Causal Models for Growing Networks

Gecia Bravo-Hermsdorff[1]   **Kayvan Sadeghi**[2]   **Lee M. Gunderson**[2]

[1]School of Informatics, University of Edinburgh, Edinburgh, Scotland, UK
[2]Department of Statistical Science, University College London, London, England, 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.

We highlight one model in particular, which we call *Paral-*

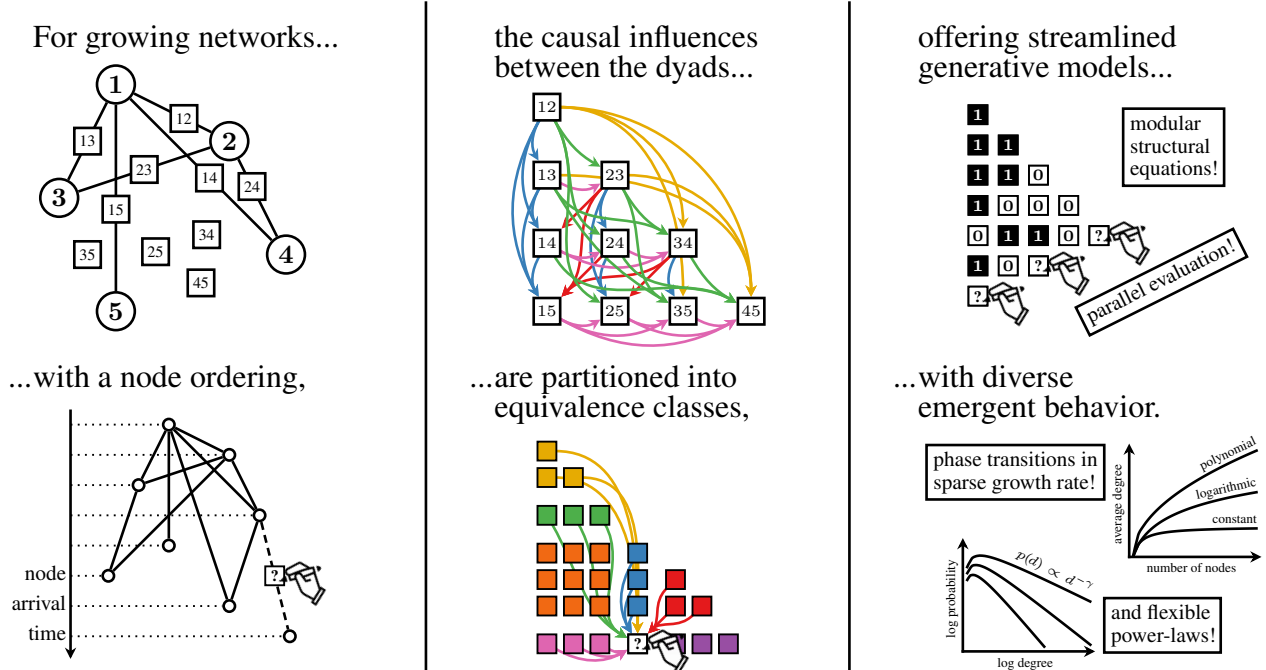[1]Such as the ants and their Aunt Hillary [Hofstadter, 1999].

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 types 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 Parallelized Affine Preferential Attachment (PAPA). 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).

lelized Affine Preferential Attachment (*PAPA*), 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) 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 distribu-

tions 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:

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

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.

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:

$\bigcirc$ i	j		"distant" nodes	$\bigcirc < i < j$
$\textcircled{i}$	j		"past" node	$\bigcirc = i < j$
i $\bigcirc$ j			"recent" nodes	$i < \bigcirc < j$
i	$\textcircled{j}$		"current" node	$i < \bigcirc = j$
i	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. 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.

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.

---

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

ⓘ ○		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 types 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]

2.4.1 Arrows Between Dyads Sharing a Node

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

The HUB causal arrows mediate the tendency to reference a

---

[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."*

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 types of causal arrows between them.

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 5.1, 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 types 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.

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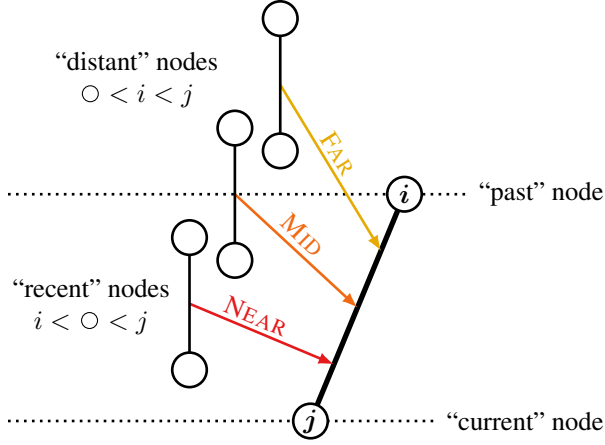
Figure 3: **Types of causal arrows between dyads that do not share a node.** Horizontal dashed lines indicate the relative ordering of the nodes in the parent dyad ($\bigcirc\bigcirc$) with respect to the nodes in the child dyad (ij).

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 types of 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 meta-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 [Lauritzen, 1996]. 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 B 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 meta-DAG contains the arrows $(ab)\longrightarrow(ij)$ and $(ij) \longrightarrow (rs)$, then it must also include the arrow $(ab)\longrightarrow(rs)$. 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.

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 B 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. Their partial ordering is shown in Figure 8.

