta_{\text{in}} + \theta_{\text{out}} - 1}$ $1 < \theta_{\text{in}} + \theta_{\text{out}} < 2$

3.2.2 And a Flexible Power-law Degree Distribution

For all three sparsity regimes, the growing networks exhibit a range of power-law scalings in their degree distributions.

Theorem 4 (*Phase transitions in the power-law of the degree distribution of the DAPA model*)**.** The asymptotic probability that a random node has degree d has a tail of the form $p(d) \propto d^{-\gamma}$, where the scaling exponent depends on either θ_{in} or θ_{out}:

constant: $\gamma = \dfrac{1 + \theta_{\text{in}}}{\theta_{\text{in}}}$ $0 < \theta_{\text{in}} + \theta_{\text{out}} \leq 1$

polynomial: $\gamma = \dfrac{2 - \theta_{\text{out}}}{1 - \theta_{\text{out}}}$ $1 \leq \theta_{\text{in}} + \theta_{\text{out}} < 2$

These results allow for a nice interpretation of the parameters. The power-law scaling exponent is determined by $0 \leq \theta_{\text{in}} \leq 1$, which is associated to the strength of preferential attachment due to HUB causal arrows. The growth rate of the edge density is modulated by $0 \leq \theta_{\text{out}} \leq 1$, which is associated to the strength of preferential attachment due to PATH causal arrows. The final and initial values of the edge density are influenced by $\alpha > 0$ and $\beta \geq 0$, which can be thought of as associated with the SELF arrows.

4 Some Applications

4.1 Inference and Generalization

This flexibility of asymptotic behaviors from a simple model is a useful property for extrapolating from limited data. For instance, consider observing a growing network that is still in its relative infancy. The average degree is increasing as a function of the number of nodes, but it is slowing down; will it converge to some constant value, or if not, at what rate might it increase? The degree distribution is currently more spread than a network with independent edges, but there is not yet a region that looks

linear on a log-log plot;[6] what might its scaling exponent be once many more nodes are added?

By fitting the parameters of a simple structural equation to initial observations, one might still be able to estimate such quantities related to the asymptotic behavior of the growing network, despite those features not yet being present.

4.2 Interventions and Counterfactuals

In this section, we illustrate how to use our framework to answer interventional and counterfactual causal queries using our beloved running examples.

Suppose you are about to submit a publication, and you want to add a few citations to help it reach a larger audience. To estimate the net effect of such strategic citations, one could fit the parameters of a causal model such as our DAPA model to the citation network.

By approximating the strength of various causal mechanisms, one can run the model forward to estimate the number of additional citations one might receive as a result. This is an example of an *interventional* question, as the answer involves quantifying (the result of performing an action) over a *distribution* of possible futures.

Now suppose you have a older publication that you really feel should have more citations, and you are deciding how much to regret not promoting it more at the time. This is an example of a *counterfactual* question, as now the answer involves quantifying the difference between one *particular* outcome (that was actually observed), and another (that *could* have occurred, but did not).

To estimate net effect of such fictional actions, one can use the structural equations of a causal model. For example, the randomness in Equation (1) can be represented explicitly by introducing an (unobserved) random variable:

$$\epsilon_{ij} \sim \mathcal{Uniform}(0, 1) \tag{4}$$
$$x_{ij} = \text{sign}(p_{ij} - \epsilon_{ij}) \tag{5}$$

From the estimated parameters of the model, and the actual observed data, one can use this form to estimate the likelihood of such counterfactual changes. Essentially, this involves performing bayesian updates to the p_{ij} and x_{ij}, while treating the ϵ_{ij} as fixed [Pearl, 2009].

---

[6]Indeed, estimating the power-law exponent from a degree distribution is a notoriously tricky procedure [Clauset et al., 2009, Artico et al., 2020]. No finite network is truly scale-free; even if there is an obvious power-law that fits the majority of the degree distribution, there are typically significant deviations from it in the frequency of nodes with small degrees.

5 Distributed Discussion

Initially, we set out to classify similar causal models for networks that grow one node at a time. This requirement turned out to be overly restrictive, and we were pleasantly surprised to find causal structures that were less rigid in the order of their generation. In particular, causal models in which the dyads only depend on one of the two "quadrants" of past dyads can be evaluated in a distributed manner (as illustrated in Fig. 1). For these models, coarse-graining the rows and columns of grid results in blocks of dyad variables with a similar causal structure. Thus, one can assign workers to different blocks of dyads to evaluate them in parallel requiring communication only when workers move to the next block. For example, in the model below, w workers can alternate between evaluating blocks of size $\frac{n}{w}$-by-$\frac{n}{w}$ and a total of $2w$ rounds of communication.

DAPA model with clustering — HUB + PATH + OLD. As an extension of the DAPA model, one could include OLD causal arrows (see bottom-right of Fig. 7 for its causal meta-DAG.). This addition would, for example, allow for the in-degrees to also exhibit a power-law. Moreover, when both OLD and PATH arrows are present, it is possible to promote clustering via triadic closure, as the similarity between the connections that nodes i and j make with the "distant" nodes $\bigcirc < i < j$ can influence the likelihood that i and j themselves form a connection.

A "bottom-up" causality — HUB + NEW. The DAPA model and its extension including OLD have a "top-down" sort of causal structure, with dyads containing older nodes influencing dyads containing newer nodes. Conversely, the causal model with HUB and NEW depends on the other "quadrant" of dyads, and instead has a sort of "bottom-up" sort of causal structure (see its causal meta-DAG at the top-right of Fig. 7). That is, the dyads containing nodes that are *closer together* in the ordering influence the outcomes of dyads containing nodes that are *further apart*. This causal meta-DAG could be useful for modeling "local" clustering between nodes that occur at similar times. While this combination can also be sampled in a distributed manner, it is incompatible with the causal models mentioned above.

Composing Causalities. This incompatibility can be resolved by composing multiple causal DAGs together, performing the acyclic computations in one, and using the outcomes to seed the other. Indeed, many apparent cycles in causal models can be "fine-grained" away, as they actually represent an underlying series of back-and-forth communication. Decomposing the behavior of complex interconnected networks into a collection causal DAGs could prove a fruitful direction for future study.

