

# Heterogeneous node copying from hidden network structure

Max Falkenberg McGillivray (✉ [mff113@ic.ac.uk](mailto:mff113@ic.ac.uk))

Imperial College London <https://orcid.org/0000-0002-2986-2494>

---

## Article

### Keywords:

**Posted Date:** April 14th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-258292/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Communications Physics on September 2nd, 2021. See the published version at <https://doi.org/10.1038/s42005-021-00694-1>.

# Heterogeneous node copying from hidden network structure

Max Falkenberg<sup>1, \*</sup>

<sup>1</sup>*Centre for Complexity Science, Imperial College London, SW7 2AZ, U.K.*

(Dated: April 14, 2021)

Node copying is an important mechanism for social network formation, yet most models assume uniform copying rules. Motivated by observations of heterogeneous triadic closure in real networks, we introduce the concept of a hidden network model – a generative two-layer model in which an observed network evolves according to the structure of an underlying hidden layer – and apply the framework to a model of heterogeneous copying. Framed in a social context, these two layers may represent a node’s inner social circle, and wider social circle, such that the model can bias copying probabilities towards, or against, a node’s inner circle of friends. Comparing the case of extreme inner circle bias to an equivalent model with uniform copying, we find that heterogeneous copying suppresses the power-law degree distributions commonly seen in copying models, and results in sparse networks with significantly higher clustering than even the most optimum scenario for uniform copying. Similarly large clustering values are found across a range of real collaboration networks, lending empirical support to the mechanism.

**Introduction.** Node copying is an important network growth mechanism [1–7]. In social networks, copying is synonymous with triadic closure, playing an important role in the emergence of high clustering [8, 9]. In biology, node copying encapsulates duplication and deletion, a key mechanism in the formation of protein-interaction networks [10–12].

Despite this range of applications, most node copying models assume uniform, or homogeneous copying, i.e., that the probability of copying any given neighbour of a node is equal. The exact formulation varies widely, but examples include “links are attached to neighbours of [node]  $j$  with probability  $p$ ” [5], or “one node [is duplicated]... edges emanating from the newly generated [node] are removed with probability  $\delta$ ” [11]. Similar uniform copying rules are found in [2, 3, 6–8, 10, 12–20].

Homogeneous copying is a sensible base assumption, often aiding a model’s analytical tractability. However, especially in a social context, there is good reason to believe that node copying may be heterogeneous. Consider for instance the social brain hypothesis, a theory which suggests that the average human has around 150 friends (Dunbar’s number), encapsulating progressively smaller sub-groups of increasing social importance [21, 22]. However, large social networks often have an average degree far exceeding Dunbar’s number [23], implying that most of these observed friends are only distant acquaintances. In this context, if individual A introduces individual B to one their friends, C, (i.e., B is *copying* A’s friend C), we may reasonably expect that C is more likely to be chosen from A’s inner social circle, than A’s wider social circle.

This is directly related to the principle of *strong triadic closure*: “If a node A has edges to nodes B and C, then the B-C edge is especially likely to form if A’s edges to B and C are both strong ties” [24]. In weighted networks where tie strength can be equated to edge weight,

empirical evidence for the strong triadic closure principle can be inferred by measuring the neighbourhood overlap between two nodes as a function of tie strength [24]; see for example [25] using mobile communication networks, or [26] using face to face proximity networks.

Unfortunately, for many networks tie strength data is unavailable or unknown. In these cases, evidence for asymmetric triadic closure may be inferred through proxy means. For instance, in academic collaboration networks it has been shown that the ratio of triadic closure varies strongly with the number of shared collaborators between nodes [27]. Although the average triadic closure ratio is small (typically  $< 10\%$ ), the ratio rapidly increases with the number of shared collaborators. However, these aggregate measures are highly coarse grained and likely only approximate real closure dynamics.

This motivates the study of simple heterogeneous copying models, such as those found in [4, 5, 9, 28, 29]. Typically these models fall into a small number of distinct categories. In the first, heterogeneity is introduced as a node intrinsic property (e.g. node fitness) in the absence of structural considerations [5]. In the second, heterogeneity is introduced via group homophily [9, 29] where the probability of triadic closure between nodes A and B is dependent on whether nodes A and B are in the same group or different groups (e.g. researchers from the same academic discipline, as opposed to different disciplines). However, intra-group copying is typically modelled uniformly. Finally, some models consider heterogeneous copying driven by the network structure around nodes A and B, without introducing node homophily [4].

In the threshold model defined in [4], the authors consider the case where node A introduces node B to one of their friends C. An edge then forms between B and C if the fraction of neighbours common to B and C exceeds some threshold  $F$ . The model demonstrates a transition from a state where networks are almost complete for small  $F$ , to a state where networks are sparse but highly clustered as  $F$  increases past a critical threshold. By using a fractional threshold for neighbour overlap, the

---

\* max.falkenberg13@imperial.ac.uk

framework avoids some issues where closure probability is associated with the absolute overlap of the local communities. However, the model is limited in its tractability and has peculiarities such as the observation that fringe communities are almost always complete.

In this paper, our aim is to extend these ideas and introduce a more general framework for heterogeneous node copying based on the concept of “hidden strong ties”. To do so, we introduce the *hidden network model* as a novel modelling framework. A multilayer network [30] where layers have identical node structure but different edge structure, the framework lets us build models with local heterogeneity in the rules of network growth, but where that heterogeneity is a property of a hidden network structure and not arbitrarily encoded using node intrinsic properties or group homophily. The concept has loose similarities to other multilayer paradigms including the use of replica nodes to model heterogeneity [31] and interdependent networks [32]. See also [19] on multilayer copying and triadic closure.

The framework is very general and can be applied in a number of contexts. For instance, the concept can be used to decompose complex single layer networks into simpler two-layer structures. One such example is second-neighbour preferential attachment; an implementation of the Barabási-Albert model where nodes attach proportionally to the number of nodes within two steps of a target node, see [33]. Using our framework, the model is decomposed into an observed network, and a hidden network where nodes are connected to all nodes which are two or fewer steps away, representing the node’s sphere of influence [34]. Here, second-neighbour preferential attachment is equivalent to conventional first-neighbour preferential attachment followed by a local copying step. As detailed in [33], structural heterogeneity that is intrinsic in such a model has profound consequences for the time dependence of network growth.

In this general context, the hidden network model has similarities to other paradigms that aim to encode implicit heterogeneity in network models. One common approach is to define “hidden variables” on nodes [35, 36]. However, whether two nodes are connected is typically encoded as a function of these node intrinsic properties, rather than the structure between nodes.

In our context where we are interested in heterogeneous growth emerging from heterogeneous structure (not heterogeneous nodes), closer parallels can be found in the study of higher-order networks where there is an explicit distinction between a size  $n$  clique of pairwise interactions, and a higher order interaction represented by an  $n$ -simplex, leading to different network dynamics [37]. Although our focus in the current paper is on emergent heterogeneity in traditional pairwise networks, how hidden structure influences observations is related to questions regarding the projection between higher order networks and their pairwise equivalents [38]. On page 8, we briefly outline how our framework may be used to construct random simplicial complexes from heterogeneous

copying with which higher order dynamics may be studied.

In the remainder of this paper, we first outline a simple uniform copying model (UCM) [2, 3] used as the foundation for our model of heterogeneous copying. We will formally introduce the concept of hidden network models and define the specific case of extreme heterogeneous copying, the correlated copying model (CCM), for which we provide analytical results following the approach in [2]. Relaxing the extreme copying case, we numerically investigate a generalised form of the correlated copying model (GCCM) which exhibits a spectrum of heterogeneity, covering both the uniform and extreme bias cases.

The GCCM generates a diverse spectrum of network structures spanning both ergodic sparse and non-ergodic dense networks, with degree distributions ranging from exponential decay, through stretched-exponentials and power-laws, to extremely fat tailed distributions with anomalous fluctuations. These networks exhibit a broad clustering spectrum from sparse networks with significantly higher clustering than their uniform equivalents, to the unusual case where networks are almost complete, but with near zero clustering. Focusing on the highly clustered sparse regime, we comment on a number of real social networks, particularly collaboration networks, which exhibit higher clustering than can be explained by even the most optimum parameter choices in uniform copying. This suggests that heterogeneous copying may be an important explanatory mechanism for social network formation.