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: $\xrightarrow{\text{PATH}}\xrightarrow{\text{PATH}} = \{\xrightarrow{\text{FAR}}\}$, 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: $\xrightarrow{\text{HUB}}\xrightarrow{\text{OLD}} = \{\xrightarrow{\text{MID}}, \xrightarrow{\text{PATH}}, \xrightarrow{\text{FAR}}\}$. Technically, this composition depends on the order: $\xrightarrow{\text{OLD}}\xrightarrow{\text{HUB}} = \{\xrightarrow{\text{MID}}\}$, but this non-commutativity 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, $\{\xrightarrow{\text{SELF}}, \xrightarrow{\text{HUB}}, \xrightarrow{\text{PATH}}\}$ composed with itself results in $\{\xrightarrow{\text{SELF}}, \xrightarrow{\text{HUB}}, \xrightarrow{\text{PATH}}, \xrightarrow{\text{FAR}}\}$. Composing $\{\xrightarrow{\text{SELF}}, \xrightarrow{\text{HUB}}, \xrightarrow{\text{PATH}}, \xrightarrow{\text{FAR}}\}$ 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 easily be found by hand, this quickly become unwieldy for generalizations of the procedure.[4]

---

[4]For example, for simple hypergraphs with cardinality 3, there

3 STREAMLINING 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 depends on the position of the child dyad variable, such a function must allow for an arbitrary number of inputs. A natural choice is to define the function in terms of summary statistics of the different types of causal parents.

In Section 3.2, we present a simple model for binary (edge or no-edge) dyad variables that implements preferential attachment using the HUB and PATH causal arrows. 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 already been described by several authors [Eggenberger and Pólya, 1923, Simon, 1955, Price, 1965], the effect itself is perhaps best exemplified by the overwhelming number of citations garnered by Barabási and Albert [1999].

The simplest statement of their model [Pósfai and Barabási, 2016] has a single parameter, m. Initialize the network with clique of m nodes. At each iteration, select m nodes proportional to their current degree, and add a new node connected to each of these selected nodes.

Many extensions to this model have been described [Ray, 2024]: 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]. Here, we draw attention to a particular line of modifications that reduce the statistical correlations between pairs of edges [Bollobás et al., 2007, Wang and Resnick, 2020].

---

are already 37 type of causal arrows between the triads (subsets of 3 nodes) with finite ancestral sets.

3.1.1 Poissonified Preferential Attachment

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 ($\bigcirc j$) 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] This does not qualitatively change the asymptotic behavior; the degree distribution has the same power-law tail $p(d) \propto d^{-3}$ and average degree $\langle d \rangle = 2m$ as before.

By allowing the number of edges to be an implicit random variable, this modification renders the dyad variables ($\bigcirc j$) conditionally independent given the current degrees of the previous nodes. However, 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 dyad variables necessary? How much dependence can we remove while still retaining the hallmark features of preferential attachment models?

3.2 PARALLELIZED AFFINE PREFERENTIAL ATTACHMENT

Motivated by this question, we note that only two of the seven types of causal arrows are essentially being asked for by preferential attachment: HUB and PATH, and that their transitive closure implies significantly less dependence on previous dyad variables (see Fig. 8). As summary statistics for the structural equation, we use node degrees of the older node: d_i^{in} corresponds to HUB, and d_i^{out} corresponds to PATH. We take the edge probabilities to be an affine function of these statistics, leading to the model we call *Parallelized Affine Preferential Attachment*, or *PAPA*, as sampling from it is highly parallelizable (see discussion in Section 5). Explicitly:

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

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

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

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 *increased diversity* in its asymptotic behavior (see Fig. 4, Theorems 3 and 4, proofs in Appendix C).

---

[5]Setting aside cases with probabilities greater than 1.
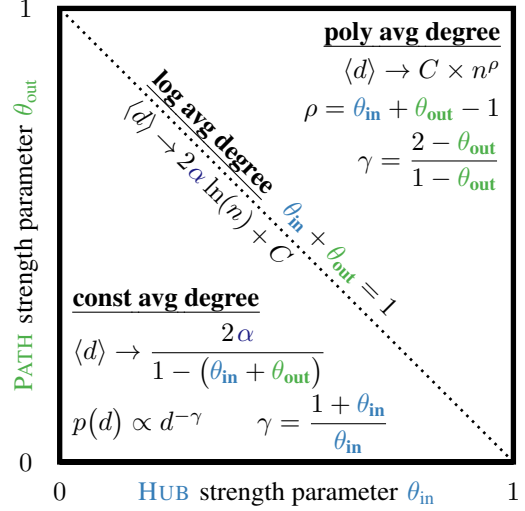
Figure 4: **Sparsity and power-laws in the PAPA model.**

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 PAPA 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 average degree of the PAPA model*). The PAPA 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 degree distributions of the growing networks have a range of power-law scalings.

Theorem 4 (*Power-law degree distributions of the PAPA 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}.

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

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

Note that these two expressions result in the same scaling exponent precisely when $0 < \theta_{\text{in}} + \theta_{\text{out}} \leq 1$.

4 SOME APPLICATIONS

In this section, we discuss how having causal DAGs as generative models for the networks leads to two advantageous features of our framework: ease of sampling and inference (Section 4.1), and ability to answer interventional and counterfactual causal queries (Section 2.5).

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 predict the the asymptotic behavior of the growing network, despite those features not yet being present.

4.2 INTERVENTIONS AND COUNTERFACTUALS

We now 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 more citations to your bibliography to help it reach a larger audience. To estimate the net effect of this strategy, one could fit the parameters of a causal model (such as the PAPA model or its extensions) to the current state of the citation network relevant to you.

---

[6]Estimating the power-law exponent of a degree distribution is notoriously tricky [Clauset et al., 2009]. 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 necessarily deviations at the extremities.

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 an 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 you published it. 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 the net effect of such "could/should have been" actions, one can use the structural equations of a causal model. For example, the randomness in Equation (1) of the PAPA model can be represented explicitly by introducing an (unobserved) random variable:

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

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].