References

Lada Adamic, Celso Brunetti, Jeffrey H Harris, and Andrei Kirilenko. Trading networks. The Econometrics Journal, 20(3):S126–S149, 2017.

Naomi A Arnold, Peijie Zhong, Cheick Tidiane Ba, Ben Steer, Raul Mondragon, Felix Cuadrado, Renaud Lambiotte, and Richard G Clegg. Insights and caveats from mining local and global temporal motifs in cryptocurrency transaction networks. Scientific Reports, 14(1): 26569, 2024.

Igor Artico, I Smolyarenko, Veronica Vinciotti, and Ernst C Wit. How rare are power-law networks really? Proceedings of the Royal Society A, 476(2241): 20190742, 2020.

Chen Avin, Hadassa Daltrophe, Barbara Keller, Zvi Lotker, Claire Mathieu, David Peleg, and Yvonne-Anne Pignolet. Mixed preferential attachment model: Homophily and minorities in social networks. Physica A: Statistical Mechanics and its Applications, 555:124723, 2020.

Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. Science, 286(5439):509–512, 1999.

Ginestra Bianconi and A-L Barabási. Competition and multiscaling in evolving networks. Europhysics letters, 54(4):436, 2001.

Béla Bollobás, Svante Janson, and Oliver Riordan. The phase transition in inhomogeneous random graphs. Random Structures & Algorithms, 31(1):3–122, 2007.

Gecia Bravo-Hermsdorff. Quantifying human priors over social and navigation networks. In International Conference on Machine Learning, pages 3063–3105. PMLR, 2023.

Gecia Bravo-Hermsdorff, David Watson, Jialin Yu, Jakob Zeitler, and Ricardo Silva. Intervention generalization: A view from factor graph models. Advances in Neural Information Processing Systems, 36, 2024.

Bernard Chazelle. The dynamics of influence systems. In 2012 IEEE 53rd Annual Symposium on Foundations of Computer Science, pages 311–320. IEEE, 2012.

Bernard Chazelle and Chu Wang. Iterated learning in dynamic social networks. Journal of Machine Learning Research, 20(29):1–28, 2019.

Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. Power-law distributions in empirical data. SIAM review, 51(4):661–703, 2009.

Bob Coecke. Compositionality as we see it, everywhere around us. In The Quantum-Like Revolution: A Festschrift for Andrei Khrennikov, pages 247–267. Springer, 2023.

David Deutsch. The beginning of infinity: Explanations that transform the world. penguin uK, 2011.

Florian Eggenberger and George Pólya. Über die statistik verketteter vorgänge. ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik, 3(4):279–289, 1923.

Nicole Eikmeier and David F Gleich. Triangle preferential attachment has power-law degrees and eigenvalues; eigenvalues are more stable to network sampling. arXiv preprint arXiv:1904.12989, 2019.

Diletta Goglia and Davide Vega. Structure and dynamics of growing networks of reddit threads. Applied Network Science, 9(1):48, 2024.

Lee Gunderson, Gecia Bravo-Hermsdorff, and Peter Orbanz. The graph pencil method: mapping subgraph densities to stochastic block models. Advances in Neural Information Processing Systems, 36, 2024.

Jenine K Harris. An introduction to exponential random graph modeling, volume 173. Sage Publications, 2013.

Douglas R Hofstadter. Gödel, Escher, Bach: an eternal golden braid. Basic books, 1999.

Paul L Krapivsky, Sidney Redner, and Francois Leyvraz. Connectivity of growing random networks. Physical Review Letters, 85(21):4629, 2000.

Steffen Lauritzen, Alessandro Rinaldo, and Kayvan Sadeghi. Random networks, graphical models and exchangeability. Journal of the Royal Statistical Society Series B: Statistical Methodology, 80(3):481–508, 2018.

Steffen L Lauritzen. Graphical models, volume 17. Clarendon Press, 1996.

Jay Lee, Manzil Zaheer, Stephan Günnemann, and Alex Smola. Preferential attachment in graphs with affinities. In Artificial Intelligence and Statistics, pages 571–580. PMLR, 2015.

Robin Lorenz and Sean Tull. Causal models in string diagrams. arXiv preprint arXiv:2304.07638, 2023.

László Lovász. Large networks and graph limits, volume 60. American Mathematical Soc., 2012.

Hosam Mahmoud. Pólya urn models. Chapman and Hall/CRC, 2008.

Barry Mazur. When is one thing equal to some other thing. Proof and other dilemmas: Mathematics and philosophy, pages 221–241, 2008.

Peter Orbanz. Subsampling large graphs and invariance in networks. arXiv:1710.04217, 2017.

Peter Orbanz and Daniel M Roy. Bayesian models of graphs, arrays and other exchangeable random structures. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(2):437–461, 2014.

Judea Pearl. A probabilistic calculus of actions. In Uncertainty in Artificial Intelligence, pages 454–462. Elsevier, 1994.

Judea Pearl. Causality. Cambridge university press, 2009.

Erol Peköz, Adrian Röllin, and Nathan Ross. Pólya urns with immigration at random times. Bernoulli, 25(1): 189–220, 2029.

Spencer Peters and Joseph Y Halpern. Causal modeling with infinitely many variables. arXiv preprint arXiv:2112.09171, 2021.

Márton Pósfai and Albert-László Barabási. Network science. Citeseer, 2016.

Derek J De Solla Price. Networks of scientific papers: The pattern of bibliographic references indicates the nature of the scientific research front. Science, 149(3683): 510–515, 1965.

Filippo Radicchi, Santo Fortunato, and Alessandro Vespignani. Citation networks. Models of science dynamics: Encounters between complexity theory and information sciences, pages 233–257, 2011.

Rounak Ray. Stochastic processes on preferential attachment models. Phd thesis, Eindhoven University of Technology, 2024. Available at https://arxiv.org/abs/2411.14111.

Thomas Richardson and Peter Spirtes. Ancestral graph markov models. The Annals of Statistics, 30(4):962–1030, 2002.