**Uniform Copying Model.** The uniform copying model (UCM) was introduced in [2, 3], see Fig. 1. At time  $t_\alpha$ , a single node,  $\alpha$ , is added to the network, and one target node,  $\beta$ , is chosen uniformly at random. We label each neighbour of  $\beta$  with the index  $\gamma_j$  where  $j \in \{1, \dots, k^\beta\}$ , and  $k^\beta$  is the degree of node  $\beta$ . For each neighbour  $\gamma_j$ , the copied edge  $(\alpha, \gamma_j)$  is added to the network independently with probability  $p$ . The network is initialised at  $t = 1$  with a single node. If  $p = 0$ , no edges are copied forming a random recursive tree. If  $p = 1$ , the UCM generates a complete graph.

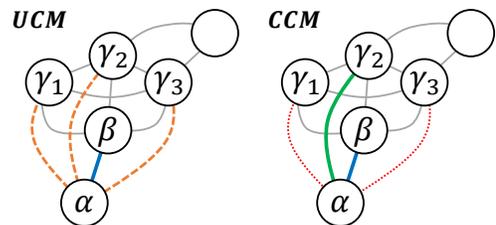


FIG. 1. A new node  $\alpha$  is added to the network and forms an edge to a random target node,  $\beta$ . UCM: All neighbours of node  $\beta$  have an equal probability  $p$  of being copied (orange dashed). CCM: Copied edges are added deterministically; neighbours of node  $\beta$  in the hidden network are copied (solid green), the remainder are not copied (red dotted).

**Hidden Network Models.** We define a hidden network model as the pair of single layer graphs  $G = (G_O, G_H)$ , comprising an observed network  $G_O = (V, E_O)$  and a hidden network  $G_H = (V, E_H)$ , where  $V$  is the set of nodes for both networks and  $E_O$  and  $E_H$  are the set of edges for each network. The set  $V$  represent the same entities in both  $G_O$  and  $G_H$ , with differences lying exclusively in the edge structure between nodes. The key feature of a hidden network model is that the evolution of  $G_O$  is dependent on  $G_H$  (or vice versa).

**Correlated Copying Model.** In the correlated copying model (CCM), see Fig. 2, the observed and hidden networks are initialised with a single node at  $t = 1$ . At  $t = t_\alpha$ , node  $\alpha$  is added to both networks and a single target node,  $\beta$ , is chosen uniformly at random. We label the  $k_O^\beta$  neighbours of  $\beta$  in  $G_O$  with the index  $\gamma_j$ . Then, in the observed network only, the copied edge  $(\alpha, \gamma_j)$  is formed with  $p_H = 1$  if the edge  $(\beta, \gamma_j) \in E_H$ ,  $p_O = 0$  otherwise. The general case with intermediate copying probabilities is discussed on page 6. No copied edges are added to the hidden network  $G_H$ . The direct edge  $(\alpha, \beta)$  is added to both  $G_O$  and  $G_H$ .

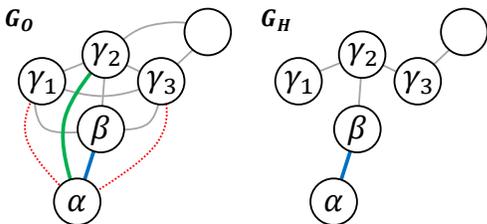


FIG. 2. The correlated copying model. A new node,  $\alpha$ , forms an edge (solid blue) to a randomly chosen target node  $\beta$ . Copied edges  $(\alpha, \gamma_j)$  (solid green) are formed in the observed network,  $G_O$ , if the edge  $(\alpha, \gamma_j)$  is present in the hidden network,  $G_H$ . Other neighbours of node  $\beta$  are not copied (dotted red).

$G_H$  evolves as a random recursive tree. Unlike the UCM, all copying in  $G_O$  is deterministic, with the only probabilistic element emerging in the choice of the target node  $\beta$ . For comparative purposes, we define the effective copying probability in the CCM as  $p_{\text{eff}} = \langle k_H^\beta / k_O^\beta \rangle$ , i.e., the fraction of the observed neighbours of node  $\beta$  which are copied by node  $\alpha$ .

Framed in a social context, we might think of  $G_O$  as an observed social network where individuals have many friends, but the quality of those friendships is unknown, with most ties being weak. In contrast, underlying every social network is a hidden structure representing the inner social circle of individuals, where a node is only connected to their closest friends [22]. Copying in the CCM is biased to this inner circle.

**Preliminaries.** The total number of edges in  $G_H$  scales as  $E_H(t) \sim t$ , with the average degree given by  $\langle k_H \rangle = 2$ . Using the degree distribution of  $G_H$ , see below,  $\langle k_H^2 \rangle = 6$ . In the observed network, each time step a single edge is added by direct attachment, and one copied

edge is added for each neighbour of the target node in  $G_H$ ,  $k_H^\beta$ . The average change in the number of edges is therefore  $\langle \Delta E_O(t) \rangle = 1 + \langle k_H^\beta \rangle = 1 + \langle k_H \rangle = 3$ , such that  $\langle E_O(t) \rangle \sim 3t$  and  $\langle k_O \rangle = 6$ .

As an alternative, note that the observed degree of node  $\alpha$  can be written as

$$(k_O)_\alpha = \sum_{\beta=1}^{(k_H)_\alpha} (k_H)_{\alpha,\beta} \quad (1)$$

where the index  $\alpha, \beta$  labels the  $(k_H)_\alpha$  unique neighbours of  $\alpha$  in  $G_H$ . Averaging both sides of Eq. (1) over all nodes we find,

$$\langle k_O \rangle = \frac{1}{t} \sum_{\alpha=1}^t \sum_{\beta=1}^{(k_H)_\alpha} (k_H)_{\alpha,\beta} = \frac{1}{t} \sum_{\ell=1}^t n_\ell (k_H)_\ell, \quad (2)$$

where  $n_\ell$  is the number of times that the degree of node  $\ell$  appears in the expanded sum. For any tree graph, node  $\ell$  will appear exactly once in Eq. (2) for each of its  $(k_H)_\ell$  neighbours. Hence,  $n_\ell = (k_H)_\ell$  and  $\langle k_O \rangle = \langle k_H^2 \rangle$ . In the supplementary material (SM), Eq. (1) is used to derive  $\langle k_O^2 \rangle \approx 62$ .

We may naively expect that the effective copying probability is  $p_{\text{eff}} = \langle k_H \rangle / \langle k_O \rangle = 1/3$ . However, for the CCM,  $p_{\text{eff}} = \langle k_H^\beta / k_O^\beta \rangle \neq \langle k_H \rangle / \langle k_O \rangle$ . We have not found a route to calculating this exactly, but simulations suggest  $p_{\text{eff}} \approx 0.374$ .

**Degree Distribution.** The hidden network evolves as a random recursive tree which has a limiting degree distribution given by

$$p_H(k_H) = 2^{-k_H}, \text{ for } k_H > 1. \quad (3)$$

In the supplement, we show that the degree distribution for the observed network can be written as the recurrence

$$p_O(k_O) = \frac{\pi_O(k_O - 1) \cdot p_O(k_O - 1) + 2^{1-k_O}}{1 + \pi_O(k_O)}, \text{ for } k \geq 2, \quad (4)$$

where the final term is the probability that at time  $t$  the newly added node has initial degree  $k_O$  and

$$\pi_O(k_O) = 1 + \langle k_H | k_O \rangle, \quad (5)$$

with  $\langle k_H | k_O \rangle$  as the average degree of nodes in the hidden network with observed degree  $k_O$ . Here, the 1 corresponds to edges that are gained from direct attachment, whereas  $\langle k_H | k_O \rangle$  corresponds to edges gained from copying. Although we have not found an exact expression for  $\langle k_H | k_O \rangle$ , we can make progress by considering the evolution of individual nodes.