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 DAGs that were less rigid in the order of their generation.

In particular, causal meta-DAGs in which the dyads only depend on one of the two "quadrants" of past dyads—such as the "PAPA w/ clustering" or "bottom-up causality" in Fig. 8—can be evaluated in a distributed manner (as illustrated in Fig. 1). For these models, coarse-graining the rows and columns of the grid of dyad variables results in blocks of dyads 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.

PAPA model with clustering — HUB + PATH + OLD. As an extension of the PAPA model, one could include OLD causal arrows (see bottom-right of Fig. 7 for its meta-DAG.). This addition would, for example, allow for the in-degrees to also exhibit a power-law (as dyads of an income node would no longer be independent). Moreover, when both

OLD and PATH causal 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 $\circ < i < j$ can influence the likelihood that i and j themselves form a connection.

A "bottom-up" causality — HUB + NEW. The PAPA model and its extension including OLD have a "top-down" 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 "bottom-up" causal structure (see top-right of Fig. 7 for its meta-DAG). 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.

5.1 SPARSE COMPOSABLE STRUCTURAL EQUATIONS

As previously mentioned (Section 2.4.1), the OLD and NEW causal arrows cannot be included in the same causal meta-DAG. Here we propose a way to do essentially that.

The choice of an affine structural equation in the PAPA model was motivated in part by its similarity to other classical growing models, such as Pólya's urn [Eggenberger and Pólya, 1923, Mahmoud, 2008], the process of Pitman-Yor [Pitman and Yor, 1997], and various Canonical Restaurant processes [Aldous et al., 2006, Ghahramani and Griffiths, 2005].

Here is another option for Equation (2), with ellipsis (...) to suggest its straightforward generalization:

$$p_{ij} = 1 - \exp\left(- \frac{\alpha + \theta_{\text{in}} d_i^{\text{in}} + \theta_{\text{out}} d_i^{\text{out}} + \cdots}{j + \beta} \right) \quad (6)$$

This simple transformation of the affine model retains all the asymptotics of the original PAPA model (since $1 - \exp(-p_{ij}) \approx p_{ij}$ for $p_{ij} \ll 1$). But now the inclusion of additional terms such as $\theta_{\text{old}} d_j^{\text{old}}$ predictably increase the probability of an edge (without becoming greater than 1). Also, different choices for the denominators might allow for better modeling of growing networks over a wide range of scales.

However, the most compelling property of a model using Equation (6) is that it allows one to effectively sample from causal models that apparently have cycles! Moreover, the algorithm is naturally asynchronous, and exploits the sparsity of the resulting network.

Here is a sketch of the algorithm for such a model. Initialize all dyads in the network as "empty". Add an edge independently for each dyad variable with probability $\exp(-\frac{\alpha}{j+\beta})$,

and set those dyads as "active". Iteratively take an active dyad, set it as "completed", and sample its HUB children independently with probability $\exp(-\frac{\theta_{\text{in}}}{j+\beta})$, and likewise for PATH children. For any sampled children dyads that are currently "empty", add that edge to the network and set those dyads as "active". Continue until there are no "active" dyads.

This perspective effectively allows for directed cycles between the dyads, essentially "fine-graining" them into a series of back-and-forth communication, while the acyclic transitions ("empty" → "active" → "completed") ensure that the algorithm will terminate for a finite network. Modeling the behavior of complex interconnected networks in this modular "event-based" manner is a promising 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.

David J Aldous, Illdar A Ibragimov, and Jean Jacod. Ecole d'Ete de Probabilites de Saint-Flour XIII, 1983, volume 1117. Springer, 2006.

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.

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.

Zoubin Ghahramani and Thomas Griffiths. Infinite latent feature models and the indian buffet process. Advances in neural information processing systems, 18, 2005.

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.

Jim Pitman and Marc Yor. The two-parameter poisson-dirichlet distribution derived from a stable subordinator. The Annals of Probability, pages 855–900, 1997.

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 preferential attachment model with poisson measurement. arXiv preprint arXiv:2008.07005, 2020.

Causal Models for Growing Networks

Gecia Bravo-Hermsdorff[1] **Kayvan Sadeghi**[2] **Lee M. Gunderson**[2]

[1]School of Informatics, University of Edinburgh, Edinburgh, Scotland, UK
[2]Department of Statistical Science, University College London, London, England, UK