Carlo Rovelli. Helgoland: Making Sense of the Quantum Revolution. Riverhead Books, New York, 2021.

Herbert A Simon. On a class of skew distribution functions. Biometrika, 42(3/4):425–440, 1955.

Remco Van Der Hofstad. Random graphs and complex networks, volume 54. Cambridge university press, 2024.

Soledad Villar, David W Hogg, Weichi Yao, George A Kevrekidis, and Bernhard Schölkopf. Towards fully covariant machine learning. arXiv preprint arXiv:2301.13724, 2023.

Tiandong Wang and Sidney I Resnick. A directed pref-
erential attachment model with poisson measurement.
arXiv preprint arXiv:2008.07005, 2020.

A Examples of Causal Meta-DAGs

In this section, we display the causal meta-DAGs associated with the various types of causal arrows for a growing network with five nodes, where the X_{ij} represent the dyad variables. Bear in mind that these structures continue indefinitely for networks with any number of nodes.
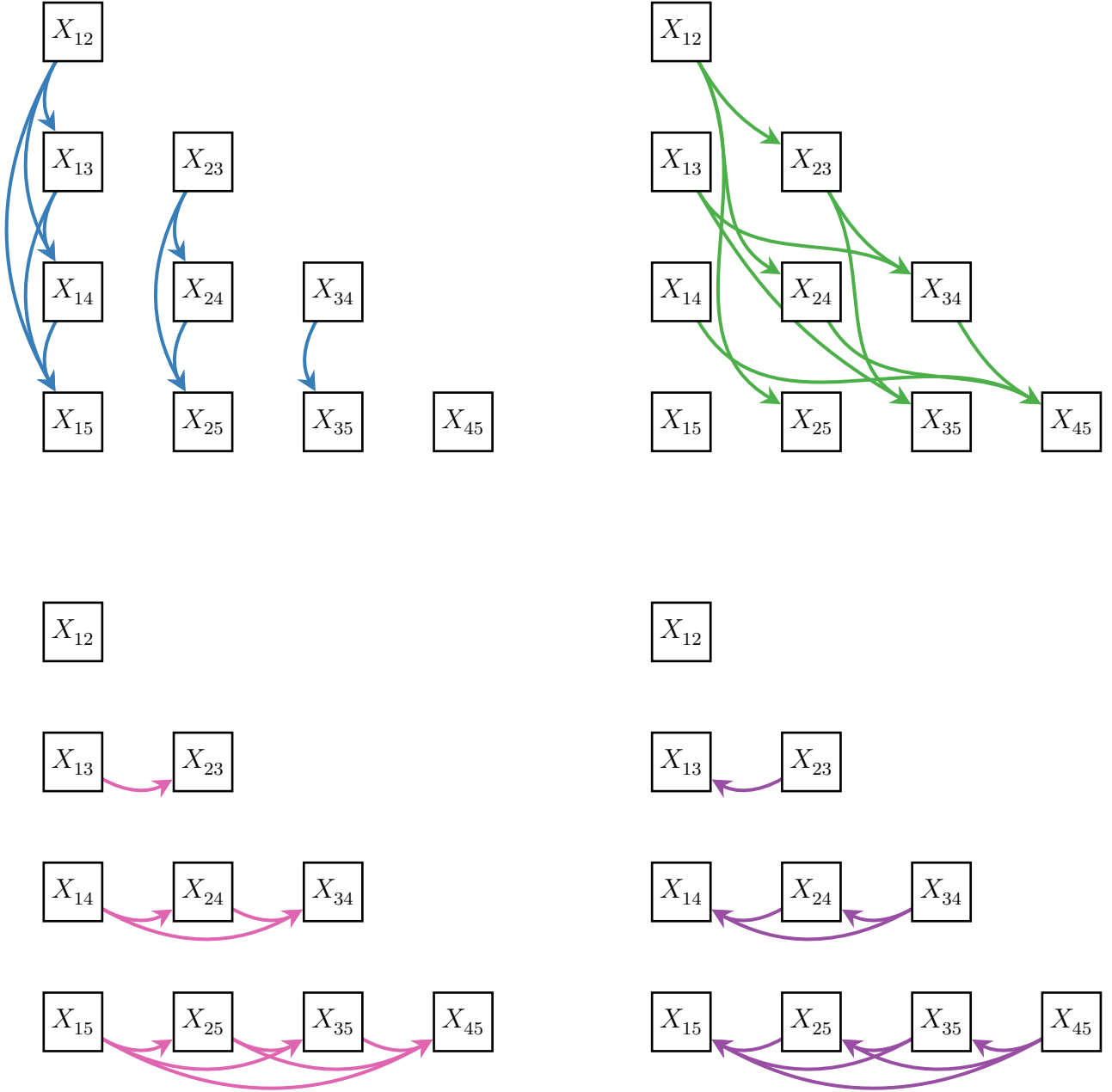