Consider node  $\alpha$  added to the network at  $t_\alpha$ . The initial conditions for node  $\alpha$  are

$$(k_H(t_\alpha))_\alpha = 1, \quad (6a)$$

$$(k_O(t_\alpha))_\alpha = 1 + \langle k_H(t_\alpha - 1) \rangle_\beta, \quad (6b)$$

where the final term is the average hidden degree of the target node  $\beta$ . In  $G_H$ , node  $\alpha$  gains edges from direct attachment only. Hence, at  $t > t_\alpha$ ,

$$\langle k_H(t) \rangle_\alpha = 1 + \sum_{j=t_\alpha}^{t-1} \frac{1}{j} = 1 + H_{t-1} - H_{t_\alpha-1}, \quad (7)$$

where  $H_n$  is the  $n^{\text{th}}$  harmonic number. In  $G_O$ , either node  $\alpha$  is targeted via direct attachment, or a copied edge is formed from the new node to node  $\alpha$  via any of the  $(k_H(t))_\alpha$  neighbours of node  $\alpha$ . Hence,

$$\begin{aligned} \langle k_O(t) \rangle_\alpha &= \langle k_O(t_\alpha) \rangle_\alpha + \sum_{j=t_\alpha}^{t-1} \frac{1 + \langle k_H(j) \rangle_\alpha}{j} \\ &= \langle k_O(t_\alpha) \rangle_\alpha + \sum_{j=t_\alpha}^{t-1} \frac{2 + H_j - H_{t_\alpha-1} - 1/j}{j}, \end{aligned} \quad (8)$$

where we have subbed in Eq. (7) and  $H_{j-1} = H_j - 1/j$ . Evaluating this sum, see supplement, we find

$$\begin{aligned} \langle k_O(t) \rangle_\alpha &= \langle k_O(t_\alpha) \rangle_\alpha + \frac{1}{2} \left[ (4 + H_{t-1} - H_{t_\alpha-1}) \right. \\ &\quad \left. \times (H_{t-1} - H_{t_\alpha-1}) - H_{t-1}^{(2)} + H_{t_\alpha-1}^{(2)} \right], \end{aligned} \quad (9)$$

where  $H_n^{(m)}$  is the  $n^{\text{th}}$  generalised Harmonic number of order  $m$ . For  $t \rightarrow \infty$ ,  $H_t^{(2)} \rightarrow \pi^2/6$ . Hence, for large  $t$  we can drop the final two terms and substitute in Eq. (7) to give

$$\langle k_O(t) \rangle_\alpha \approx \langle k_O(t_\alpha) \rangle_\alpha + \frac{1}{2} (\langle k_H(t) \rangle_\alpha + 3) (\langle k_H(t) \rangle_\alpha - 1). \quad (10)$$

Noting, that Eq. (10) is a monotonically increasing function of  $k_H$  for  $k_H > 1$ , we assume that we can drop the index  $\alpha$  and the time dependence giving the average observed degree of nodes with specific hidden degree as

$$\langle k_O | k_H \rangle \approx \langle \tilde{k}_O | k_O \rangle + \frac{1}{2} (k_H + 3) (k_H - 1), \quad (11)$$

where  $\langle \tilde{k}_O | k_O \rangle$  denotes the average initial observed degree of nodes with current degree  $k_O$ . Finally, we make the approximation that  $\langle k_H | k_O \rangle \approx \langle k_O | k_H \rangle^{-1}$  where the exponent denotes the inverse function. This gives

$$\pi_O(k_O) = 1 + \langle k_H | k_O \rangle \approx \sqrt{2(k_O + 2 - \langle \tilde{k}_O | k_O \rangle)}. \quad (12)$$

To proceed, let us solve the degree distribution at  $k_O = 2$ . Although the average initial condition  $\langle \tilde{k}_O \rangle = 1 + \langle k_H \rangle = 3$ , in this case  $\langle \tilde{k}_O | 2 \rangle = 2$ . Therefore

$$p_O(2) = -\pi_O(2) \cdot p_O(2) + 2^{-1} = -p_O(2) \cdot \sqrt{2(2)} + 2^{-1}, \quad (13)$$

giving  $p_O(2) = 1/6$ . Since  $\langle \tilde{k}_O | k_O \rangle$  has an almost negligible effect on  $\pi_O(k_O)$  for  $k_O > 2$ , for simplicity we set  $\langle \tilde{k}_O | k_O \rangle = 2$ . We can now rewrite Eq. (4) as

$$p_O(k_O) = \frac{p_O(k_O - 1) \sqrt{2(k_O - 1)} + 2^{1-k_O}}{1 + \sqrt{2k_O}}, \quad \text{for } k_O > 2. \quad (14)$$

Although computing this recurrence shows good agreement with simulations, see Fig. 3, we have not found a closed form solution to Eq. (14).

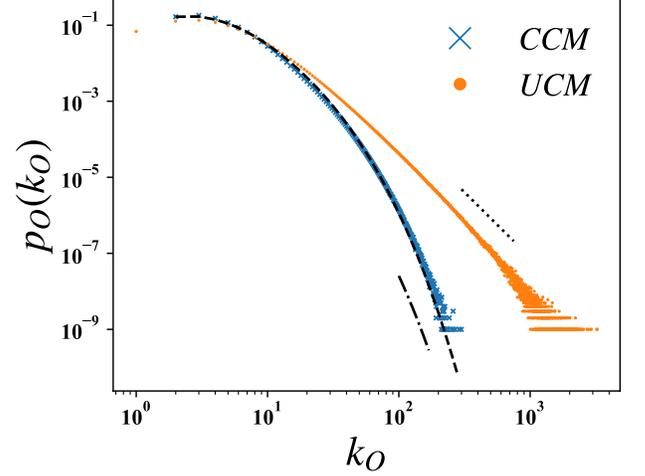


FIG. 3. The degree distribution for the CCM (blue crosses) and UCM ( $p = p_{\text{eff}} = 0.374$ ; orange points) at  $t = 10^7$ , averaged over 100 networks. Dashed line: CCM recurrence relation in Eq. (14). Dot-dashed: stretched exponential approximation. Dotted: power-law scaling.

As an approximation, we return to Eq. (9) and note that  $H_{t-1} - H_{t_\alpha-1} \approx \ln(t/t_\alpha)$ . Substituting this into Eq. (9) and dropping small terms

$$\langle k_O(t > t_\alpha) \rangle_\alpha \approx 2 \ln(t/t_\alpha) + \frac{\ln^2(t/t_\alpha)}{2}, \quad (15)$$

which inverted gives

$$\ln(t/t_\alpha) \approx -2 + \sqrt{2(k_O + 2)} \approx \sqrt{2k_O}, \quad \text{for } k \gg 2. \quad (16)$$

We have dropped the expectation value and define  $t_\alpha$  as the time a node was created such that its degree at time  $t$  is approximately  $k_O$ . Exponentiating each side and taking the reciprocal,

$$\frac{t_\alpha}{t} \approx e^{-\sqrt{2k_O}}. \quad (17)$$

Finally, by substituting this approximation into the cumulative degree distribution we find

$$\tilde{p}_O(k_O) = \sum_{k'_O=2}^{k_O} p_O(k'_O) \approx 1 - \frac{t_\alpha}{t} \approx 1 - e^{-\sqrt{2k_O}}, \quad (18)$$

which corresponds to a Weibull (stretched exponential) distribution, suppressing the power-law scaling observed in the UCM, see Fig. 3.

The approximation for the cumulative degree distribution stems from the observation that, on average, nodes with  $k'_O > k_O$  were added to the network at  $t' < t_\alpha$ , whereas nodes with  $k'_O < k_O$  were added to the network at  $t' > t_\alpha$ . Both Eq. (16) and Eq. (18) are close to the scaling expected from sub-linear preferential attachment with an exponent  $1/2$ , see [39].

**Clique Distribution.** In a simple undirected graph, a clique of size  $n$  is a subgraph of  $n$  nodes which forms a complete subgraph. A clique of size  $n = 2$  is an edge, whereas  $n = 3$  is a triangle. Here we calculate the exact scaling for the number  $n$  cliques,  $Q_n(t)$ , in  $G_O$ .

Let us first consider the case of triangles. At  $t = t_\alpha$ , there are two mechanisms by which a new triangle forms:

1. **Direct triangles.** The new node,  $\alpha$ , forms a direct edge to the target node,  $\beta$ , and forms copied edges to each of the  $k_H^\beta$  neighbours of node  $\beta$ , labelled with the index  $\gamma_j$ . The combination of the direct edge  $(\alpha, \beta)$ , the copied edge  $(\alpha, \gamma_j)$ , and the existing edge  $(\beta, \gamma_j)$  creates one triangle,  $(\alpha, \beta, \gamma_j)$ , for each of the  $k_H^\beta$  neighbours.
2. **Induced triangles.** If node  $\alpha$  forms copied edges to both node  $\gamma_j$ , and to node  $\gamma_{j'}$ ,  $j \neq j'$ , the triangle  $(\alpha, \gamma_j, \gamma_{j'})$  is formed if  $(\gamma_j, \gamma_{j'}) \in E_O$ .