A EXAMPLES OF CAUSAL META-DAGS

In this section, we display the causal meta-DAGs associated with the various types of causal arrows (defined in Table 1) 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.
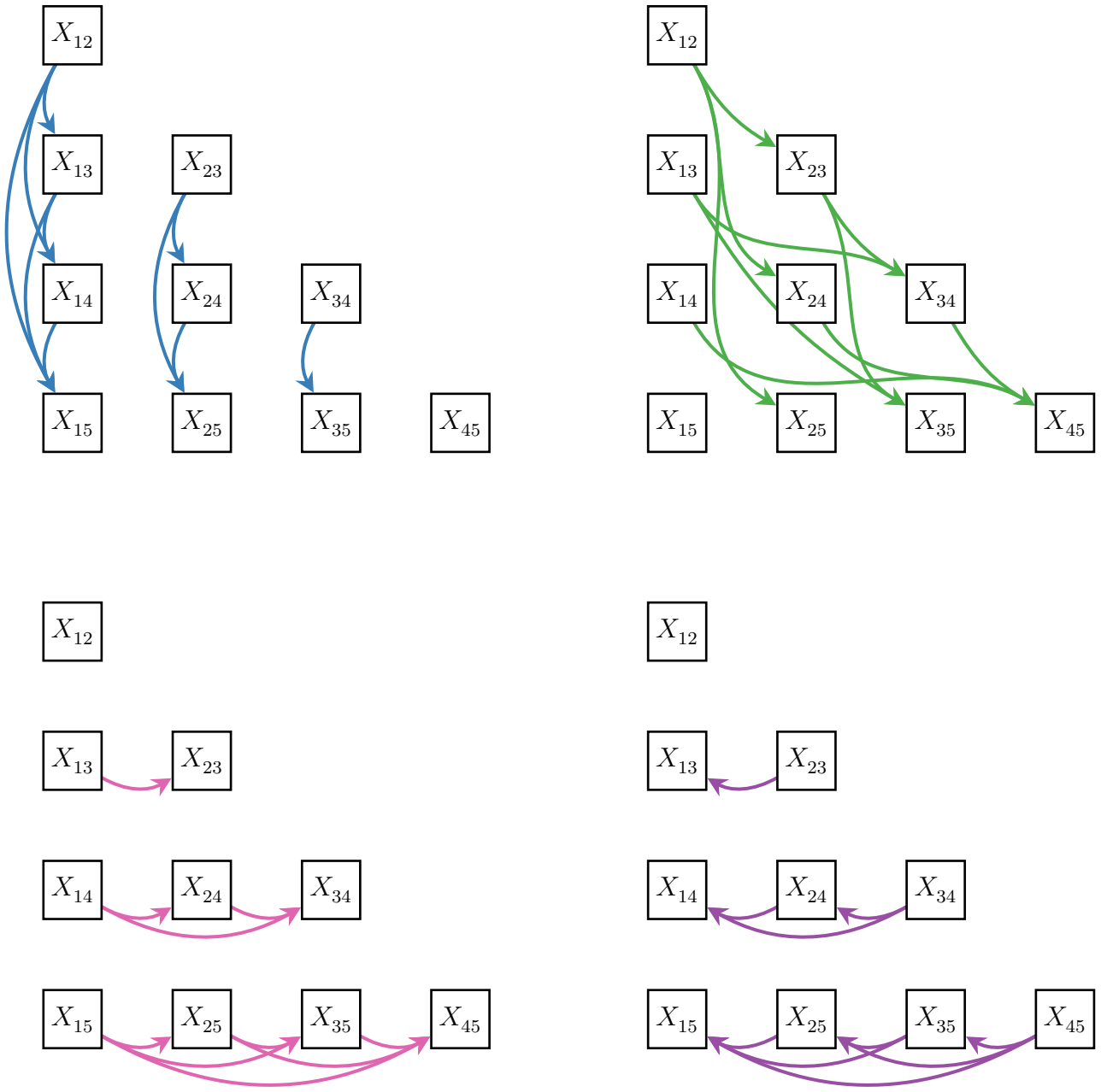
Figure 5: **Causal graphs for each of the types of causal arrows between dyads that share a node.**
Causal meta-DAGs between dyads of a growing network with 5 nodes that are compatible with network models having the following types of causal arrows: HUB (*top-left*); PATH (*top-right*); OLD (*bottom-left*); and NEW (*bottom-right*).
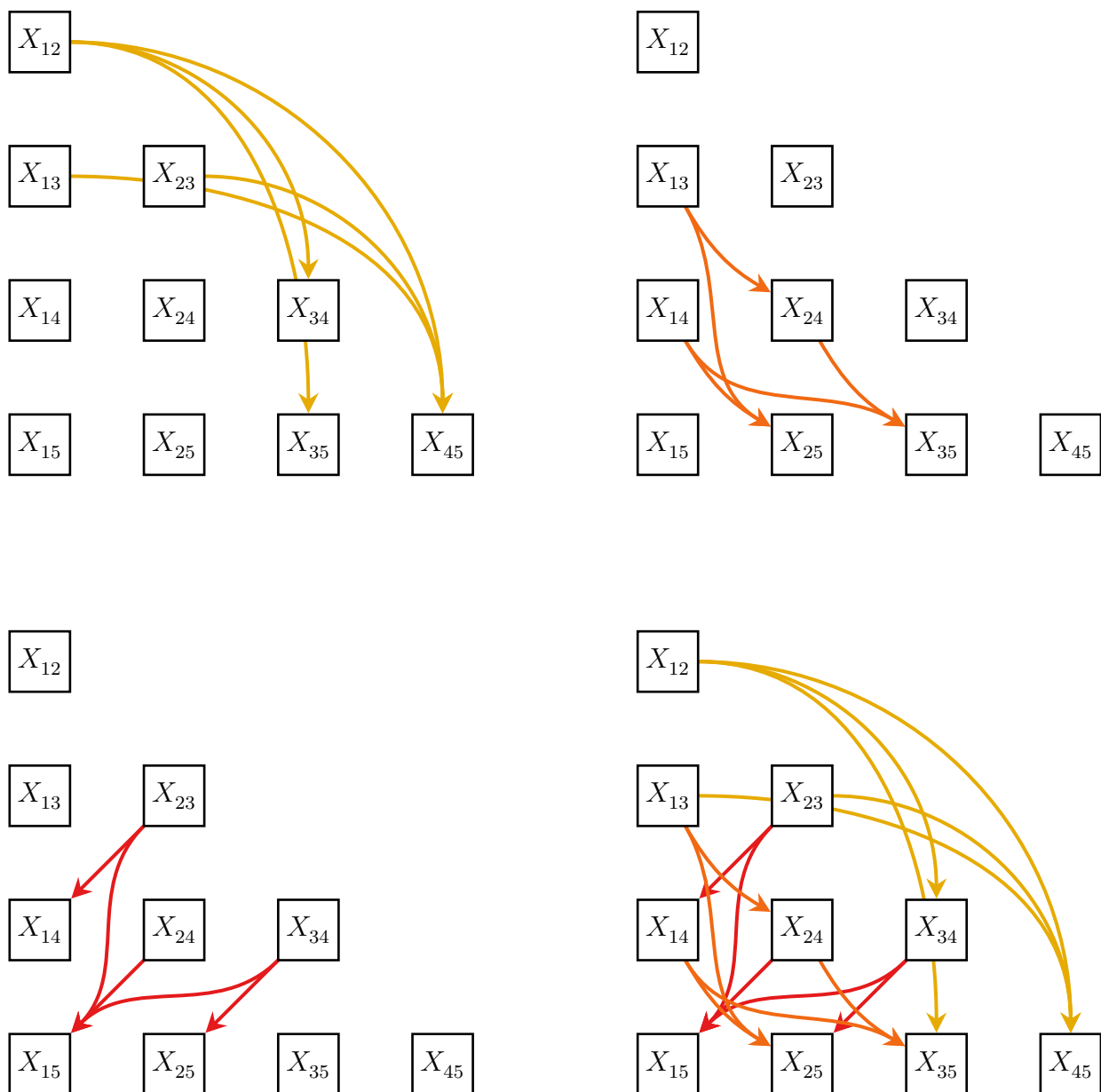
Figure 6: **Causal graphs for each of the types of causal arrows between dyads that do not share nodes.**