Figure 5: **Causal graphs for each of the types of causal arrows between dyads that share a node.**
Causal meta-DAGs between dyads of a growing network with 5 nodes that are compatible with network models having the following types of causal arrows: HUB (*top-left*); PATH (*top-right*); OLD (*bottom-left*); and NEW (*bottom-right*).
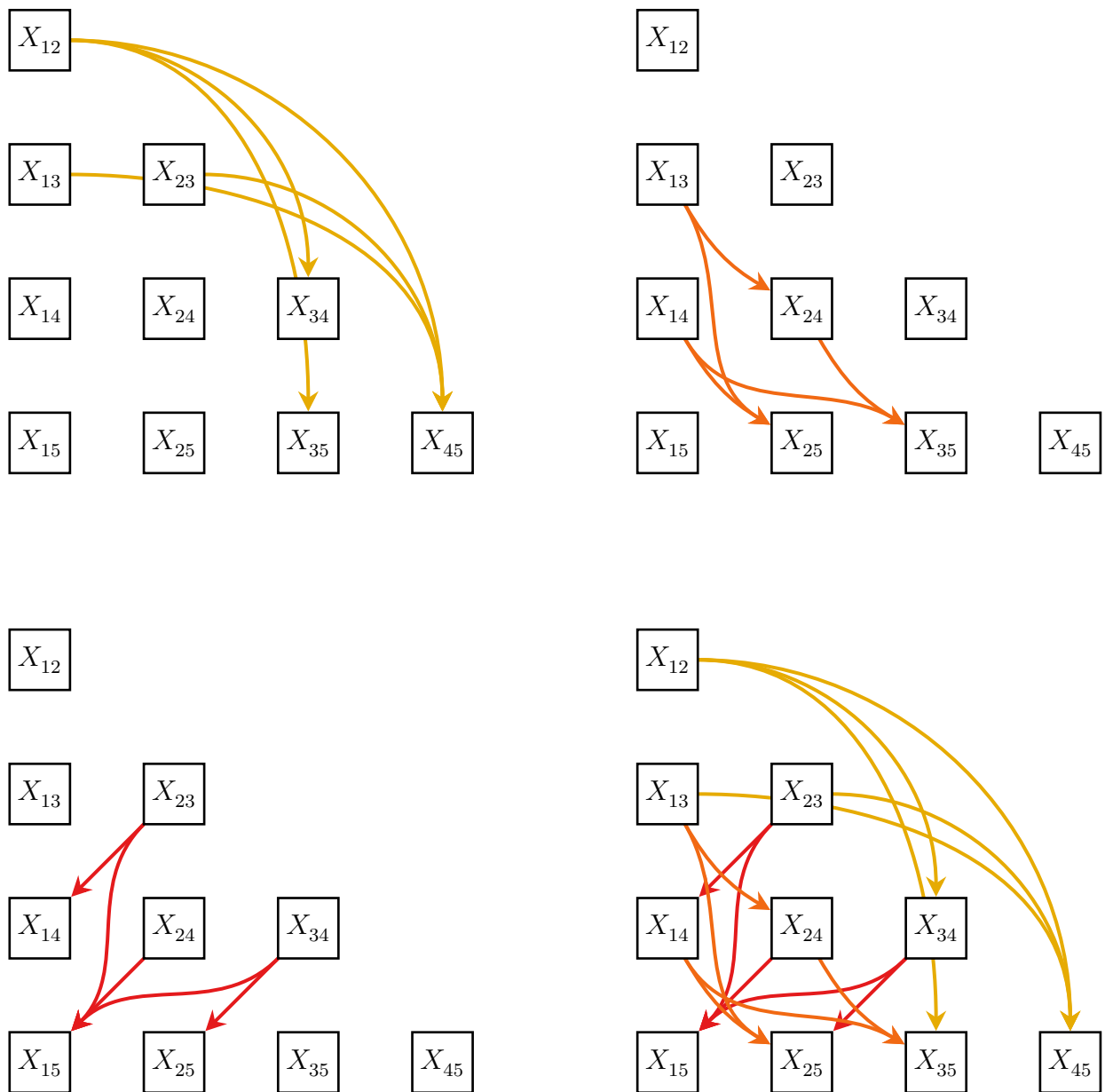
Figure 6: **Causal graphs for each of the types of causal arrows between dyads that do not share nodes.**
Causal meta-DAGs between dyads of a growing network with 5 nodes that are compatible with network models having the following types of causal arrows: FAR (*top-left*); MID (*top-right*); NEAR (*bottom-left*); and the three together (*bottom-right*).
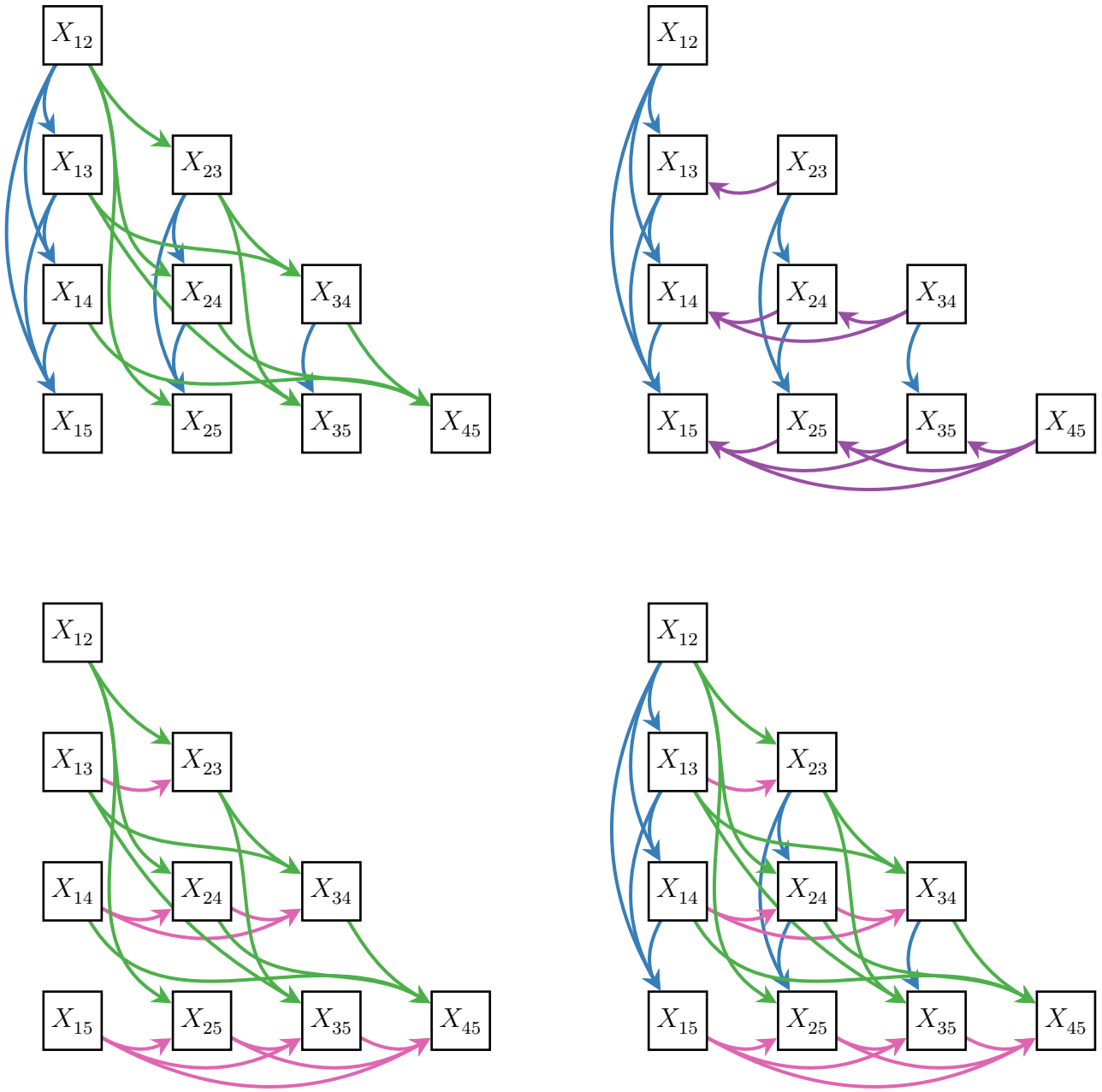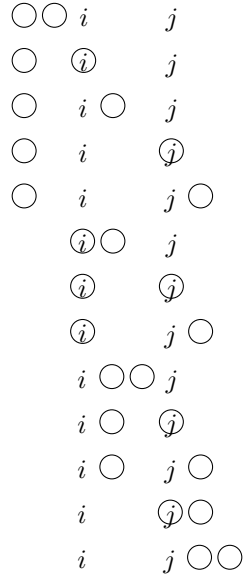
Figure 7: **Causal graphs with multiple types of causal arrows between dyads.**
Causal meta-DAGs between dyads of a growing network with 5 nodes that are compatible with network models having the following types of causal arrows: HUB and PATH (*top-left*); HUB and NEW (*top-right*); PATH and OLD (*bottom- left*); and HUB, PATH, and OLD (*bottom-right*).