Combining these mechanisms, the change in the number of triangles can be written as

$$\Delta Q_3(t_\alpha) = \Delta Q_3^D(t_\alpha) + \Delta Q_3^I(t_\alpha), \quad (19)$$

where the first and second terms on the right correspond to direct and induced triangles respectively. One new direct triangle is formed for each of the  $k_H^\beta$  neighbours of node  $\beta$ ,  $\Delta Q_3^D = k_H^\beta$ . For induced triangles, the copied edge  $(\alpha, \gamma_j)$  is only formed if  $(\beta, \gamma_j) \in E_H$ . Additionally, all pairs of nodes which are next-nearest neighbours in  $G_H$  must be nearest neighbours in  $G_O$ . Hence, the edge  $(\gamma_j, \gamma_{j'})$  must exist in the observed network if both  $\gamma_j$  and  $\gamma_{j'}$  are copied. As a result, one induced triangle is formed for each pair of copied edges  $(\alpha, \gamma_j)$  and  $(\alpha, \gamma_{j'})$  such that

$$\Delta Q_3^I = \binom{k_H^\beta}{2} = \frac{(k_H^\beta)^2 - k_H^\beta}{2}. \quad (20)$$

Extending the argument to general  $n$  we can write

$$\Delta Q_n(t_\alpha) = \Delta Q_n^D(t_\alpha) + \Delta Q_n^I(t_\alpha), \quad (21)$$

where direct cliques are those which include the edge  $(\alpha, \beta)$ . For a clique of size  $n$ , the number of direct cliques is given by the number of ways in which  $n - 2$  nodes can be chosen from  $k_H^\beta$  nodes,

$$\Delta Q_n^D(t_\alpha) = \binom{k_H^\beta}{n-2}, \quad (22)$$

whereas the number of induced cliques is given by the number of ways in which  $n - 1$  nodes can be chosen,

$$\Delta Q_n^I(t_\alpha) = \binom{k_H^\beta}{n-1}. \quad (23)$$

As  $t \rightarrow \infty$ , the average change in clique number is

$$\langle \Delta Q_n(t) \rangle = \sum_{k_H=1}^{\infty} p_H(k_H) \left[ \binom{k_H}{n-2} + \binom{k_H}{n-1} \right], \quad (24)$$

where  $p_H(k_H)$  is the probability that the randomly chosen target node  $k_H^\beta = k_H$ . To avoid ill-defined binomials, we rewrite Eq. (24) as

$$\langle \Delta Q_n(t) \rangle = p_H(n-2) + \sum_{k_H=n-1}^{\infty} p_H(k_H) \cdot \binom{k_H+1}{n-1}, \quad (25)$$

where we have combined the two terms into a single binomial. After substiting in  $p_H(k_H)$  and solving the sum,

$$\langle \Delta Q_n(t) \rangle = 2^{2-n} + \sum_{k_H=n-1}^{\infty} 2^{-k_H} \cdot \binom{k_H+1}{n-1} = 4. \quad (26)$$

Consequently, for large  $t$  we find the curious result that the number of  $n$  cliques scales as

$$Q_n(t) \sim 4t, \text{ for } n > 2, \quad (27)$$

independent of the clique size.

**Clustering.** Transitivity is a global clustering measure defined as

$$\tau_{G_O} = 3 \times \frac{\#(\text{triangles in } G_O)}{\#(\text{twigs in } G_O)}, \quad (28)$$

where a twig is any three nodes connected by two edges. The number of twigs is equivalent to the number of star graphs of size 2,  $S_2$ , where a star graph of size  $n$  is a subgraph with 1 central node and  $n$  connected neighbours. The number of subgraphs of size 2 is related to the degree distribution by

$$S_2(t) = t \sum_{k_O \geq 2} \binom{k_O}{2} \cdot p_O(k_O) = t \cdot \frac{\langle k_O^2 \rangle - \langle k_O \rangle}{2}, \quad (29)$$

where we have used the property that  $p_O(k < 2) = 0$ . Recalling that  $\langle k_O \rangle = 6$  and  $\langle k_O^2 \rangle \approx 62$ , the number of twigs scales as  $S_2 \sim 28t$ , such that

$$\tau_{G_O} = \frac{3Q_3}{S_2} \sim \frac{3 \cdot 4t}{28t} = \frac{3}{7}. \quad (30)$$

The local clustering coefficient,  $cc(\alpha)$ , is defined as the number of edges between the  $(k_O)_\alpha$  neighbours of  $\alpha$ , normalised by the number of edges in a complete subgraph of size  $(k_O)_\alpha$ . For the CCM,

$$cc(\alpha) = \frac{\binom{(k_H)_\alpha}{2} + \sum_{\beta=1}^{(k_H)_\alpha} \binom{(k_H)_{\alpha,\beta}}{2}}{\binom{(k_O)_\alpha}{2}}, \quad (31)$$

where the first term corresponds to the complete subgraph of the  $(k_H)_\alpha$  neighbours of  $\alpha$  in  $G_H$ , and the sum contributes the edges from one complete subgraph formed by node  $\alpha, \beta$  and its  $(k_H)_{\alpha, \beta} - 1$  neighbours, excluding  $\alpha$ . The global clustering coefficient,  $CC(G_O)$ , is defined as the average of Eq. (31) over all nodes in the network. In simulations,  $CC(G_O) \approx 0.771$  for large  $t$ .

**General Correlated Copying Model.** The general correlated copying model (GCCM), is defined analogously to the CCM, starting with observed and hidden networks initialised at  $t = 1$ . For practical reasons, we initialise the graph with three nodes which form a complete graph in  $G_O$ , and a wedge in  $G_H$ . This ensures that the initial graph contains some edges found in  $G_H$ , and some edges found only in  $G_O$ .

At  $t = t_\alpha$ , node  $\alpha$  is added to both networks and a single target node,  $\beta$ , is chosen uniformly at random. We label the  $k_O^\beta$  neighbours of  $\beta$  in  $G_O$  with the index  $\gamma_j$ . In the observed network, the copied edge  $(\alpha, \gamma_j)$  is formed with probability  $p_H$  if the edge  $(\beta, \gamma_j) \in E_H$  (inner circle copying), and probability  $p_O$  otherwise (outer circle copying). The direct edge  $(\alpha, \beta)$  is added to both  $G_O$  and  $G_H$ . Copied edges added to  $G_O$  are added to  $G_H$  independently with probability  $q$ . If  $p_H > 0$  or  $p_O > 0$ , and  $q > 0$ , the hidden network is no longer a random tree.

The GCCM encapsulates a wide spectrum of heterogeneous copying. Setting  $p_H = 1, p_O = 0$  and  $q = 0$  reduces the GCCM to the CCM, whereas setting  $p_H = p_O = p$ , for any  $q$ , reduces the GCCM to the UCM. We have discussed the social motivation for the case where  $p_H > p_O$ , representing a copying bias towards the inner social circle of a node. However, the GCCM can also be tuned to the reverse case where  $p_H < p_O$ , resulting in a bias against inner circle nodes. We are not aware of a physical motivation for this latter case. However, the unusual structural diversity of these anti-correlated networks warrants their discussion here.

Figure 4 shows numerical results for (a) the effective copying probability, (b) the densification exponent, (c) the average local clustering coefficient, and (d) the transitivity, for the GCCM with  $q = 0$  and  $10^4$  nodes.

The effective copying probability corresponds to the fraction of target node neighbours which appear to be copied in the observed network. Formally, we can write the average effective copying probability at time  $t$  as

$$p_{\text{eff}}(t) = \left\langle \frac{p_H k_H^\beta + p_O (k_O^\beta - k_H^\beta)}{k_O^\beta} \right\rangle, \quad (32)$$

where  $\beta$  is the index of the target node at time  $t$ , the first term represents edges copied from node  $\beta$ 's inner circle, and the second term represents edges copied from the outer circle.

The dashed contour in Fig. 4(b) corresponds to an effective copying probability of 0.5, calculated by averaging over the preceding  $10^4$  time steps. We note that  $p_{\text{eff}} = 0$  if  $p_H = p_O = 0$  (random tree),  $p_{\text{eff}} = 1$  if  $p_H = p_O = 1$

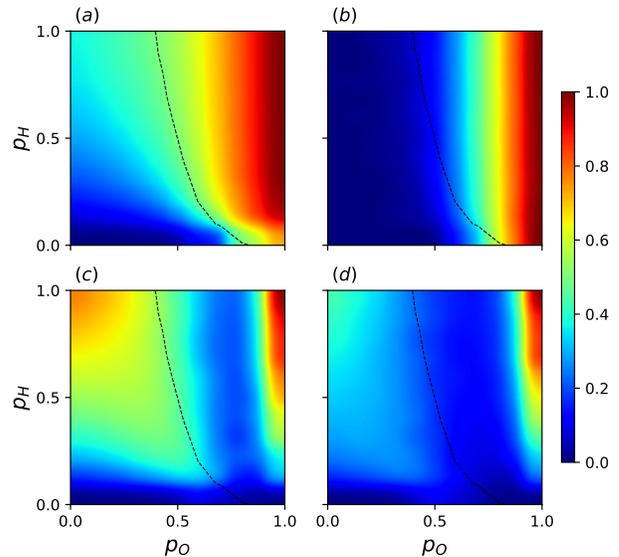


FIG. 4. Numerical results for the GCCM with  $q = 0$  and  $10^4$  nodes as a function of the hidden copying probability,  $p_H$ , and the outer copying probability,  $p_O$ . (a) The effective copying probability. (b) The densification exponent. (c) The average local clustering coefficient,  $CC(G_O)$ . (d) The transitivity,  $\tau_{G_O}$ . Black dashed contour: effective copying probability of 0.5 at  $t = 10^4$ . Values have been smoothed for clarity.