Causal meta-DAGs between dyads of a growing network with 5 nodes that are compatible with network models having the following types of causal arrows: FAR (*top-left*); MID (*top-right*); NEAR (*bottom-left*); and the three together (*bottom-right*).
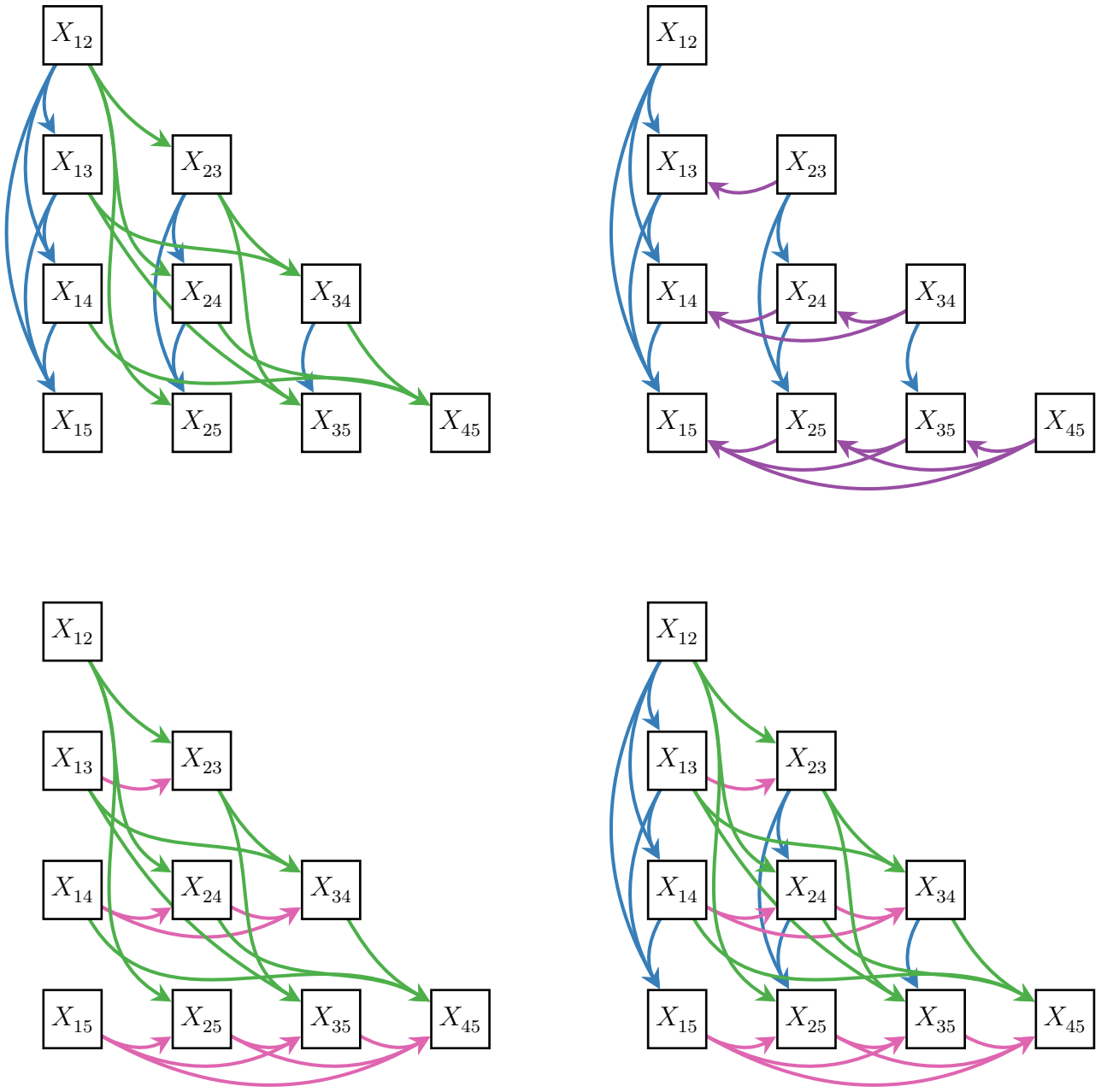
Figure 7: **Causal graphs with multiple types of causal arrows between dyads.**
Causal meta-DAGs between dyads of a growing network with 5 nodes that are compatible with network models having the following types of causal arrows: HUB and PATH (*top-left*); HUB and NEW (*top-right*); PATH and OLD (*bottom-left*); and HUB, PATH, and OLD (*bottom-right*).