B How Relations Relate

With respect to the total ordering of the nodes, there are 13 ways in which two dyads can relate:

$$
\begin{array}{llll}
\bigcirc\bigcirc\; i & & j & \\
\bigcirc & \textcircled{i} & j & \\
\bigcirc & i\;\bigcirc & j & \\
\bigcirc & i & \textcircled{j} & \\
\bigcirc & i & j\;\bigcirc & \\
& \textcircled{i}\bigcirc & j & \\
& \textcircled{i} & \textcircled{j} & \\
& \textcircled{i} & j\;\bigcirc & \\
& i\;\bigcirc\bigcirc\; j & & \\
& i\;\bigcirc & \textcircled{j} & \\
& i\;\bigcirc & j\;\bigcirc & \\
& i & \textcircled{j}\bigcirc & \\
& i & j\;\bigcirc\bigcirc & \\
\end{array}
$$

However, we would like our meta-DAGs to have finite ancestral sets, so it is easy to sample from them. This excludes the five options with circles appearing after j, as they represent causal arrows from infinitely many dyads with indices greater than j, leaving the options displayed in Table 1, which we study throughout the paper.

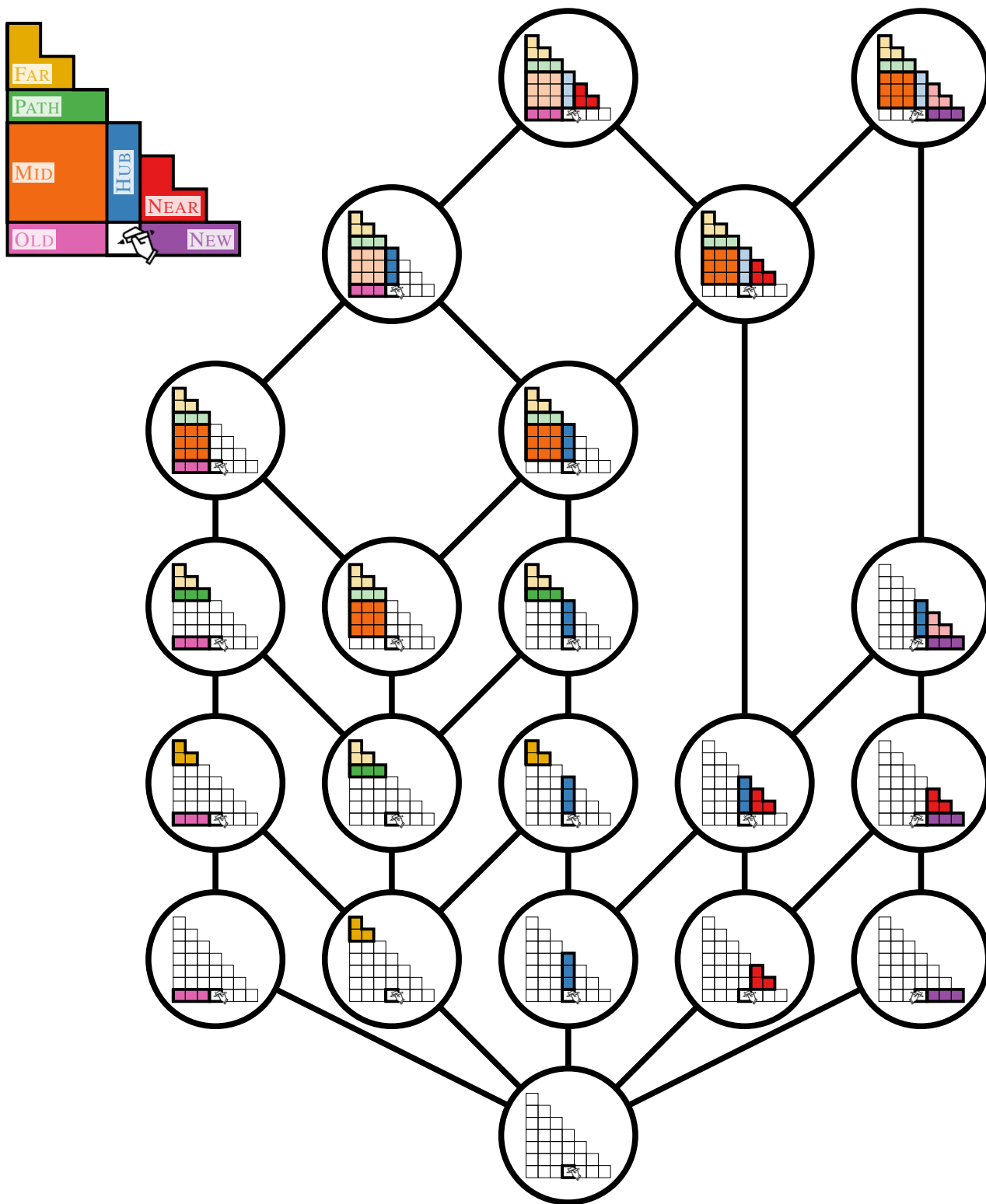# C Hasse Diagram with the 21 Invariant Causal Models