(complete graph), and  $p_{\text{eff}} = p$  if  $p_H = p_O = p$  (UCM). In general, the rise in  $p_{\text{eff}}$  is faster with increasing  $p_O$  than increasing  $p_H$ , although for  $p_H = 0$  we find very small  $p_{\text{eff}}$ , even for large  $p_O$ . However, this observation is somewhat deceptive since, if the GCCM is in the dense regime and  $p_H \neq p_O$ ,  $p_{\text{eff}}$  is not stationary. Calculated over longer time frames, we note that the effective copying probability appears to slowly converge to the outer copying probability,  $p_{\text{eff}} \rightarrow p_O$ , since for  $t \rightarrow \infty$ , the ratio of the number of edges in the hidden network to the number of edges in the observed network tends to zero. This suggests that the dashed  $p_{\text{eff}} = 0.5$  contour will converge to the  $p_O = 0.5$  line as  $t \rightarrow \infty$ .

We test whether the GCCM is in the sparse or dense regime explicitly by tracking the growth in the number of edges in the observed network. Let us define the densification exponent,  $\delta$ ,  $0 \leq \delta \leq 1$ , using  $E_O(t) \propto t^{1+\delta}$ , which relates the number of edges in the observed network to the number of nodes  $t$ . If  $\delta \approx 0$ , the GCCM is sparse. If  $\delta = 1$ , the GCCM grows as a complete graph. For intermediate values, the GCCM undergoes densification. For the UCM, the transition from the sparse to dense regime is known to take place at  $p = 0.5$  [3]. We have not analytically calculated the transition for the GCCM, but may intuitively expect the transition at  $p_O = 0.5$  since the hidden network is a random tree. This seems to be supported by the numerical values of  $\delta$  in Fig. 4(b), although the transition from zero to non-zero  $\delta$  is shifted

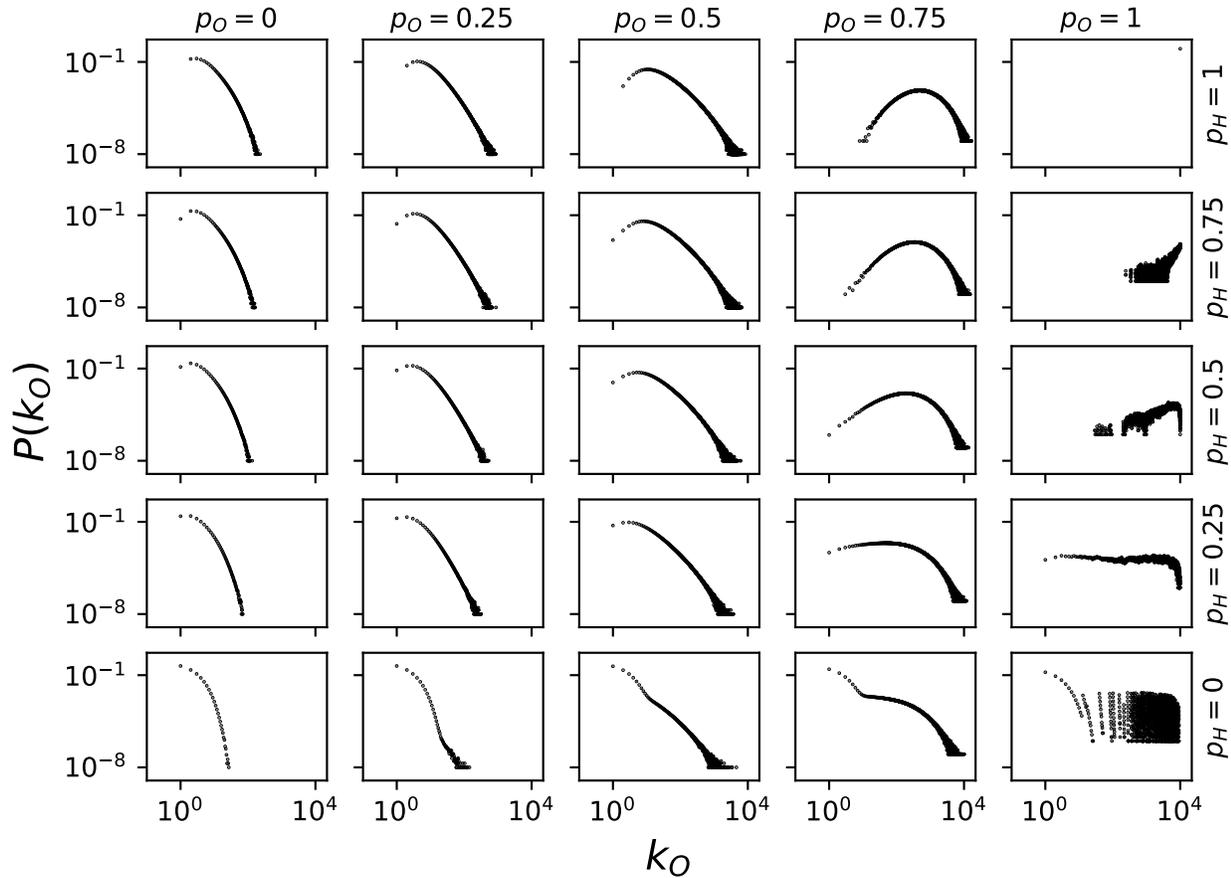


FIG. 5. The observed degree distribution in the GCCM for various values of the outer copying probability,  $p_O$  (left to right), and the hidden copying probability,  $p_H$  (bottom to top), for  $q = 0$ . For  $p_O \in \{0, 0.25, 0.5\}$ ,  $10^6$  nodes per network. For  $p_O = 0.75$ ,  $10^5$  nodes per network. For  $p_O = 1$ ,  $10^4$  nodes per network. Distributions are averaged over 100 instances. In the dense regime (large  $p_O$ , right column), network growth is non-ergodic leading to anomalous scaling and noisy degree distributions, even after averaging over many instances. The bottom left distribution (exponential decay) corresponds to a random recursive tree, see Eq. 3. The top right corresponds to a complete graph of  $N$  nodes. The top left corresponds to the CCM, see Fig. 3. The diagonal from the bottom left to the top right corresponds to the UCM.

to slightly larger  $p_O$  for  $p_H = 0$ , and to smaller  $p_O$  for  $p_H = 1$ ; this shift is likely to disappear as  $t \rightarrow \infty$ .

Figures 4(c) and (d) show the average local clustering coefficient,  $CC(G_O)$ , and transitivity (global clustering),  $\tau_{G_O}$ , for the GCCM. Patterns are similar between the two figures, although local clustering generally exceeds global clustering in the sparse regime. For the UCM it is known that, in the dense regime,  $\tau_{G_O}$  slowly converges to zero as  $t \rightarrow \infty$ , unless  $p = 1$  [2]. In contrast, the local clustering appears to remain non-zero.

As expected, clustering is minimised at  $p_H = p_O = 0$  (random tree) and maximised for a complete graph,  $p_H = p_O = 1$ . However, in the sparse regime we find that the maximum clustering is found at  $p_H = 1$ ,  $p_O = 0$  which corresponds to the CCM. In [2], the authors note that local and global clustering for the UCM is not a monotonically increasing function of the copying prob-

ability  $p$ , with a local maxima in the sparse regime at non-zero  $p$ . This bimodal clustering is also present in the GCCM. Interesting behaviour is observed in the anti-correlated regime for  $p_H \approx 0$ , where we find near zero clustering values. In particular if  $p_H = 0$  and  $p_O = 1$ , we observe the unusual property that  $\delta \approx 1$ , such that the network scales as (but is not) a complete graph, yet both the local and global clustering are approximately zero.