B   HASSE DIAGRAM WITH THE 21 INVARIANT CAUSAL MODELS

There are 21 causal meta-DAGs with finite ancestral sets whose interventional structure are invariant to node deletion and contraction (Theorem 2, Sections 2.5 and 2.6). They form a partially ordered set (poset), with the partial order relation given by inclusion of the seven types causal arrows.

Figure 8 display a Hasse diagram of their poset, starting from the bottom of the figure and ordering the columns from left to right, the type of causal arrows that define each of the 21 causal meta-DAGs:

- *Bottom row:*
    - no causal arrows.
- *2ⁿᵈ row:*
    - OLD
    - FAR
    - HUB
    - NEAR
    - NEW
- *3ʳᵈ row:*
    - FAR/OLD
    - PATH (FAR)
    - FAR/HUB
    - HUB/NEAR
    - NEAR/NEW
- *4ᵗʰ row:*
    - OLD/PATH (FAR)
    - MID (PATH, FAR)
    - HUB/PATH (FAR)
    - HUB/NEW (NEAR)
- *5ᵗʰ row:*
    - OLD/MID (PATH/FAR)
    - MID/HUB (PATH/FAR)
- *6ᵗʰ row:*
    - OLD/HUB (MID/PATH/FAR)
    - MID/NEAR (PATH/FAR/HUB)
- *Top row:*
    - OLD/NEAR (MID/PATH/FAR/HUB)
    - MID/NEW (PATH/FAR/HUB/NEAR)

where a type of causal arrow in parenthesis indicates that it is implied by the addition of the new type of causal arrow.
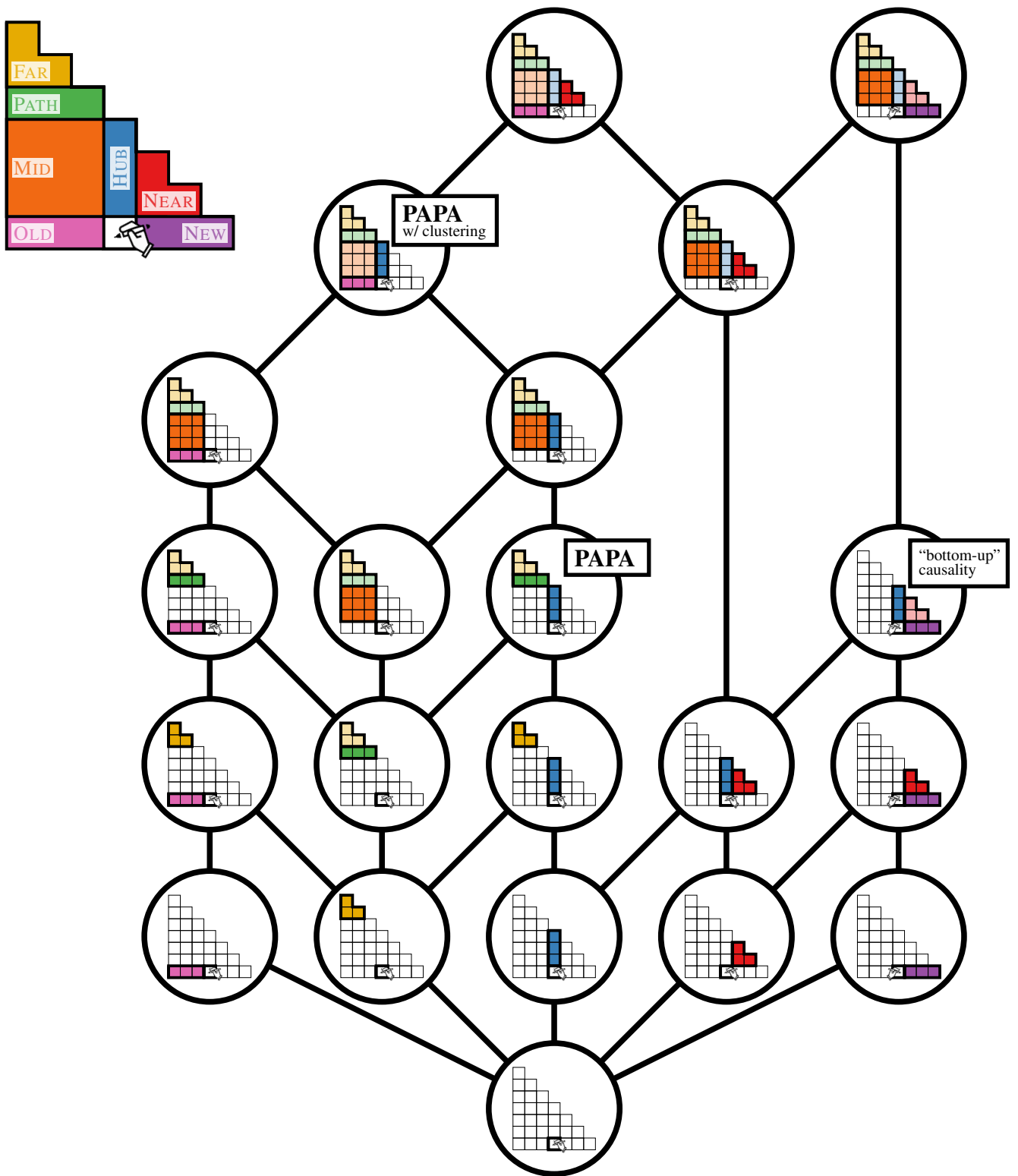
Figure 8: **Hasse diagram of the 21 transitively-closed deletion-invariant causal meta-DAGs with finite ancestral sets.**