Figure 8: **Hasse diagram of the 21 transitively-closed deletion-invariant causal meta-DAGs with finite ancestral sets.**

D   Proof of Theorems for DAPA model

The Distributed Affine Preferential Attachment (DAPA) model is defined as:

$$x_{ij} \sim \text{Bernoulli}(p_{ij}) \tag{6}$$

$$p_{ij} = \frac{\alpha + \theta_{\text{in}} d_i^{\text{in}} + \theta_{\text{out}} d_i^{\text{out}}}{j - 2 + \alpha + \beta} \tag{7}$$

$$d_i^{\text{in}} = \sum_{\circ = i+1}^{j-1} x_{i\circ} \qquad d_i^{\text{out}} = \sum_{\circ=1}^{i-1} x_{\circ i} \tag{8}$$

where $x_{ij} = 1$ $(x_{ij} = 0)$ indicates an edge (no edge) between nodes i and j.

This model exhibits three sparsity regimes (depending on θ_{in} and θ_{out}) and a flexible power law (depending on θ_{in} for the constant average degree regime and on θ_{out} for the polynomial degree regime).

D.1   Proof of Theorem 3 *(Phase transitions in the sparsity of the DAPA model)*

Our DAPA model exhibits three qualitatively different asymptotic behaviors for the average degree $\langle d(n) \rangle$

$$\text{constant:} \qquad \frac{2\alpha}{1 - (\theta_{\text{in}} + \theta_{\text{out}})} \qquad 0 < \theta_{\text{in}} + \theta_{\text{out}} < 1$$

$$\text{logarithmic:} \qquad 2\alpha \log(n) + C \qquad \theta_{\text{in}} + \theta_{\text{out}} = 1$$

$$\text{polynomial:} \qquad C \times n^{\theta_{\text{in}} + \theta_{\text{out}} - 1} \qquad 1 < \theta_{\text{in}} + \theta_{\text{out}} < 2$$

D.1.1   Constant degree regime (when: $0 \leq \theta_{\text{in}} + \theta_{\text{out}} < 1$)

In this regime, the average degree converges to a constant:

$$\langle d(n) \rangle = \frac{2\alpha}{1 - (\theta_{\text{in}} + \theta_{\text{out}})} + o(1) \tag{9}$$

For a node j, its in-degree d_j^{in} and out-degree d_j^{out} are

$$d_j^{\text{in}} = \sum_{k=j+1}^{n} x_{jk} \qquad \text{and} \qquad d_j^{\text{out}} = \sum_{i=1}^{j-1} x_{ij} \tag{10}$$

After node n has decided all its connections, the total number of edges in the network is

$$E(n) = \sum_{i=1}^{n} d_i^{\text{in}} = \sum_{i=1}^{n} d_i^{\text{out}} \tag{11}$$

The expected number of edges added at each step[7] is

$$\langle E(n+1) - E(n) \rangle = \sum_{i=1}^{n} \frac{\alpha + \theta_{\text{in}} d_i^{\text{in}} + \theta_{\text{out}} d_i^{\text{out}}}{n + \alpha + \beta - 1}$$

$$= \frac{\alpha n + (\theta_{\text{in}} + \theta_{\text{out}}) E(n)}{n + \alpha + \beta - 1} \tag{12}$$

---

[7]To simplify notation, here, we use n for both the step/iteration of the model and the number of nodes.

We make an ansatz of constant average degree

$$\langle E(n) \rangle = C_1 \times n + g(n) \tag{13}$$

where $g(n) = o(n)$ is subdominant.

We will first solve for C_1 to obtain the asymptotic average degree $\langle d \rangle = 2C_1$, then we will verify our assumption that $g(n) = o(n)$.

Taking the expectation of (12),

$$\langle E(n+1) \rangle - \langle E(n) \rangle = \frac{\alpha n + (\theta_{\text{in}} + \theta_{\text{out}})\langle E(n) \rangle}{n + \alpha + \beta - 1} \tag{14}$$

using our ansatz (13),

$$g(n+1) - g(n) + \underbrace{C_1 = \left(\alpha + (\theta_{\text{in}} + \theta_{\text{out}})\left(C_1 + \frac{g(n)}{n} \right) \right)}_{\text{equate these constant terms}} \times \frac{n}{n + \alpha + \beta - 1} \tag{15}$$

and equating the $\Theta(1)$ (constant) terms to solve for C_1,

$$C_1 = \alpha + (\theta_{\text{in}} + \theta_{\text{out}})C_1 \tag{16}$$

$$= \frac{\alpha}{1 - (\theta_{\text{in}} + \theta_{\text{out}})} \tag{17}$$

we obtain the asymptotic average degree claimed in (9).

To check verify our ansatz $g(n) = o(n)$, substitute (17) into (15)

$$\underbrace{g(n+1) - g(n) = \frac{(\theta_{\text{in}} + \theta_{\text{out}})g(n)}{n}}_{\text{equate these dominant terms}} - \frac{\alpha + \beta - 1}{n} + \mathcal{O}\left(\frac{g(n)}{n^2} \right) \tag{18}$$

to conclude that $g(n) = \mathcal{O}(n^{\theta_{\text{in}} + \theta_{\text{out}}})$.