Extracting the degree distributions for the GCCM for various  $p_H$  and  $p_O$  shows similarly diverse behaviour, see Fig. 5. Each distribution is averaged over 100 instances, but points are left deliberately unbinned to illustrate the significant fluctuations observed in the dense regime. For  $p_H = p_O = 0$  (bottom left) the GCCM reduces to a random recursive tree, see Eq. (3). The CCM case with  $p_H = 1$ ,  $p_O = 0$  (top left) follows Eq. (14), where the tail can be approximated as a stretched exponential. This

distribution is also shown in Fig. 3. Along the diagonal where  $p_H = p_O$  (UCM), the degree distribution has a power-law tail in the sparse regime, and exhibits anomalous scaling in the dense regime ( $p \geq 0.5$ ). For  $p_H = p_O = 1$ , the GCCM reduces to a complete graph and all nodes have degree  $t - 1$ .

For  $p_O = 0$ , the power-law scaling observed in the UCM is completely suppressed, with a gradual transition from exponential decay to a stretched exponential tail as  $p_H$  is increased from 0 to 1. In the sparse regime with  $p_O \neq 0$ , all degree distributions appear fat tailed with only small deviations from the power-laws observed for the UCM. However, unusual scaling is observed for  $p_H = 0$ ,  $p_O \neq 0$ , where the distributions exhibit initial exponential decay at small  $k_O$ , attributable to the hidden network, before a second fat-tailed regime starting at intermediate  $k_O$ .

In the dense regime, all degree distributions exhibit anomalous scaling, such that individual instances are not self-averaging. For  $p_O = 0.75$ , the tail of the degree distributions is largely consistent across all  $p_H$ . However, the probability of finding nodes with small degree is large for  $p_H = 0$ , and is gradually suppressed as  $p_H \rightarrow 1$ . These effects are most pronounced for  $p_O = 1$  where the modal degree is 1 for  $p_H = 0$ , and  $t - 1$  for  $p_H = 1$ , with a gradual transition in between. Remarkably, throughout this transition the degree distribution appears almost uniform at  $p_H = 0.25$ , where the probability of finding nodes with any given degree is approximately constant up until the large  $k_O$  limit. However, this effect is only observed when averaging over many instances, with a much smaller degree range observed in individual networks.

Similar results are shown in the SM for  $q > 0$  where clustering is enhanced if  $p_H > p_O$  and suppressed if  $p_H < p_O$ , relative to the UCM. However, the clustering maxima slowly decrease with increasing  $q$ . In the limiting case of  $q = 1$ , the GCCM is independent of  $p_O$  and equivalent to the UCM with  $p = p_H$ . As  $q$  increases from  $0 \rightarrow 1$ , the GCCM becomes increasingly uniform, and the phase spaces shown in Fig. 4 are stretched from the  $p_H = p_O$  line outwards towards  $p_O = 0$  and  $p_O = 1$ . The  $p_H = p_O$  line (UCM) is invariant under changes in  $q$ .

One potential application of the  $q \neq 0$  case is for generating random simplicial complexes (see [37]) by combining the hidden and observed networks into a single structure, see Fig. 6. Such a construction may be interesting since it explicitly distinguishes between cliques of strong ties, where all nodes are within each other's inner circle, and cliques of weak ties.

Let us define the observed network,  $G_O$  as the network projection of the simplicial complex  $S$ , i.e.,  $S$  and  $G_O$  share the same set of edges. For every clique of size  $n$  that exists in  $G_O$ , if that same clique exists in  $G_H$ , then the clique maps to a simplex of size  $n$  in  $S$ . Otherwise, if the clique is not present in  $G_H$ , the corresponding simplex is also not present in  $S$ .

For the example shown in Fig. 6, the clique formed by the edges  $((a, b), (a, c), (a, d), (b, c), (b, d), (c, d))$  exists

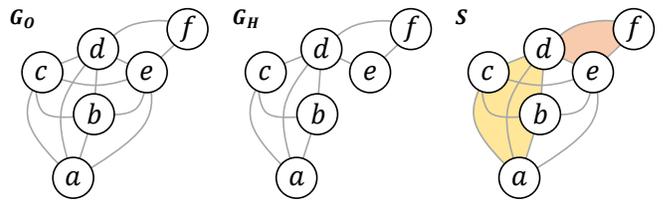


FIG. 6. Generating the simplicial complex  $S$  by combining the observed,  $G_O$ , and hidden network,  $G_H$ . The complex  $S$  contains all the edges from  $G_O$ . Cliques of size  $n$  which exists in both  $G_O$  and  $G_H$  map to a simplex of size  $n$  in  $S$ .

in both  $G_O$  and  $G_H$ . Hence, in  $S$ , the clique maps to the  $n = 4$  simplex  $(a, b, c, d)$  and all the corresponding sub-simplices. The same mechanism results in the  $n = 3$  simplex  $(d, e, f)$ . If a clique does not exist in both  $G_O$  and  $G_H$ , such as the triangle formed by the edges  $((a, b), (a, e), (b, e))$ , the corresponding simplex  $(a, b, e)$  is not found in  $S$ .

By tuning the GCCM copying parameter  $q$ , this mechanism can be adjusted from the trivial  $q = 0$  state where no simplices of size  $n > 2$  are found, to a state where every clique in  $G_O$  maps to an equivalently sized simplex in  $S$  for  $q = 1$ . Investigating the structure of, and dynamics on these simplicial complexes may be of future interest.

**Comparing Copying Models.** We have introduced a simple model of heterogeneous node copying, motivated by arguments that triadic closure may not be structurally homogeneous in real networks.

Comparing the CCM, for which we have analytical results, to the UCM with the equivalent effective copying probability ( $p = p_{\text{eff}} = 0.374$ ) we find significant differences in network structure. Most notably, both the average local clustering coefficient,  $CC(G_O)$ , and the transitivity,  $\tau_{G_O}$ , are significantly larger in the CCM than the UCM. The CCM suppresses the power-law tail observed in the UCM for the sparse regime, and consequently, the degree variance observed in the CCM is smaller than for the UCM. CCM:  $\sigma^2(k_O) \approx 26$ ; UCM:  $\sigma^2(k_O) \approx 192$ . The CCM also has the unusual property, not found in the UCM, that the growth in the number of cliques of size  $n$  scales independently of  $n$ .

The above comparison uses a single effective copying probability, but key differences are robust for variable  $p$  in the sparse regime. Specifically, the UCM degree distribution always exhibits a power-law tail, and the largest measured clustering coefficients fall below the values seen for the CCM, see Tab. I. Relaxing the CCM to the GCCM, we note that for large  $p_H$  and small  $p_O$ , the measured clustering values regularly exceed those observed in the UCM, with the UCM only reaching similar values far into the dense regime.

Whether such extreme bias is plausible in real networks is uncertain. However, observations in academic collaboration networks suggest that extreme bias may be possible [27]. For instance, Fig. 4 in [27] shows that the ratio of triadic closure between two nodes is approximately

	$p_{\text{eff}}$	$CC(G_O)$	$\tau_{G_O}$	Degree Dist.
CCM	0.37	0.77	0.43	Str. Exp.
UCM (Max $CC(G_O)$ )	0.38	0.52	0.20	Power-law
UCM (Max $\tau_{G_O}$ )	0.22	0.40	0.28	Power-law

TABLE I. The effective copying probability, average local clustering coefficient, transitivity and degree distribution for the CCM, and for the UCM in the sparse regime with the largest  $CC(G_O)$ , and largest  $\tau_{G_O}$  values. Calculated from 50 simulations with  $10^5$  nodes.

	Nodes	Edges	$CC(G_O)$	$\tau_{G_O}$
arXiv Astro coauthors	18.8K	198.1K	0.63	0.32
arXiv GR coauthors	5.2K	14.5K	0.53	0.63
arXiv CM coauthors	23.1K	93.4K	0.63	0.26
arXiv HEP coauthors	22.9K	2.7M	0.81	0.31
DBLP coauthors	540.5K	15.2M	0.80	0.65
NetSci coauthors	379	914	0.74	0.43
Hollywood collaborations	1.1M	56.3M	0.77	0.31
DNC Email corecipients	906	12.1K	0.61	0.56

TABLE II. A selection of sparse undirected networks which may plausibly grow via a copying mechanism. These networks exhibit larger average local clustering coefficients ( $CC(G_O)$ ) and transitivity ( $\tau_{G_O}$ ) than may be expected from homogeneous copying. Network source data available at [40].

zero if the number of shared collaborators is zero, rises rapidly as the number of shared collaborators increases, and plateaus at a ratio of one.

A second clue towards heterogeneous copying is the observation of very large clustering values in real networks. A selection of these networks and their clustering coefficients is shown in Tab. II. Stressing that both the UCM and GCCM are toy models of node copying, the networks in Tab. II exhibit average local and/ or global clustering far exceeding even the most optimistic values for the UCM. In contrast, the listed clustering values are relatively similar to what may plausibly emerge from

heterogeneous copying, although even the clustering observed for the extreme CCM case falls below some of the values shown in Tab. II, implying that other factors are also important.