C PROOF OF THEOREMS FOR PAPA MODEL

The Distributed Affine Preferential Attachment (PAPA) model (Section 3.2) is defined as:

$$x_{ij} \sim \text{Bernoulli}(p_{ij}) \tag{7}$$

$$p_{ij} = \frac{\alpha + \theta_{\text{in}} d_i^{\text{in}} + \theta_{\text{out}} d_i^{\text{out}}}{j - 2 + \alpha + \beta} \tag{8}$$

$$d_i^{\text{in}} = \sum_{\circ=i+1}^{j-1} x_{i\circ} \qquad d_i^{\text{out}} = \sum_{\circ=1}^{i-1} x_{\circ i} \tag{9}$$

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

The parameters θ_{in} and θ_{out} correspond to HUB and PATH causal arrows, respectively. If one uses the ordering of the nodes to orient the edges of the growing network, HUB arrows refer to the in-degree of the node $\circ$ of the dyad $(\circ j)$ being decided (i.e., dyads of the type $(\circ i)$, where $\circ < i$). Whereas PATH arrows refer to the out-degree of the node $\circ$ (i.e., dyads of the type $(i\circ)$, where $i < \circ$). However, orienting the edges is not necessary, and the final statements of our theorems refer to the undirected degrees of the nodes.

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). For the logarithmic average degree regime, which is at the transition between the constant and polynomial regimes, the dominant dependency is on the parameter α. And the expression for the power-law exponent in this transitional regime agrees with the expression for both the constant and polynomial regimes.

C.1 PROOF OF THEOREM 3 (*Phase transitions in the asymptotic growth rate of the average degree of the PAPA model*)

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

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

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

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

We now derive the expressions for the average degree in the three regimes, starting with the constant regime. To simplify notation on our derivations, we will use n for both the step/iteration of the model and the number of nodes. (Indeed, although this model is parallelizable, one can always simulate it sequentially with a single node arriving at each step.)

C.1.1 Constant degree regime: $0 \leq \theta_{\text{in}} + \theta_{\text{out}} < 1$

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

$$\langle d(n) \rangle = \frac{2\alpha}{1 - (\theta_{\text{in}} + \theta_{\text{out}})} + o(1) \tag{10}$$

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

$$d_j^{\text{in}} = \sum_{k=j+1}^{n} x_{jk} \qquad \text{and} \qquad d_j^{\text{out}} = \sum_{i=1}^{j-1} x_{ij} \tag{11}$$

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

$$E(n) = \sum_{i=1}^{n} d_i^{\text{in}} = \sum_{i=1}^{n} d_i^{\text{out}} \tag{12}$$

The expected number of edges added at each step is

$$
\begin{aligned}
\langle E(n+1) - E(n) \rangle &= \sum_{i=1}^{n} \frac{\alpha + \theta_{\text{in}} d_i^{\text{in}} + \theta_{\text{out}} d_i^{\text{out}}}{n + \alpha + \beta - 1} \\
&= \frac{\alpha n + (\theta_{\text{in}} + \theta_{\text{out}}) E(n)}{n + \alpha + \beta - 1}
\end{aligned}
\tag{13}
$$

We make an ansatz of constant average degree

$$\langle E(n) \rangle = C_1 \times n + g(n) \tag{14}$$

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 (13),

$$\langle E(n+1) \rangle - \langle E(n) \rangle = \frac{\alpha n + (\theta_{\text{in}} + \theta_{\text{out}}) \langle E(n) \rangle}{n + \alpha + \beta - 1} \tag{15}$$

using our ansatz (14),

$$g(n+1) - g(n) + C_1 = \underbrace{\left(\alpha + (\theta_{\text{in}} + \theta_{\text{out}}) \left(C_1 + \frac{g(n)}{n} \right) \right)}_{\text{equate these constant terms}} \times \frac{n}{n + \alpha + \beta - 1} \tag{16}$$

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

$$
\begin{aligned}
C_1 &= \alpha + (\theta_{\text{in}} + \theta_{\text{out}}) C_1 \tag{17} \\
&= \frac{\alpha}{1 - (\theta_{\text{in}} + \theta_{\text{out}})} \tag{18}
\end{aligned}
$$

we obtain the asymptotic average degree claimed in (10).

To check verify our ansatz $g(n) = o(n)$, substitute (18) into (16)

$$\underbrace{g(n+1) - g(n) = \frac{(\theta_{\text{in}} + \theta_{\text{out}}) g(n)}{n}}_{\text{equate these dominant terms}} - \frac{\alpha + \beta - 1}{n} + \mathcal{O}\left(\frac{g(n)}{n^2} \right) \tag{19}$$

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 (10) is physically meaningful.

C.1.2 Logarithmic degree regime: $\theta_{\text{in}} + \theta_{\text{out}} = 1$

As $\theta_{\text{in}} + \theta_{\text{out}} \to 1$, the average degree predicted by (10) diverges, and when $\theta_{\text{in}} + \theta_{\text{out}} = 1$, the average degree is no longer bounded. Instead, it grows logarithmically in n:

$$\langle d(n) \rangle = 2\alpha \log(n) + C + o(1) \tag{20}$$

Similar to equation (14), we make an ansatz but now of logarithmic average degree

$$\langle E(n)\rangle = C_1 \times n \log n + g(n) \tag{21}$$

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

The change in $n \log n$ is approximately

$$(n+1)\log(n+1) - n\log(n) = \log(n) + 1 + \mathcal{O}\left(\tfrac{1}{n^2}\right). \tag{22}$$

Substituting this into equation (15),

$$g(n+1) - g(n) + \underbrace{C_1 \times \left(\log(n) + 1\right) = \left(\alpha + (\theta_{\text{in}} + \theta_{\text{out}})\left(C_1 \log(n) + \frac{g(n)}{n}\right)\right)}_{\text{equate these logarithmic and constant terms}} \times \left(1 + \mathcal{O}\left(\tfrac{1}{n}\right)\right) \tag{23}$$

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

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

$$g(n+1) - g(n) = \frac{g(n)}{n} + \mathcal{O}\left(\tfrac{\log n}{n}\right) \tag{24}$$

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 networks 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].