So our ansatz $g(n) = o(n)$ is valid for $\theta_{\text{in}} + \theta_{\text{out}} < 1$, which incidentally is precisely when equation (9) is physically meaningful.

D.1.2 Logarithmic degree regime (when: $\theta_{\text{in}} + \theta_{\text{out}} = 1$)

As $\theta_{\text{in}} + \theta_{\text{out}} \to 1$, the average degree predicted by (9) diverges, and when $\theta_{\text{in}} + \theta_{\text{out}} = 1$, the average degree is no longer bounded. Instead, it grows logarithmically in n:

$$\langle d(n) \rangle = 2\alpha \log(n) + C + o(1) \tag{19}$$

Similar to equation (13), we make an ansatz but now of logarithmic average degree

$$\langle E(n) \rangle = C_1 \times n \log n + g(n) \tag{20}$$

where $g(n) = o(n \log n)$ is subdominant.

The change in $n \log n$ is approximately

$$(n+1)\log(n+1) - n\log(n) = \log(n) + 1 + \mathcal{O}\left(\frac{1}{n^2} \right). \tag{21}$$

Substituting this into equation (14),

$$g(n+1) - g(n) + \underbrace{C_1 \times \left(\log(n) + 1 \right) = \left(\alpha + (\theta_{\text{in}} + \theta_{\text{out}})\left(C_1 \log(n) + \frac{g(n)}{n} \right) \right)}_{\text{equate these logarithmic and constant terms}} \times \left(1 + \mathcal{O}\left(\frac{1}{n} \right) \right) \tag{22}$$

Equating the logarithmic terms, we recover the condition that $\theta_{\text{in}} + \theta_{\text{out}} = 1$. And equating the constant terms, we can solve for the constant $C_1 = \alpha$.

To verify our ansatz $g(n) = o(n \log n)$, we consider the lower-order terms:

$$g(n+1) - g(n) = \frac{g(n)}{n} + \mathcal{O}\left(\frac{\log n}{n}\right) \tag{23}$$

The solution can be written as $g(n) = C \times n + f(n)$, where $f(n) = o(n)$.

Note that C is not determined by the asymptotic balance. This suggests that different instances of network generated with the same parameters may limit to different values of C. This is similar to the situation with the classic Pólya urn [Eggenberger and Pólya, 1923], where any asymptotic ratio is equally stable, resulting in a distribution over these ratios for any fixed set of parameters [Mahmoud, 2008, Peköz et al., 2029].

D.1.3   Sub-linear degree regime (when: $1 < \theta_{\text{in}} + \theta_{\text{out}} < 2$)

When $1 < \theta_{\text{in}} + \theta_{\text{out}} < 2$, the average degree grows as a (sublinear) power of n:

$$\langle d(n) \rangle \propto C n^\rho, \qquad \rho = \theta_{\text{in}} + \theta_{\text{out}} - 1 \tag{24}$$

Again, we make the appropriate ansatz:

$$\langle E(n) \rangle = C_1 \times n^{1+\rho} + g(n) \tag{25}$$

where $g(n) = o(n^{1+\rho})$ is subdominant. The change in $n^{1+\rho}$ is approximately

$$(n+1)^{1+\rho} - n^{1+\rho} = (1+\rho)n^\rho + \mathcal{O}\left(n^{\rho-1}\right). \tag{26}$$

Substituting this into equation (14),

$$g(n+1) - g(n) + \underbrace{C_1 \times (1+\rho)n^\rho = \left(\alpha + (\theta_{\text{in}} + \theta_{\text{out}})\left(C_1 n^\rho + \frac{g(n)}{n}\right)\right)}_{\text{equate these } \mathcal{O}(n^\rho) \text{ terms}} \times \left(1 + \mathcal{O}\left(\tfrac{1}{n}\right)\right) \tag{27}$$

and equating the dominant $\mathcal{O}\left(\frac{1}{n}\right)$ terms, we can solve for the exponent $\rho = \theta_{\text{in}} + \theta_{\text{out}} - 1$.

Note that this does not fix C_1. To see why, let us attempt to verify our ansatz $g(n) = o(n \log n)$. We again equate the remaining lower-order terms:

$$g(n+1) - g(n) = (1+\rho)\frac{g(n)}{n} + \mathcal{O}\left(n^{\rho-1}\right) \tag{28}$$

In this case, it seems as though our ansatz is not verified, with $g(n)$ being the same order as the "dominant" part of the solution: $g(n) = C_2 \times n^{1+\rho} + o(n^{1+\rho})$. However, this is in fact not a contradiction — this is the asymptotic analysis telling us that the original constant C itself is not determined.

Recapitulating the sequence of results:

- When $\theta_{\text{in}} + \theta_{\text{out}} < 1$, the average degree asymptotes to a fixed constant.
- When $\theta_{\text{in}} + \theta_{\text{out}} = 1$, the average degree grows logarithmically in n, but with an arbitrary additive constant.
- When $\theta_{\text{in}} + \theta_{\text{out}} > 1$, the average degree grows as a sublinear power of n, with an arbitrary multiplicative constant.

$\square$

D.2 Proof of Theorem 4 *(Phase transitions in the power-law of the degree distribution of the DAPA model)*

In the DAPA model, the asymptotic distribution of node degrees exhibits a power-law tail $p(d) \propto d^{-\gamma}$, where the scaling exponent depends on either θ_{in} or θ_{out}:

$$\text{constant:} \qquad \gamma = \frac{1 + \theta_{\text{in}}}{\theta_{\text{in}}} \qquad\qquad 0 < \theta_{\text{in}} + \theta_{\text{out}} \leq 1 \tag{29}$$

$$\text{polynomial:} \qquad \gamma = \frac{2 - \theta_{\text{out}}}{1 - \theta_{\text{out}}} \qquad\qquad 1 \leq \theta_{\text{in}} + \theta_{\text{out}} < 2 \tag{30}$$

To show this, we will first characterize the out-degrees $d_j^{\text{out}} = \sum_{i=1}^{j-1} x_{ij}$ (connections that a node j makes with previous nodes). These out-degrees (and the node arrival time j) serve as the initial conditions for the growth of the in-degrees $d_j^{\text{in}} = \sum_{k=j+1}^{n} x_{jk}$ (connections that node j makes with later nodes).