We must of course acknowledge that the networks listed in Tab. II have been chosen precisely due to their high clustering; there are many other networks which could plausibly grow via copying which do not exhibit unusually large clustering coefficients. However, the networks provided do suggest that heterogeneous copying may be important in some contexts.

**Conclusion.** We have introduced a general model of heterogeneous copying, implemented using a hidden network model. In the case of extreme copying bias, we have derived analytical results and have demonstrated significant differences to similar models with uniform copying rules. In particular, power-law degree distributions observed in uniform copying can be suppressed under heterogeneous copying, and networks are significantly more clustered if copying is biased towards a node’s inner circle. Although a systematic study of copying in real networks is necessary, evidence suggests that heterogeneous copying may be relevant in a social context.

The heterogeneous copying model is just one simple application of a hidden network model. In general, the framework allows us to deconstruct network growth heterogeneities in a non-arbitrary way, focusing on structural rather than node heterogeneity, and poses questions concerning the role of hidden information in network growth. Exploring these general cases and answering these questions is a key aim in upcoming work.

**Data accessibility.** A python implementation of the CCM and UCM is available at [41].

**Acknowledgements.** I am grateful to Tim S. Evans, Chester Tan and Kim Christensen for a number of useful discussions, and to Tim S. Evans and Chester Tan for proof reading the manuscript. I acknowledge a Ph.D. studentship from the Engineering and Physical Sciences Research Council through Grant No. EP/N509486/1.

- 
- [1] M. E. J. Newman, The structure and function of complex networks, *SIAM Review* **45**, 167 (2003).
- [2] U. Bhat, P. Krapivsky, R. Lambiotte, and S. Redner, Densification and structural transitions in networks that grow by node copying, *Physical Review E* **94**, 062302 (2016).
- [3] R. Lambiotte, P. Krapivsky, U. Bhat, and S. Redner, Structural transitions in densifying networks, *Physical Review Letters* **117**, 218301 (2016).
- [4] U. Bhat, P. L. Krapivsky, and S. Redner, Emergence of clustering in an acquaintance model without homophily, *Journal of Statistical Mechanics: Theory and Experiment* **2014**, P11035 (2014).
- [5] G. Bianconi, R. K. Darst, J. Iacovacci, and S. Fortunato, Triadic closure as a basic generating mechanism of communities in complex networks, *Physical Review E* **90**, 042806 (2014).
- [6] J. Davidsen, H. Ebel, and S. Bornholdt, Emergence of a small world from local interactions: Modeling acquaintance networks, *Physical Review Letters* **88**, 128701 (2002).
- [7] M. K. Hassan, L. Islam, and S. A. Haque, Degree distribution, rank-size distribution, and leadership persistence in mediation-driven attachment networks, *Physica A: Statistical Mechanics and its Applications* **469**, 23 (2017).
- [8] R. Toivonen, J.-P. Onnela, J. Saramäki, J. Hyvönen, and K. Kaski, A model for social networks, *Physica A: Statistical Mechanics and its Applications* **371**, 851 (2006).
- [9] A. Asikainen, G. Iñiguez, J. Ureña-Carrión, K. Kaski,

- and M. Kivelä, Cumulative effects of triadic closure and homophily in social networks, *Science Advances* **6**, eaax7310 (2020).
- [10] N. Farid and K. Christensen, Evolving networks through deletion and duplication, *New Journal of Physics* **8**, 212 (2006).
- [11] R. Pastor-Satorras, E. Smith, and R. V. Solé, Evolving protein interaction networks through gene duplication, *Journal of Theoretical Biology* **222**, 199 (2003).
- [12] F. Chung, L. Lu, T. G. Dewey, and D. J. Galas, Duplication models for biological networks, *Journal of Computational Biology* **10**, 677 (2003).
- [13] P. L. Krapivsky and S. Redner, Network growth by copying, *Physical Review E* **71**, 036118 (2005).
- [14] C. Steinbock, O. Biham, and E. Katzav, Distribution of shortest path lengths in a class of node duplication network models, *Physical Review E* **96**, 032301 (2017).
- [15] A. Vázquez, A. Flammini, A. Maritan, and A. Vespignani, Modeling of protein interaction networks, *Complexity* **1**, 38 (2003).
- [16] P. Holme and B. J. Kim, Growing scale-free networks with tunable clustering, *Physical Review E* **65**, 026107 (2002).
- [17] A. Vázquez, Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations, *Physical Review E* **67**, 056104 (2003).
- [18] T. P. Peixoto, Disentangling homophily, community structure and triadic closure in networks, arXiv preprint arXiv:2101.02510 (2021).
- [19] F. Battiston, J. Iacovacci, V. Nicosia, G. Bianconi, and V. Latora, Emergence of multiplex communities in collaboration networks, *PloS One* **11**, e0147451 (2016).
- [20] S. R. Goldberg, H. Anthony, and T. S. Evans, Modelling citation networks, *Scientometrics* **105**, 1577 (2015).
- [21] R. I. M. Dunbar, The social brain hypothesis, *Evolutionary Anthropology: Issues, News, and Reviews* **6**, 178 (1998).
- [22] P. Mac Carron, K. Kaski, and R. Dunbar, Calling dunbar's numbers, *Social Networks* **47**, 151 (2016).
- [23] C. R. McClain, Practices and promises of facebook for science outreach: Becoming a "nerd of trust", *PLoS biology* **15**, e2002020 (2017).
- [24] D. Easley, J. Kleinberg, *et al.*, *Networks, crowds, and markets* (Cambridge university press Cambridge, 2010) Chap. Strong and Weak Ties.
- [25] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási, Structure and tie strengths in mobile communication networks, *Proceedings of the national academy of sciences* **104**, 7332 (2007).
- [26] C. Scholz, M. Atzmueller, M. Kibanov, and G. Stumme, Predictability of evolving contacts and triadic closure in human face-to-face proximity networks, *Social Network Analysis and Mining* **4**, 217 (2014).
- [27] J. Kim and J. Diesner, Over-time measurement of triadic closure in coauthorship networks, *Social Network Analysis and Mining* **7**, 9 (2017).
- [28] L. Zhou, Y. Yang, X. Ren, F. Wu, and Y. Zhuang, Dynamic network embedding by modeling triadic closure process, in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32 (2018).
- [29] T. Raducha, B. Min, and M. San Miguel, Coevolving nonlinear voter model with triadic closure, *EPL (Europhysics Letters)* **124**, 30001 (2018).
- [30] G. Bianconi, *Multilayer networks: structure and function* (Oxford university press, 2018).
- [31] D. Cellai and G. Bianconi, Multiplex networks with heterogeneous activities of the nodes, *Physical Review E* **93**, 032302 (2016).
- [32] J. Gao, S. V. Buldyrev, H. E. Stanley, and S. Havlin, Networks formed from interdependent networks, *Nature Physics* **8**, 40 (2012).
- [33] M. Falkenberg, J.-H. Lee, S.-i. Amano, K.-i. Ogawa, K. Yano, Y. Miyake, T. S. Evans, and K. Christensen, Identifying time dependence in network growth, *Physical Review Research* **2**, 023352 (2020).
- [34] In [33], the hidden network is referred to as the influence network.
- [35] M. Boguná and R. Pastor-Satorras, Class of correlated random networks with hidden variables, *Physical Review E* **68**, 036112 (2003).
- [36] H. Hartle, F. Papadopoulos, and D. Krioukov, Dynamic hidden-variable network models, arXiv preprint arXiv:2101.00414 (2021).
- [37] F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J.-G. Young, and G. Petri, Networks beyond pairwise interactions: structure and dynamics, *Physics Reports* 10.1016/j.physrep.2020.05.004 (2020).
- [38] J.-G. Young, G. Petri, and T. P. Peixoto, Hypergraph reconstruction from network data, arXiv preprint arXiv:2008.04948 (2020).
- [39] P. L. Krapivsky, S. Redner, and F. Leyvraz, Connectivity of growing random networks, *Physical Review Letters* **85**, 4629 (2000).
- [40] R. A. Rossi and N. K. Ahmed, The network data repository with interactive graph analytics and visualization, in *AAAI* (2015).
- [41] Code: [github.com/MaxFalkenberg/RandomCopying](https://github.com/MaxFalkenberg/RandomCopying).