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

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

$$\langle d(n)\rangle \propto C n^\rho, \qquad \rho = \theta_{\text{in}} + \theta_{\text{out}} - 1 \tag{25}$$

Again, we make the appropriate ansatz:

$$\langle E(n)\rangle = C_1 \times n^{1+\rho} + g(n) \tag{26}$$

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

The change in $n^{1+\rho}$ is approximately

$$(n+1)^{1+\rho} - n^{1+\rho} = (1+\rho)n^\rho + \mathcal{O}\left(n^{\rho-1}\right). \tag{27}$$

Substituting this into equation (15),

$$g(n+1) - g(n) + \underbrace{C_1 \times (1+\rho)n^\rho = \left(\alpha + (\theta_{\text{in}} + \theta_{\text{out}})\left(C_1 n^\rho + \frac{g(n)}{n}\right)\right)}_{\text{equate these } \mathcal{O}(n^\rho) \text{ terms}} \times \left(1 + \mathcal{O}\left(\tfrac{1}{n}\right)\right) \tag{28}$$

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

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:

$$g(n+1) - g(n) = (1+\rho)\frac{g(n)}{n} + \mathcal{O}\left(n^{\rho-1}\right) \tag{29}$$

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.

$\square$

C.2 PROOF OF THEOREM 4 (*Power-law degree distributions of the PAPA model*)

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

$$\text{constant:} \qquad \gamma = \frac{1 + \theta_{\text{in}}}{\theta_{\text{in}}} \qquad\qquad 0 < \theta_{\text{in}} + \theta_{\text{out}} \le 1 \tag{30}$$

$$\text{polynomial:} \qquad \gamma = \frac{2 - \theta_{\text{out}}}{1 - \theta_{\text{out}}} \qquad\qquad 1 \le \theta_{\text{in}} + \theta_{\text{out}} < 2 \tag{31}$$

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).

C.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} ($1 \le i < j$) are conditionally independent given the previous entries x_{ab} ($1 \le a < b < j$), and the sum of independent Bernoulli variables does not exhibit a power-law distribution.

Denote these (conditionally) independent probabilities as p_{ij}. In our proof of the sparsity of the model above, 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

$$\text{constant:} \qquad \langle d_j^{\text{out}} \rangle \sim \frac{\alpha}{1 - (\theta_{\text{in}} + \theta_{\text{out}})} \qquad 0 < \theta_{\text{in}} + \theta_{\text{out}} < 1 \tag{32}$$

$$\text{logarithmic:} \qquad \langle d_j^{\text{out}} \rangle \sim \alpha \log(j) + C \qquad\qquad \theta_{\text{in}} + \theta_{\text{out}} = 1 \tag{33}$$

$$\text{polynomial:} \qquad \langle d_j^{\text{out}} \rangle \sim C \times j^{\theta_{\text{in}} + \theta_{\text{out}} - 1} \qquad 1 < \theta_{\text{in}} + \theta_{\text{out}} < 2 \tag{34}$$

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

C.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:

$$\langle d_j^{\text{in}} \rangle(n+1) - \langle d_j^{\text{in}} \rangle(n) = \frac{\alpha + \theta_{\text{out}} d_j^{\text{out}} + \theta_{\text{in}} \langle d_j^{\text{in}} \rangle(n)}{n + \alpha + \beta - 1} \qquad \langle d_j^{\text{in}} \rangle(j) = 0 \tag{35}$$

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

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

where $\Gamma(n) := (n-1)!$ is the gamma function.

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

$$\langle d_j \rangle(n) \approx \frac{\alpha + \theta_{\text{out}} \langle d_j^{\text{out}} \rangle}{\theta_{\text{in}}} \left(\left(\frac{n}{j} \right)^{\theta_{\text{in}}} - 1 \right) + \langle d_j^{\text{out}} \rangle \tag{37}$$

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:

$$p(d) \propto \left| \frac{d}{dj} \langle d_j \rangle \right|^{-1} \tag{38}$$

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

$$\gamma = \frac{\frac{d}{dj} \log \left| \frac{d}{dj} \langle d_j \rangle \right|}{\frac{d}{dj} \log \langle d_j \rangle} = \frac{\langle d_j \rangle \frac{d^2}{dj^2} \langle d_j \rangle}{\left(\frac{d}{dj} \langle d_j \rangle \right)^2} \tag{39}$$

Substitute the expressions for $\langle d_j^{\text{out}} \rangle$ from equations (32) and (34) into equation (37). When $1 \ll j \ll n$, the dominant term has the following scalings:

$$\text{constant:} \qquad \langle d_j \rangle(n) \propto n^{\theta_{\text{in}}} j^{-\theta_{\text{in}}} \qquad\qquad 0 < \theta_{\text{in}} + \theta_{\text{out}} < 1 \tag{40}$$

$$\text{polynomial:} \qquad \langle d_j \rangle(n) \propto n^{\theta_{\text{in}}} j^{\theta_{\text{out}} - 1} \qquad\qquad 1 < \theta_{\text{in}} + \theta_{\text{out}} < 2 \tag{41}$$

Substituting these into equation (39), we obtain the following power-law exponents:

$$\text{constant:} \qquad \gamma = \frac{1 + \theta_{\text{in}}}{\theta_{\text{in}}} \qquad\qquad 0 < \theta_{\text{in}} + \theta_{\text{out}} \leq 1 \tag{42}$$

$$\text{polynomial:} \qquad \gamma = \frac{2 - \theta_{\text{out}}}{1 - \theta_{\text{out}}} \qquad\qquad 1 \leq \theta_{\text{in}} + \theta_{\text{out}} < 2 \tag{43}$$

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

$\square$