D.2.1 The out-degrees (connections with previous nodes)

The out-degrees do *not* exhibit a power law in this model. This is because the outcomes of these x_{ij} ($1 \leq i < j$) are conditionally independent given the previous entries x_{ab} ($1 \leq a < b < j$), and the sum of independent Bernoulli variables does not exhibit a power-law distribution.

Denote these (conditionally) independent probabilities as p_{ij}. In the proof of the sparsity of the model in Appendix D.1, we used the fact that the sum of these probabilities $\sum_{i=1}^{j-1} p_{ij}$ is the expected change in edges at each step j. Moreover, as the network grows, this sum does not change much between each step. Thus, the expected out-degree $d_j^{\text{out}} = \sum_{i=1}^{j-1} x_{ij}$ of node j is approximately

$$\text{constant:} \qquad \langle d_j^{\text{out}} \rangle \sim \frac{\alpha}{1 - (\theta_{\text{in}} + \theta_{\text{out}})} \qquad\qquad 0 < \theta_{\text{in}} + \theta_{\text{out}} < 1 \tag{31}$$

$$\text{logarithmic:} \qquad \langle d_j^{\text{out}} \rangle \sim \alpha \log(j) + C \qquad\qquad \theta_{\text{in}} + \theta_{\text{out}} = 1 \tag{32}$$

$$\text{polynomial:} \qquad \langle d_j^{\text{out}} \rangle \sim C \times j^{\theta_{\text{in}} + \theta_{\text{out}} - 1} \qquad\qquad 1 < \theta_{\text{in}} + \theta_{\text{out}} < 2 \tag{33}$$

with a variance upper-bounded by this average (as they are the sum of independent Bernoulli random variables).

D.2.2 The in-degrees (connections with later nodes)

For a node j with a given out-degree d_j^{out}, the expected in-degree grows according to the difference equation:

$$\langle d_j^{\text{in}} \rangle (n+1) - \langle d_j^{\text{in}} \rangle (n) = \frac{\alpha + \theta_{\text{out}} d_j^{\text{out}} + \theta_{\text{in}} \langle d_j^{\text{in}} \rangle (n)}{n + \alpha + \beta - 1} \qquad \langle d_j^{\text{in}} \rangle (j) = 0 \tag{34}$$

As this evolution does not depend on the outcome of any other edges in the network, we can write the solution in closed form:

$$\langle d_j^{\text{in}} \rangle (n) = \frac{\alpha + \theta_{\text{out}} d_j^{\text{out}}}{\theta_{\text{in}}} \left(\frac{\Gamma(\alpha + \beta + j - 1)}{\Gamma(\alpha + \beta + \theta_{\text{in}} + j - 1)} \frac{\Gamma(\alpha + \beta + \theta_{\text{in}} + n - 1)}{\Gamma(\alpha + \beta + n - 1)} - 1 \right) \tag{35}$$

For $1 \ll j < n$, the ratios of gamma functions can be approximated, and the expected degree as a function of j and n is

$$\langle d_j \rangle (n) \approx \frac{\alpha + \theta_{\text{out}} \langle d_j^{\text{out}} \rangle}{\theta_{\text{in}}} \left(\left(\frac{n}{j} \right)^{\theta_{\text{in}}} - 1 \right) + \langle d_j^{\text{out}} \rangle \tag{36}$$

We can extract the power law of the degree distribution from the dependence of the expected degree on the node index j. Since $\langle d_j \rangle$ is monotonically decreasing in j, the probability density will be proportional to the reciprocal of the magnitude

of the derivative with respect to j:

$$p(d) \propto \left| \frac{d}{dj} \langle d_j \rangle \right|^{-1} \tag{37}$$

For a power-law degree distribution $p(d) \propto d^{-\gamma}$, the exponent is the change in $\log(p(d))$ with respect to $\log(d)$:

$$\gamma = \frac{\frac{d}{dj} \log \left| \frac{d}{dj} \langle d_j \rangle \right|}{\frac{d}{dj} \log \langle d_j \rangle} = \frac{\langle d_j \rangle \frac{d^2}{dj^2} \langle d_j \rangle}{\left(\frac{d}{dj} \langle d_j \rangle \right)^2} \tag{38}$$

Substitute the expressions for $\langle d_j^{\text{out}} \rangle$ from equations (31) and (33) into equation (36). When $1 \ll j \ll n$, the dominant term has the following scalings:

$$\text{constant:} \qquad \langle d_j \rangle(n) \propto n^{\theta_{\text{in}}} j^{-\theta_{\text{in}}} \qquad\qquad 0 < \theta_{\text{in}} + \theta_{\text{out}} < 1 \tag{39}$$

$$\text{polynomial:} \qquad \langle d_j \rangle(n) \propto n^{\theta_{\text{in}}} j^{\theta_{\text{out}}-1} \qquad\qquad 1 < \theta_{\text{in}} + \theta_{\text{out}} < 2 \tag{40}$$

Substituting these into equation (38), we obtain the following power-law exponents:

$$\text{constant:} \qquad \gamma = \frac{1 + \theta_{\text{in}}}{\theta_{\text{in}}} \qquad\qquad 0 < \theta_{\text{in}} + \theta_{\text{out}} \le 1 \tag{41}$$

$$\text{polynomial:} \qquad \gamma = \frac{2 - \theta_{\text{out}}}{1 - \theta_{\text{out}}} \qquad\qquad 1 \le \theta_{\text{in}} + \theta_{\text{out}} < 2 \tag{42}$$

Notice that both expressions give the same scaling when $\theta_{\text{in}} + \theta_{\text{out}} = 1$ for the intermediate logarithmic regime.

$\square$