# Figures

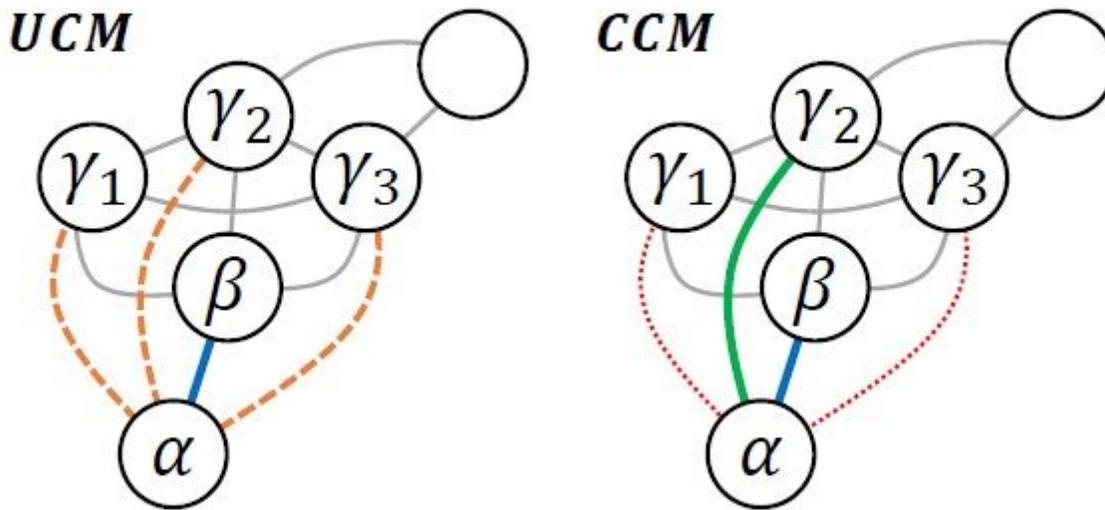


Figure 1

A new node  $\alpha$  is added to the network and forms an edge to a random target node,  $\beta$ . UCM: All neighbors of node  $\beta$  have an equal probability  $p$  of being copied (orange dashed). CCM: Copied edges are added deterministically; neighbors of node  $\beta$  in the hidden network are copied (solid green), the remainder are not copied (red dotted).

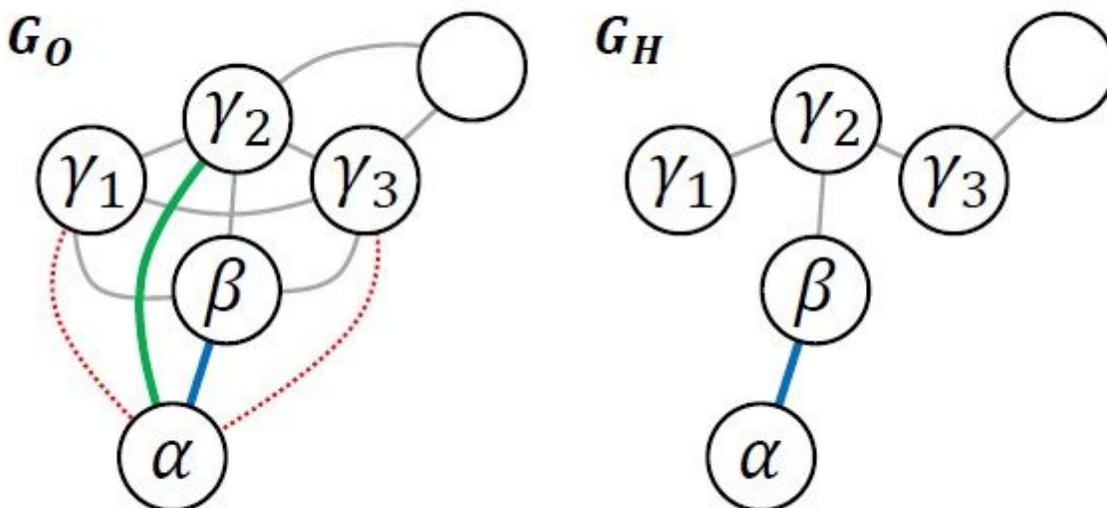
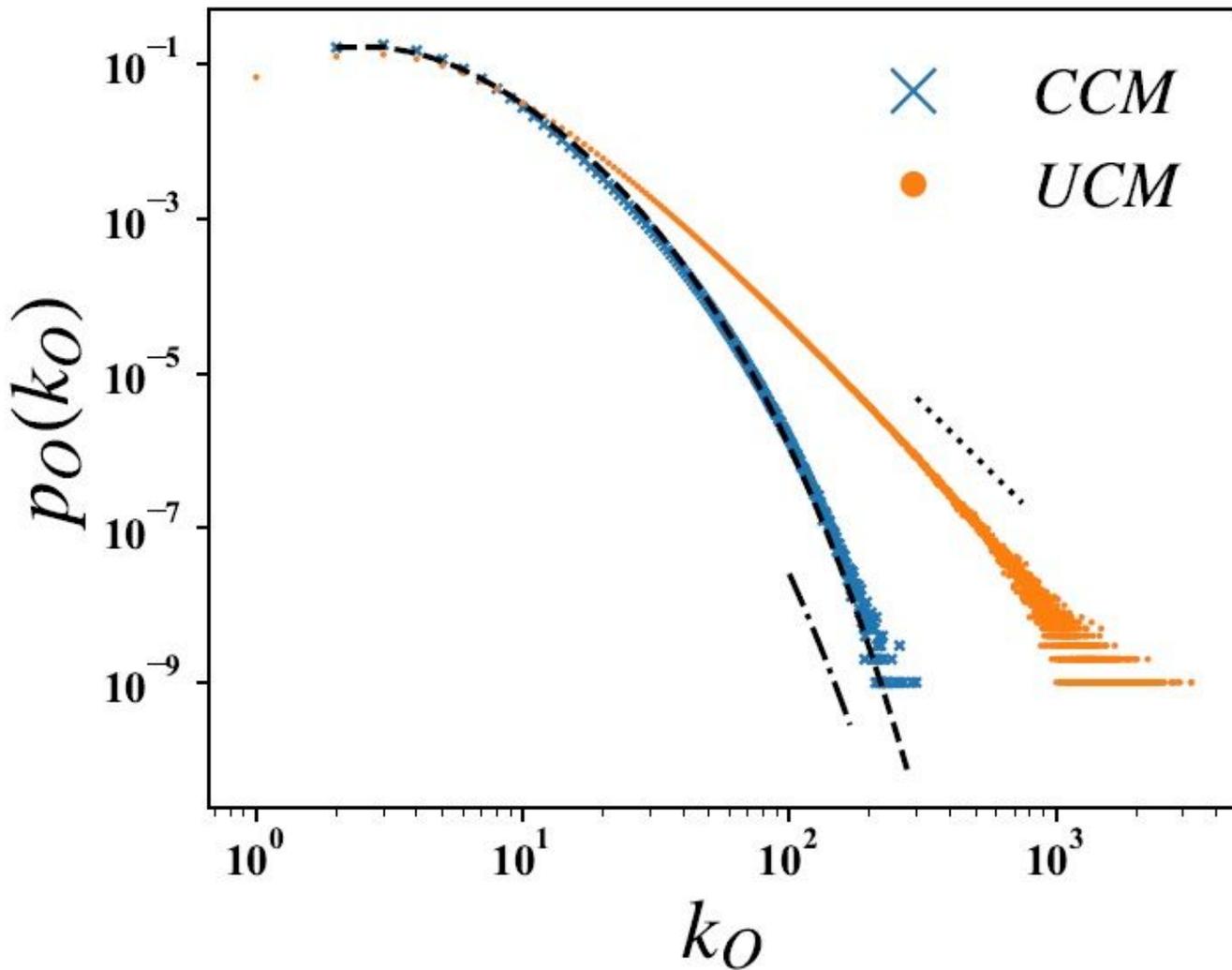


Figure 2

The correlated copying model. A new node,  $a$ , forms an edge (solid blue) to a randomly chosen target node  $B$ . Copied edges ( $a; j$ ) (solid green) are formed in the observed network,  $G_O$ , if the edge ( $a; j$ ) is present in the hidden network,  $G_H$ . Other neighbors of node  $B$  are not copied (dotted red).



**Figure 3**

The degree distribution for the CCM (blue crosses) and UCM ( $p = p_{eff}$ ; orange points) at  $t = 107$ , averaged over 100 networks. Dashed line: CCM recurrence relation in Eq. (14). Dot-dashed: stretched exponential approximation. Dotted: power-law scaling.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [InnerCirclePaperCommsPhysicsSupplement.pdf](#)

- [24521supp55446qrk6d3convrt.pdf](#)
- [24521supp55446qrk6d3convrt.pdf](#)